

Article

Identification and Analysis of Unstructured, Linear B-Cell Epitopes in SARS-CoV-2 Virion Proteins for Vaccine Development

Andrés Corral-Lugo ^{1,†}, Mireia López-Siles ^{1,†}, Daniel López ², Michael J. McConnell ^{1,*‡} and Antonio J. Martín-Galiano ^{1,*‡}

¹ Intrahospital Infections Unit, National Centre for Microbiology, Instituto de Salud Carlos III (ISCIII), 28220 Madrid, Spain; acorral@isciii.es (A.C.-L.); mireia.lopez@isciii.es (M.L.-S.)

² Immune Presentation and Regulation Unit, Instituto de Salud Carlos III, 28220 Madrid, Spain; dlopez@isciii.es

* Correspondence: michael.mcconnell@isciii.es (M.J.M.); mgaliano@isciii.es (A.J.M.-G.); Tel.: +34-91-822-3869 (M.J.M.); +34-91-822-3976 (A.J.M.-G.)

† These authors contributed equally to this work.

‡ These authors contributed equally to this work.

Received: 18 May 2020; Accepted: 17 July 2020; Published: 20 July 2020



Abstract: The efficacy of SARS-CoV-2 nucleic acid-based vaccines may be limited by proteolysis of the translated product due to anomalous protein folding. This may be the case for vaccines employing linear SARS-CoV-2 B-cell epitopes identified in previous studies since most of them participate in secondary structure formation. In contrast, we have employed a consensus of predictors for epitopic zones plus a structural filter for identifying 20 unstructured B-cell epitope-containing loops (uBCEPs) in S, M, and N proteins. Phylogenetic comparison suggests epitope switching with respect to SARS-CoV in some of the identified uBCEPs. Such events may be associated with the reported lack of serum cross-protection between the 2003 and 2019 pandemic strains. Incipient variability within a sample of 1639 SARS-CoV-2 isolates was also detected for 10 uBCEPs which could cause vaccine failure. Intermediate stages of the putative epitope switch events were observed in bat coronaviruses in which additive mutational processes possibly facilitating evasion of the bat immune system appear to have taken place prior to transfer to humans. While there was some overlap between uBCEPs and previously validated SARS-CoV B-cell epitopes, multiple uBCEPs had not been identified in prior studies. Overall, these uBCEPs may facilitate the development of biomedical products for SARS-CoV-2.

Keywords: coronavirus; COVID-19; epitope prediction; epitope switch; immunoinformatics; unfolded protein response; SARS-CoV

1. Introduction

In December 2019, Wuhan, China became the center of an outbreak of febrile respiratory illness due to a new type of coronavirus [1]. Genome sequencing showed genetic similarities to other coronaviruses found in humans [2], particularly with Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), resulting in the designation of this virus as SARS-CoV-2 by the World Health Organization (WHO) [3]. The number of globally confirmed cases according to the WHO reached 9,129,146 and 473,797 deaths by 24 June 2020.

Human coronaviruses, first characterized in the 1960s, are responsible for upper respiratory tract infections [4], and can infect both human and animal hosts [4,5]. In addition to the current pandemic, coronaviruses have previously caused two outbreaks involving respiratory infections with

great repercussions. In 2003, SARS-CoV affected more than 8000 people in 25 countries across five continents [6] and, in 2012, the Middle East Respiratory Syndrome coronavirus infected more than 1000 patients with a mortality rate of more than 35% [7].

Similarly to other coronaviruses, SARS-CoV-2 has a ~30 kb single-stranded, positive-sense RNA genome containing genes that encode homologs for at least four main structural proteins of the viral particle: spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins [8]. Functional and immunological properties of these proteins have been well-studied in SARS-CoV since 2003. For instance, the S protein is a glycoprotein that forms homotrimers and consists of two functional subunits responsible for binding to the host cell receptor (S_1 subunit) and for the fusion of the viral and cellular membranes (S_2 subunit), as also confirmed in SARS-CoV-2 [9]. The distal S_1 subunit contains the receptor-binding domain (S_{1B}) necessary for attachment to the ACE2 receptor and entry into type II pneumocytes and other host cells. Active immunization with full-length and truncated S protein [10], S protein peptides [11], and chimeric versions of the S protein [12] has been characterized for SARS-CoV. DNA constructs encoding the S protein produce neutralizing antibodies against the virus [13,14]. Therefore, the S protein of both SARS-CoV and SARS-CoV-2 is a major target for vaccine development. The small multifunctional E protein is important for pathogenesis and different steps in the virus life cycle (assembly, budding, and envelope formation). Although only a small portion of E is incorporated into the virion, it is abundantly expressed inside the infected cell [15]. The E protein is recognized by SARS-CoV convalescent sera [16]; however, limited information is available regarding its antigenic properties.

The transmembrane M glycoprotein is the most abundant structural protein in the mature coronavirus virion and has been suggested to play a major role in envelope formation in SARS-CoV [17]. IgM and IgG antibodies against the M protein are present in sera from patients with SARS-CoV [18], and high titers of these antibodies are elicited in rabbits immunized with the N-terminal region of the protein [10,19,20]. The N protein contains an amino-terminal RNA binding domain and a carboxyl-terminal dimerization domain. This protein is involved in envelope formation, regulation of viral RNA synthesis, packaging of viral RNA, and may play an important role in overcoming host defense by suppressing RNA interference mechanisms [21,22]. The SARS-CoV N protein is highly immunogenic and antigenic sites have been described throughout the entire sequence [18,23,24]. Moreover, N protein is an early diagnostic marker for SARS-CoV because it is detectable in clinical samples as early as one day after the onset of symptoms [25].

It is clear that the fight against SARS-CoV-2 requires a number of clinical approaches that are not currently available. The identification of viral B-cell epitopes, which are the key elements that trigger the protective humoral immune response, can facilitate the design and development of vaccines, rapid diagnostic tests, and antibody-based therapeutics. In addition, characterization of such epitopes can contribute to our understanding of mutational changes that affect the ability of the immune response to provide cross protection against related viruses. Multiple immunoinformatic approaches have been developed for the prediction of B-cell epitopes based on different criteria that aim to capture the intrinsic complexity of the binding between the antigenic epitope and the antibody paratope [26]. However, given that the sensitivity for detection of linear epitopes using computational approaches has been estimated to be around 60% [27], the integration of several methods may identify physiologically relevant B-cell epitopes more accurately. In addition to the discontinuous and structured nature of many B-cell epitopes [28], numerous examples of linear epitopes located in loops and inducing immunoprotective humoral responses and their recognition by antibodies in laboratory assays have been described for relevant viral pathogens [29–31], highlighting their potential utility in biomedical applications.

Given the inherent difficulty of experimentally mapping B-cell epitopes [32], several studies have used computational approaches to predict B-cell epitopes from SARS-CoV [13,33–37]. Multiple B-cell epitopes from SARS-CoV have been experimentally validated since 2003 [38] and included in the Immune Epitope Database (IEDB), a central repository that stores, catalogs, and assists in the prediction

and analysis of epitopes [39]. In contrast to SARS-CoV, there is still relatively little information available regarding B-cell epitopes from SARS-CoV-2. Nevertheless, both *in vitro* [40] and *in silico* analyses of proteins S [41–45], M and N [46] have been conducted to determine sequence variation, antigenic regions, and targets of the immune responses in SARS-CoV-2.

A myriad of vaccine initiatives for SARS-CoV-2 based on different technologies are currently being undertaken [47]. One of the most promising approaches is the utilization of DNA vaccines coding for viral epitopic regions from different antigens. However, these highly-complex chimeric proteins include incomplete sections of protein folds that may adopt aberrant structural arrangements when present in isolation. This situation may trigger the cellular unfolded protein response (UPR) in host cells. During the UPR, hydrophobic residues normally buried in the context of the full-length antigen may be exposed and sensed in the endoplasmic reticulum by the GRP78 chaperone, which labels the anomalous protein for cytoplasm back-translocation and ubiquitin-mediated proteolysis [48]. The presence of misfolded sequences in multiepitope vaccines may therefore result in antigen degradation, potentially decreasing the ability of the antigen to induce a robust immune response.

To address this problem, and in contrast to recent reports predicting B-cell epitopes in SARS-CoV-2, the present study focuses on the identification and characterization of unstructured B-cell epitope-containing loops (uBCEs) of virion proteins in order to avoid triggering the UPR. These sequences may therefore be ideally suited for the development of diagnostic, therapeutic, and preventive technologies for SARS-CoV-2 infection, particularly multiepitope vaccines.

2. Materials and Methods

SARS-CoV and SARS-CoV-2 proteins utilized in this study were those in the reference proteome provided by NCBI entries NC_004718 [49] and NC_045512 [50], respectively. For a detailed bibliography of the databases and methodology used, see Supplementary Table S1.

2.1. Identification of uBCEs

Linear B-cell epitopes (BCEs) were predicted by the SVM2 model of APPPred with a score ≥ 0 ; ABCPred with a score ≥ 0.8 ; Bepipred applying a score of ≥ 0.35 ; Bepipred 2.0, applying a score of ≥ 0.5 ; Kolaskar's antigenicity applying a score threshold ≥ 0.988 ; LBEEP using window length of 15 residues and a score of 0.7; and SVMtrip, applying a score of ≥ 0.35 . In addition, an eighth method consisted of collectively analyzing the average value resultant from six physicochemical predictors related to B-cell antigenicity: Emini accessibility, Janin surface exposure, Karplus & Schulz flexibility, Parker hydrophilicity, Pellenquer turns and Ponnuswamy polarity, applying an average value of 1.2 as threshold.

Sequence spans of ≥ 6 residues (including a maximum of 1 internal non-predicted residue), supported by at least four out of eight algorithms were considered as candidate linear B-cell epitopes. These preliminary epitopes were filtered by resolved or predicted secondary protein structure, and only those with ≥ 6 residues in non-folded regions were denoted unstructured BCEs (uBCEs). Finally, the entire unstructured loop containing each uBCE was considered as uBCEs.

Antigenicity was predicted using the iLBE (<http://kurata14.bio.kyutech.ac.jp/iLBE/>) and VaxiJen 2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) servers.

2.2. Structural and Accessory Analysis of Protein Antigens

Secondary structure analysis was carried out using available three-dimensional structures that covered most of the S and N proteins: 6M3M (unpublished) (41–174 residues, 100% identical) and 2CJR [51] (250–360 residues, 95% identical) PDB entries for N protein; and the 6VXX_A entry (14–1211 residues, 100% identical) for S protein [9]. For E and M proteins, structures were not available and models were deemed low-quality by Swiss Model curators (<https://swissmodel.expasy.org/repository/species/2697049>), and therefore were not considered. For regions and proteins without available experimental structural information, secondary structure was predicted by PSI-PRED 4.0. Protein

domains were identified with Pfam 32.0 applying gathering thresholds; transmembrane helices (TMHs) were predicted by Phobius; signal peptides were predicted by SignalP 5.0 applying the Eukarya model; coiled-coil regions were predicted by COILS using the MTIDK matrix, and applying a threshold of 0.5 with 14, 21 and 28 residue windows; disordered regions were predicted by the long model of IUPred2, using 0.5 as threshold, and by DISOPRED 3.1, finally considering disordered regions as those with ≥ 5 residues predicted by at least one of these methods.

2.3. Phylogenetic Analyses

All non-redundant coronavirus homologs of the Identical Protein Groups database for the four selected proteins were detected by BLAST using the reference S, E, M, and N SARS-CoV-2 sequences as queries. A 70% identity threshold, 90% alignment length coverage, and E value $< 10^{-5}$ were applied. Partial hit sequences and those containing 'X's were removed. Updated SARS-CoV-2 protein sequences were obtained from the coding region files provided by the NCBI. Then, the dataset was reduced on a 95% identity and a 95% reciprocal protein length coverage basis using CD-HIT. Alignments were carried out with Clustal Omega. The evolutionary history was calculated by the Neighbor-Joining method using Mega 7. Branch lengths were proportional to the tree evolutionary distances calculated using the Poisson correction method. The rate variation among sites was modeled with a gamma distribution applying a shape parameter of 1. Phylogeny was tested by 1000 bootstrap replicates. SARS-CoV and SARS-CoV-2 clades were identified through the location of their respective reference sequences.

Sequences from SARS-CoV and SARS-CoV-2 clades were re-aligned with Clustal Omega and SDPs identified with SDPpred applying 1000 shuffles and a Bernoulli estimator cutoff of 0.

2.4. Epitope Collection in SARS-CoV-2

For SARS-CoV, experimentally determined and predicted epitopes of S, E, M, and N proteins were obtained from the IEDB. The following filters were applied: "linear epitope", "B-cell assay", "severe acute respiratory syndrome-related coronavirus organism (ID: 694009, SARS)", "severe acute respiratory syndrome disease data (ID: DOID:2945, SARS)", "any MHC restriction", and "any host". The published studies describing these epitopes were identified and screened in order to include only epitopes that had been validated as being able to stimulate an antibody response in experimental models. Published epitopes of SARS-CoV-2 were identified based on the bibliography available in PubMed up to 31 March 2020. The following keywords were used: "S, E, M or N protein", "epitope", "SARS-CoV-2 or COVID-19". Only those epitopes with less than 100 amino acids were considered.

2.5. Statistical Analyses

To compare length and percentage of residues in unstructured regions of previously reported SARS-CoV-2 epitopes, the Shapiro–Wilk test was firstly used to assess data normality and Levene's test used to assess for equality of variances. Given the non-normal distribution and the lack of homoscedasticity, the non-parametric Kruskal–Wallis test was applied. Significance of the occurrence of SDPs and indels in uBCELS with respect to the rest of the protein sequence was assessed by the chi-square test. Statistical analyses were conducted via the SPSS 15.0 statistical package for Windows (IBM Corp., Armonk, NY, USA). Significance levels were established for p -values ≤ 0.05 .

3. Results

3.1. SARS-CoV-2 Epitope Catalogue

The four SARS-CoV-2 structural proteins were predicted as antigens by the iLBE and VaxiJen methods (data not shown). A screen of SARS-CoV-2 linear B-cell epitopes reported in the literature up to Apr 08th 2020 was performed. A total of 37 epitopes from seven studies were included (Supplementary Table S2). Of these, more than 85% were located in the S protein ($n = 32$), with the remaining in the N ($n = 2$) or M ($n = 3$) proteins (Figure 1a). The predominance of epitopes identified

in the S protein may be explained by bias of the included studies towards utilization of only the S protein and the fact that the S protein may be more suited for epitope identification given its length, structure, and higher solvent exposure in the virus capsid. To the best of our knowledge, no linear B-cell epitopes have been reported for the SARS-CoV-2 E protein.

Up to ten epitopes were identified in the included studies, with a median length of 12 amino acids. Epitopes ranged from five to 75 residues, with different studies demonstrating significant variations when their epitope lengths were compared ($p < 0.001$), indicating a mixture of pure epitopes and wide epitopic zones. Thirty protein epitopes (81.1%) from these studies showed significant overlap of their sequence (between 10 and 100%) with protein structured elements (either alpha-helices or beta-strands), with no significant difference between studies included in the analysis ($p > 0.928$).

For the S protein, the only protein analyzed in more than one study, we further characterized the agreement of reported epitopes for the seven articles included in the analysis. This meta-analysis revealed that residues were identified as contributors of predicted epitope elements by, at most, two independent studies (Figure 1b). The three-dimensional positioning of these residues further indicated that they do not aggregate either into domains or regions, but are dispersed over the entirety of the S protein structure (Figure 1c). Together, these analyses indicate that there is little agreement between studies regarding the linear SARS-CoV-2 B-cell epitopes in the published literature. Furthermore, the proposed epitopes identified in these studies tend to include portions of structured elements, which may misfold and trigger the UPR when expressed in the endoplasmic reticulum.

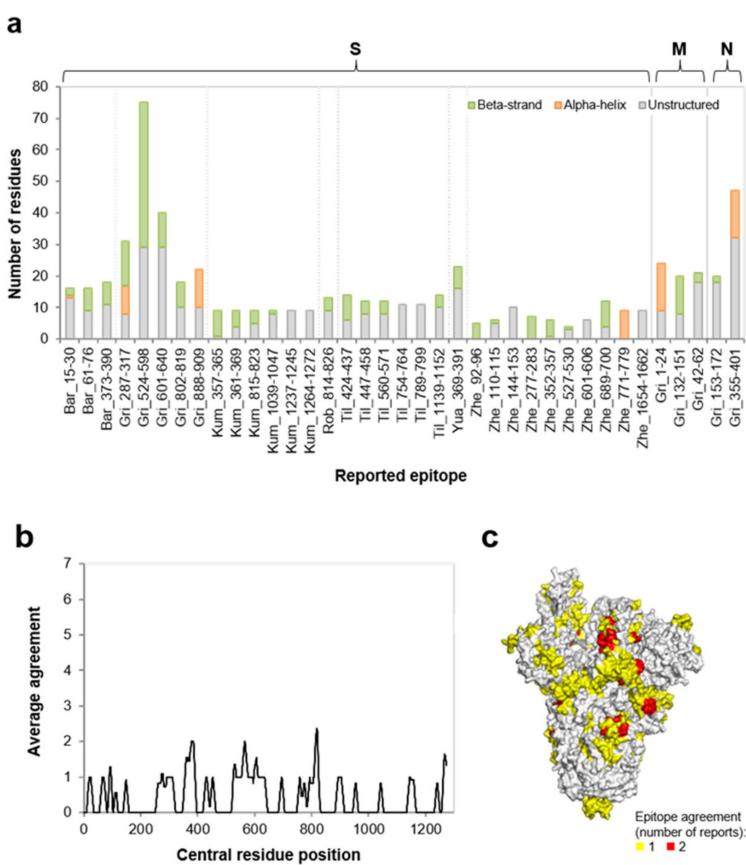


Figure 1. Characteristics of B-cell epitopes on SARS-CoV-2 proteins reported in the literature. (a) length and predicted secondary structure content of reported epitopes ($n = 37$) by protein and study. Report prefixes: Bar [44], Gri [46], Kum [41], Rob [45], Til [42], Yua [40], and Zhe [43]. Experimentally reported or predicted alpha-helices and beta-strands were considered; (b) average agreement of reported epitopes for S protein between the seven studies. The average value of a sliding window of

11 residues was calculated and this value assigned to the residue in the central position; (c) surface view of the S protein three-dimensional structure showing color-ranked epitope agreement.

3.2. Unstructured Epitope Selection to Design Antigenic Peptides and Chimera Proteins

A plausible strategy for generating antibodies against SARS-CoV-2 using peptide or multipeptide antigens involves the utilization of unstructured sections harboring B-cell epitopes from the four virion proteins. Loops containing linear B-cell epitopes were identified in these four proteins using a three-stage computational pipeline (Figure 2). In a first step, and given the low sensitivity of the detection of these epitopes, a consensus of eight available prediction methods was designed (see Methods). This inclusive approach is expected to cover technical predictive aspects as well as physicochemical nuances of these types of epitopes. Second, epitopes satisfying the initial criteria were placed into structural context. To avoid potentially triggering the UPR by misfolded sequences, only predicted linear BCEs located in loops, i.e., uBCEs, were selected. All uBCEs contained at least six unstructured residues, since most B-cell linear epitopes fall within the 6–10 residue range [52]. A total of 21 unstructured B-cell epitopes were identified using this approach. These were extended to the whole disordered region to render the corresponding uBCEs (Table 1). Eleven uBCEs were detected in the S, one in M, and eight in N proteins for a total of 20 uBCEs since two epitopes were located in the same loop. In addition, a structured exception involving a B-cell epitope in an alpha-helical section in the E protein (BCEH-E1) was also considered (Table 1).

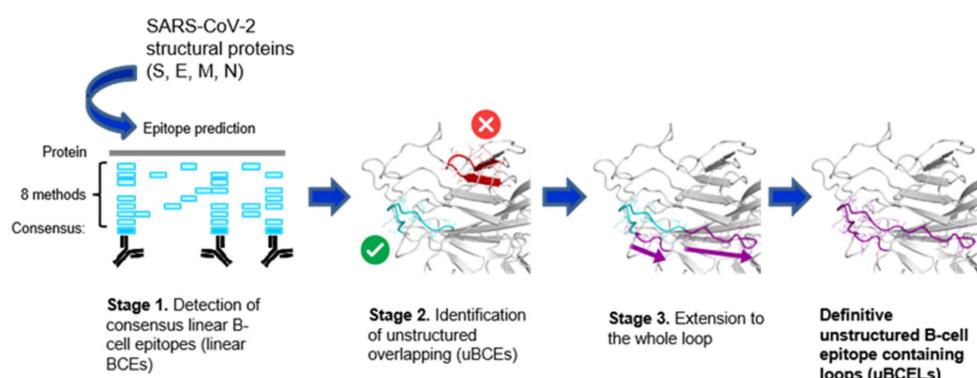


Figure 2. In silico pipeline for the identification of uBCEs. A schematic three-stage flowchart to define uBCEs in SARS-CoV-2 virion proteins is shown. First, eight bioinformatics methods were collectively utilized to locate, by consensus, the linear B-cell epitopes in S, E, M, and N proteins. Then, those B-cell epitopes that co-located to regular secondary structure elements, either alpha-helices or beta-strands, were rejected. Finally, the location of B-cell epitopes that satisfied the former criteria was extended to cover the complete loop.

Comparison of these uBCEs to reported SARS-CoV-2 epitopes through a residue coverage matrix showed that overlap between uBCEs and previously identified epitopes was poor or null (Figure 3a). Jaccard indices ≤ 0.25 , i.e., overlapping residues were below a quarter of the total, were obtained in all cases except for the N-protein when compared to the dataset from Grifoni and colleagues [46] that reached 0.41.

Table 1. Epitopes identified in SARS-CoV-2 virion proteins.

Protein	uBECL or BCEH ^a	uBCE ^b Location	uBCEL or BCEH Location	Flanking SS ^c	uBCEL Sequence ^d
S	uBCEL-S1	21–28	16–28	SP-B1	vnlttRTQLPPAY
	uBCEL-S2	71–81	68–85	B3-B4	ihvSGTNGTKRFDNpvlp
	uBCEL-S3	404–412	402–429	B25-B26	irGDEVRQIAPgqtgkiadynyklpddf
	uBCEL-S4	440–445	440–450	B26-B27	NLDSDKVggyn
		459–470			
	uBCEL-S5	473–480	455–491	B27-B28	IfrkSNLKPFERDISTeYQAGSTPCngvegfncyfp
	uBCEL-S6	615–630	615–642	B38-B39	VNCTEVPVAIHADQLTpwrwystgsnv
	uBCEL-S7	676–687	676–689	B43-B44	TQTNSPRRARSVAs
	uBCEL-S8	783–797	783–803	H3-B48	AQVKQIYKTPPIKDFggfnfs
	uBCEL-S9	1125–1131	1125–1131	B60-B61	NCDVVIG
	uBCEL-S10	1137–1147	1136–1147	B61-H12	TVYDPLQPPELDS
	uBCEL-S11	1240–1246	1238–1246	H15p-H16p	tsCCSCLKG
E	BCEH-E1	57–68	38–75	H3p	rlcayccnivnsvlkpsfYVYSRVKNLNSSRvpdllv
M	uBCEL-M1	209–215	209–222	B10-Ct	DHSSSSDniallvq
N	uBCEL-N1	16–48	18–55	B1p-B2	GGPSDSTGSQNQNGERSGARSKQRRPQGLPNNTASWFTA
	uBCEL-N2	59–78	59–78	B2-H1	HGKEDLKFPFRGQGVINTNS
	uBCEL-N3	158–170	135–170	B8-B9	tegalntpkdhigtrnpannaaiVLQLPQGTTLPKG
	uBCEL-N4	173–208	173–213	B9-H2p	AEGSRGGSQASSRSSRSRNSSRNSTPGSSRGTSParagn
	uBCEL-N5	235–247	235–247	H2p-H3p	SGKGQQQQGQTVT
	uBCEL-N6	276–287	276–287	H4-H5	RRGPEQTQGNFG
	uBCEL-N7	339–344	339–344	B11-H9	LDDKDP
	uBCEL-N8	363–383	363–383	H10-H11p	FPPTEPKDKKKKADETQALP

^a uBCEL: unstructured B-cell epitope loop. BCEH: B-cell epitope helix. ^b uBCE: unstructured B-cell epitopes. ^c SS: secondary structure; B: beta-strand (plus its index in the protein); H: alpha-helix (plus its index in the protein); SP: signal peptide; p: predicted by PSIPRED; Ct: C-terminus. ^d The uBCE zone is in capitals. The rest of the loop is in lowercase.

It is of interest to assess how uBCEls have evolved between SARS-CoV and SARS-CoV-2 with respect to the rest of the protein. In the S protein, uBCEls covered 15.7% of the protein sequence (excluding the 15mer signal peptide). However, they were significantly enriched (36.7%) in insertions and specificity determining positions (SDPs)—those positions with conserved exclusive residues in each clade, which greatly determine phylogenetic tree structure [53] ($p < 0.0001$). This may suggest higher evolutionary dynamics in these sequences, perhaps due in part to the selective pressure for epitope switching.

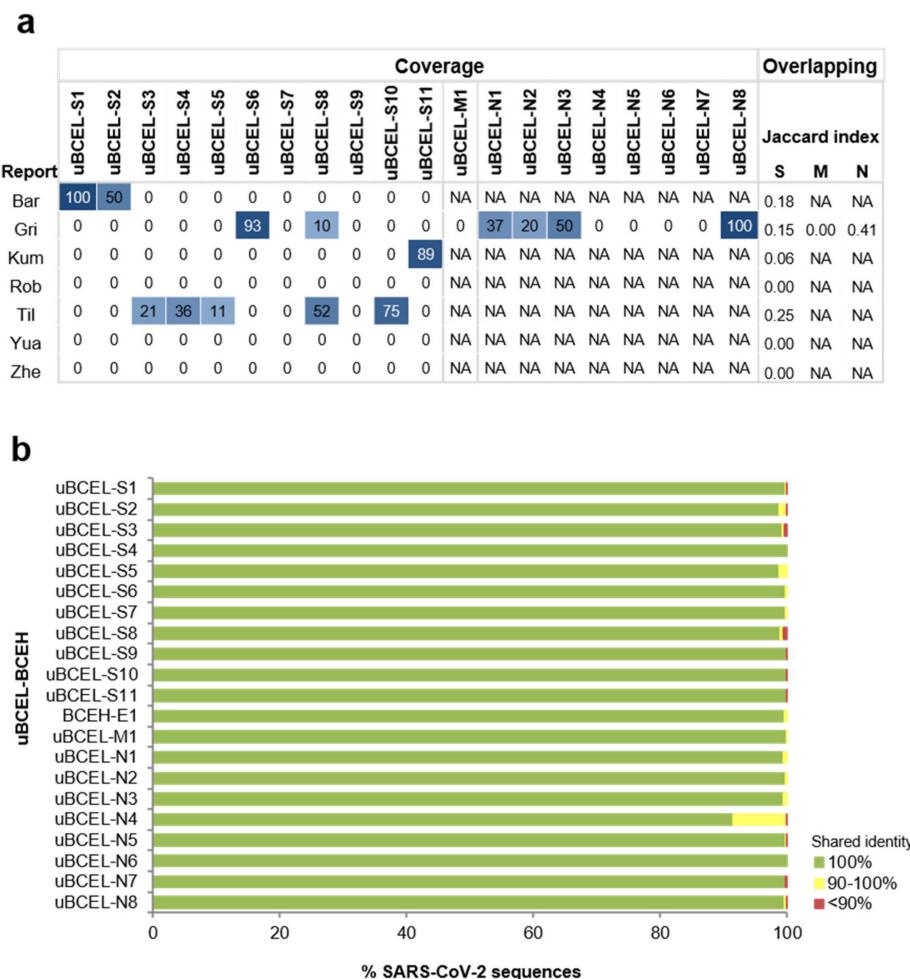


Figure 3. Characteristics of uBCEls. (a) heatmap of uBCEls coverage by protein and study. Coverage is also ranked by color intensity. Report prefixes: Bar [44], Gri [46], Kum [41], Rob [45], Til [42], Yua [40], and Zhe [43]. Not applicable (NA) labels are indicated for reports that did not propose epitopes for M or N proteins. Jaccard indices for residue overlapping are indicated on right; (b) conservation of uBCEls in SARS-CoV-2 sequenced genomes. Genomes were acquired applying the “COVID-19” organism in the NCBI nucleotide database (Last accession: 1 May 2020).

In contrast, when uBCEL sequence variability was inspected within an updated sampling (Last accession: 1 May 2020) of 1639 SARS-CoV-2 isolates with complete genome sequences, only four uBCEls demonstrated residue changes in more than 1% of isolates (Figure 3b and Table 2). Among these, the I68- deletion in uBCEL-S2, V483A in uBCEL-S5, and the S197L, R203K-G204R, and T205I in uBCEL-N4 variants were observed in 10 or more isolates. All these predominant lineages were detected for the first time in isolates obtained between 29 January 2020 and 15 March 2020. In particular, the R203K-G204R variant in N was particularly prevalent since it was detected in samples involving nine countries from three continents.

Table 2. Prevalent changes observed within uBCEls in SARS-CoV-2. Occurrence and geotemporal data is provided for residue variants found in ≥ 2 isolates.

uBCEL	Change(s)	n	Date of First Isolation	Geolocation
uBCEL-S2	I68-	11	15/03/2020	USA: WA
	N74K	2	20/01/2020	Brasil; China
	D80Y	2	31/03/2020	USA: WA
	G476S	7	10/03/2020	USA: WA
	V483A	11	05/03/2020	USA: WA
	Q677H	2	19/03/2020	USA: UT
	T791I	6	26/02/2020	Taiwan
	BCEH-E1	2	19/03/2020	USA: WA
	P67S	2	17/03/2020	USA: NY; USA: WA
	A152S	2	13/03/2020	USA: UT
uBCEL-N3	S180I	2	31/03/2020	USA: WA
	S183Y	4	17/03/2020	USA
	R185C	5	15/03/2020	USA
	R185L	2	19/03/2020	USA
	S188L	3	18/03/2020	USA
	S188P	2	13/03/2020	Taiwan
	S190I	3	17/03/2020	USA: NY
	S196L	6	29/02/2020	USA
	S197L	17	26/02/2020	Greece; Spain; USA
	S202N	7	30/01/2020	China; USA
uBCEL-N4	R203K,G204R	62	27/02/2020	Czech Republic; Greece; India; Israel; Peru; Spain; Sri Lanka; Taiwan; USA
	T205I	10	29/01/2020	China; USA
	A208G	4	16/03/2020	USA: WA; USA: NY
	P344S	2	?/01/2020	Japan
uBCEL-N7	E367-	2	16/03/2020	SA: UT; USA: WA

3.3. S Protein uBCEL Analysis

The S protein is by far the largest and most complex antigenic polypeptide in the SARS-CoV-2 virion. This protein has two subunits, three predicted domains according to the Pfam database that detect independent protein sections through hidden Markov models, a signal peptide, one TMH, at least one coiled-coil region, and numerous alpha-helices and beta-strands (Figure 4a). Four of the eleven uBCEls harbor N-linked-glycosylation sites, a potentially important factor used to evade the immune response [54]. In addition, three uBCEls (-S3, -S4 and -S5) fell in the S1B domain (323–502 residues) and include six of the fourteen residues that directly interact with the host ACE2 receptor. The uBCEL-S1, -S2, -S4, -S5, and -S7 are among the most divergent regions (showing 4–22 changes and up to seven inserted residues) between SARS-CoV and SARS-CoV-2 (Figure 4b,c). These mutational hotspots affect the apical half of the protein (Figure 4d).

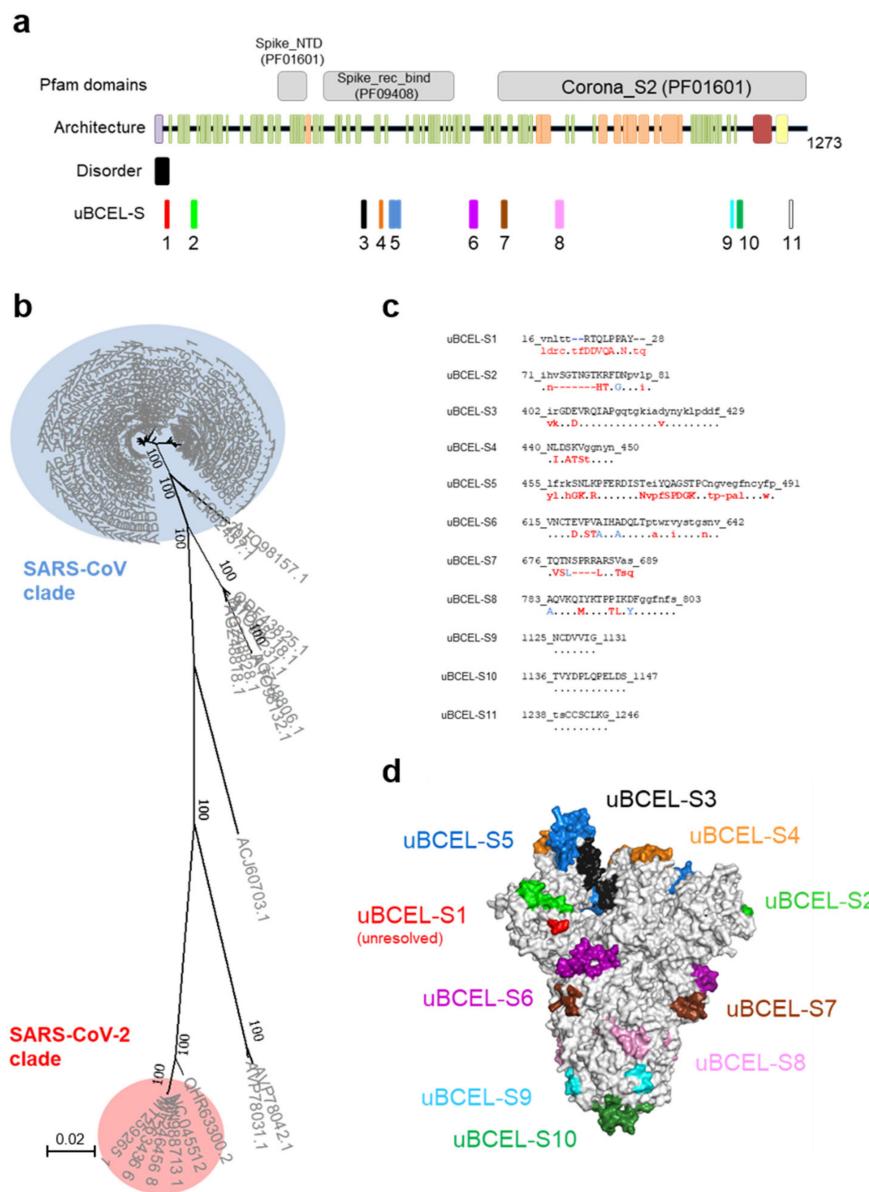


Figure 4. Architecture and analysis of predicted uBCEls on spike S protein. (a) a graphical depiction of S protein with Pfam domains, architecture, disorder and uBCEls localization. Color blocks: Pfam domains (grey), alpha-helices (orange), beta-strand (green), transmembrane helices (yellow), signal peptide (purple), coiled-coils (red) and disordered regions (black). In addition, each uBCEL is shown in its own color; (b) phylogenetic tree of 89 non-redundant coronavirus S sequences calculated by the Neighbor-Joining method. Bootstrap values of 100 are indicated; (c) SARS-CoV-2 uBCELS sequences (including deletions) and changes observed in relation to SARS-CoV S protein: capital letters indicate epitopes, residues conserved in $\geq 90\%$ sequences (dots), changed to unique option ($\geq 90\%$, red), ambiguous changes (two or more residue option in $> 10\%$ sequences, blue), and deletions (dashes); (d) structural mapping of uBCEls on the surface view of the modeled S protein homotrimer. Each uBCEL is depicted in the same colors as in Figure 4a. uBCEL-S11 falls out of the resolved section of the protein and therefore is not shown.

3.4. E Protein Epitope Analysis

The E protein is a small polypeptide with a predicted TMH and 38 residues oriented toward the exterior of the virion (Figure 5a). Nevertheless, this region is predicted to contain three alpha-helices that may interact to produce a minifold. The additional absence of disorder strongly suggests that

this section is completely folded. Within the 38-mer minifold, there is a 12-mer within the predicted alpha-helix 3 with high predicted antigenicity, BCEH-E1. This is the only structured exception included in our analyses. The SARS-CoV-2 minifold sequence only showed three changes and one deletion with respect to the SARS-CoV clade sequences, although all fall outside the epitope (Figure 5b,c).

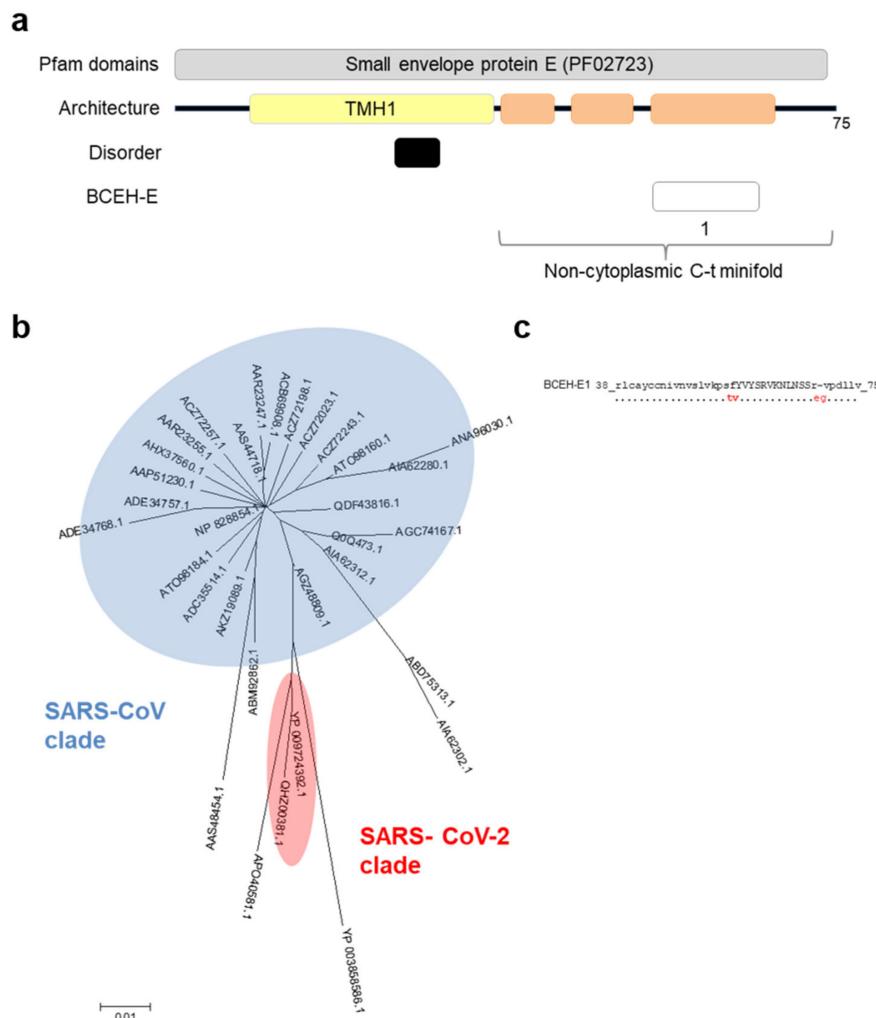


Figure 5. Architecture and analysis of BCEH of envelope E protein. (a) a graphical depiction of E protein with Pfam domains, architecture, disorder and BCEH localization. Color blocks: Pfam domains (grey), alpha-helices (orange), transmembrane helices (yellow), disordered regions (black), and BCEH (white); (b) phylogenetic tree of 32 non-redundant coronavirus E sequences calculated by the Neighbor-Joining method. No bootstrap value reached a value of 100; (c) SARS-CoV-2 BCEH-E sequence and changes observed in relation to SARS-CoV E protein: capital letters indicate epitopes, residues conserved $\geq 90\%$ sequences (dots), changed to unique option ($\geq 90\%$, red), deletions (dashes).

3.5. M Protein Epitope Analysis

The M protein is also a peripheral membrane protein, with 3 TMHs and a globular carboxyl-terminal section of 112 residues enriched in predicted beta-strands (Figure 6a). Although some loops in this region are evident, the protein is predicted to have low disorder. Only one uBCEL was detected, close to the C-terminus, with two conservative amino acid changes with respect to the SARS-CoV homologs (Figure 6b,c).

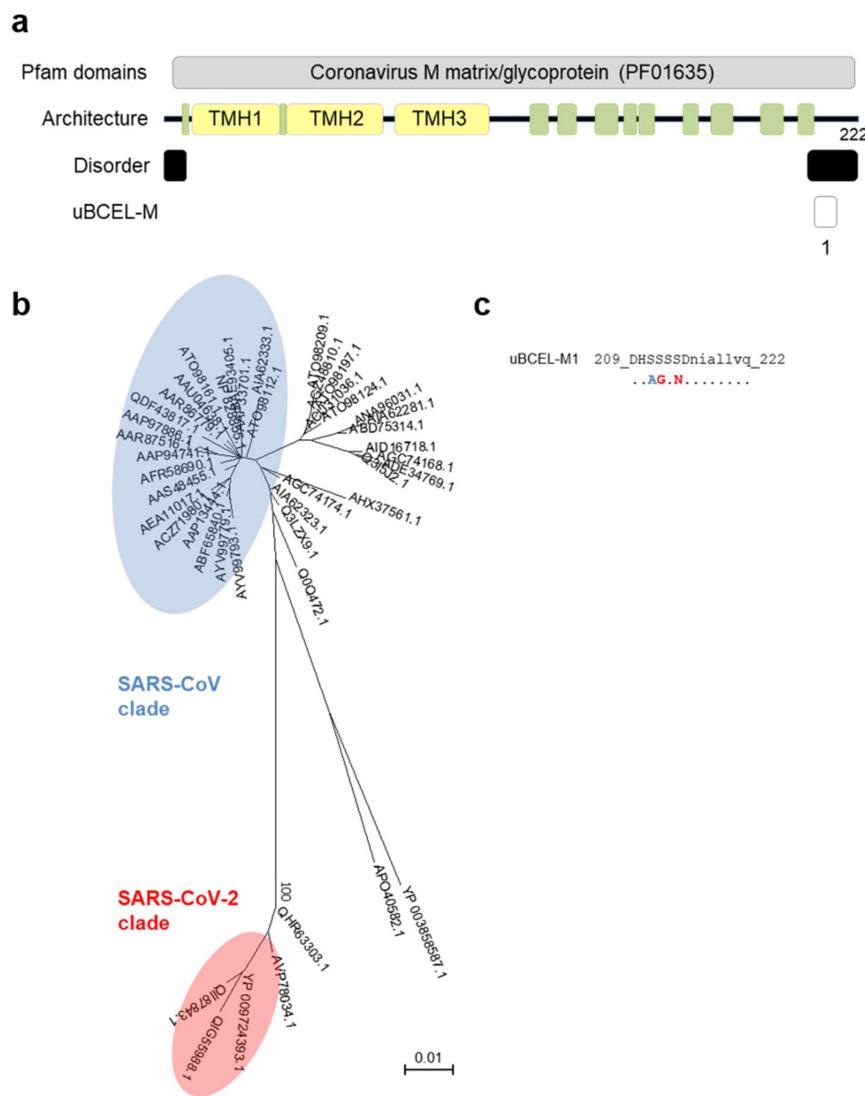


Figure 6. Architecture and analysis of uBCEls of matrix glycoprotein M. (a) a graphical depiction of M protein with Pfam domains, architecture, disorder, and uBCEls localization. Color blocks: beta-strand (green), transmembrane helices (yellow), disordered regions (black), and uBCEls (white); (b) phylogenetic tree of 44 non-redundant coronavirus M sequences calculated by the Neighbor-Joining method. Bootstrap values of 100 are indicated; (c) SARS-CoV-2 uBCEL-M sequence and changes observed in relation to SARS-CoV M protein: capital letters indicate epitopes, residues conserved $\geq 90\%$ sequences (dots), changed to unique option ($\geq 90\%$, red), ambiguous changes (two or more residue option in $> 10\%$ sequences, blue).

3.6. N Protein Epitope Analysis

In contrast to E and M proteins, predictions based on residue content as well as partial structural information indicate that the N protein is a remarkably disordered and antigenic polypeptide (Figure 7a). Three out of the eight uBCEls identified in this protein were ≥ 35 residues in length. Five uBCEls were highly conserved (≤ 1 change) compared to the SARS-CoV clade, whereas uBCEL-N1, -N4 and -N8 showed 3–5 unambiguous changes, most of them in the predicted epitope sequences (Figure 7b,c). The disordered nature of N is evident in uBCEL-N2 and -N3, which intermingled in the resolved RNA-binding domain (Figure 7d).

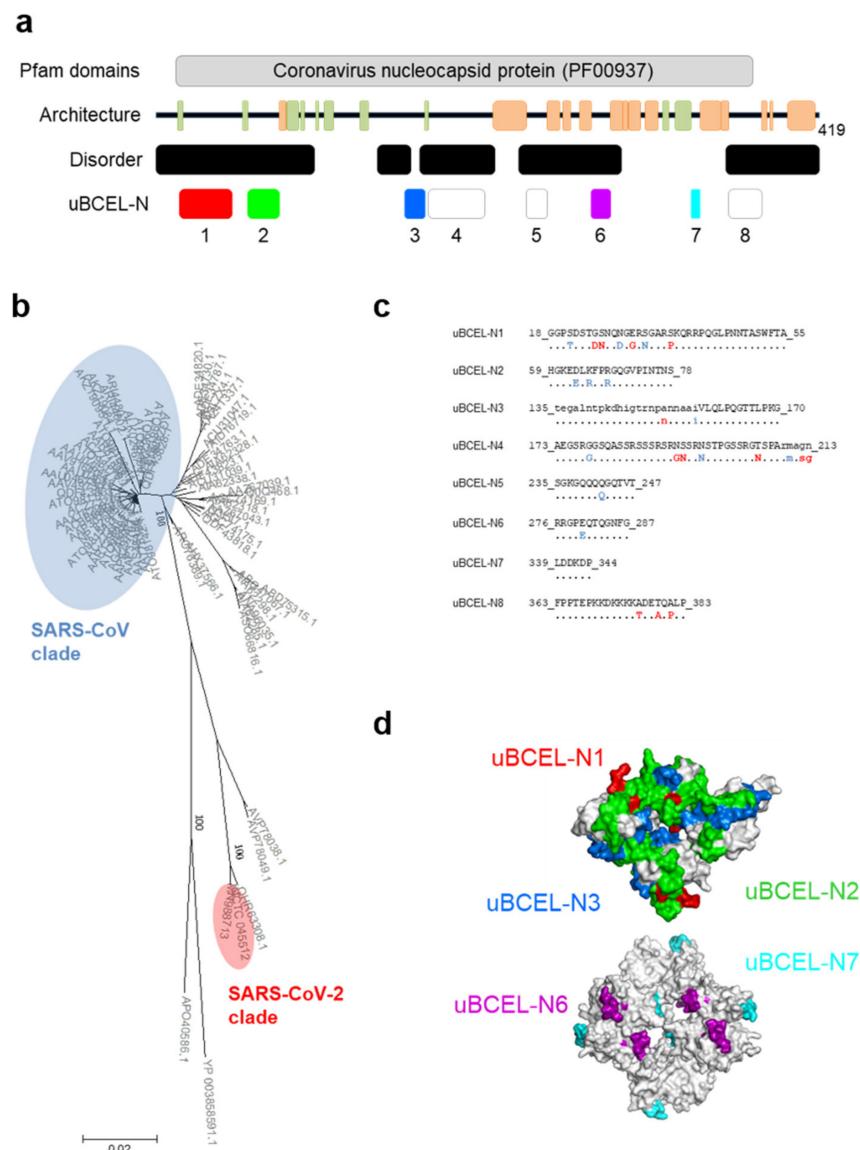


Figure 7. Architecture and analysis of uBCELS of nucleocapsid N protein. (a) a graphical depiction of N protein with Pfam domains, architecture, disorder and uBCELS localization. Color blocks: Pfam domains (grey), alpha-helices (orange), beta-strand (green). In addition, each uBCEL is shown in its own color; (b) phylogenetic tree of 71 non-redundant coronavirus N sequences calculated by the Neighbor-Joining method. Bootstrap values of 100 are indicated; (c) SARS-CoV-2 uBCEL-N sequence and changes observed in relation to SARS-CoV N protein: capital letters indicate epitopes, residues conserved $\geq 90\%$ sequences (dots), changed to unique option ($\geq 90\%$, red), ambiguous changes (two or more residue option in $> 10\%$ sequences, blue); (d) structural mapping of uBCELS on the surface view of resolved three-dimensional structures for the nucleotide-binding (above) and oligomerizing (below) domains. Each uBCEL is depicted in the same colors as in Figure 7a. The uBCEL-N4, uBCEL-N5 and uBCEL-N8 are located in regions of unsolved structure and therefore are not shown.

3.7. Assessment of the Agreement between uBCELS in SARS-CoV-2 and Linear B-Cell Epitopes Previously Reported for SARS-CoV

Given the recent nature of the SARS-CoV-2 pandemic, the body of antigenic knowledge regarding this coronavirus is still scarce. In contrast, SARS-CoV has been analyzed over almost two decades. To determine the degree of locational novelty in the epitopes proposed in this study, they were overlaid with SARS-CoV epitopes available in immunologic and bibliographic databases (Figure 8). A total

of 117 validated epitopes for SARS-CoV were identified in the four structural proteins: S ($n = 64$), E ($n = 2$), M ($n = 11$) and N ($n = 40$) (Supplementary Table S2). Twenty-four epitopes have been reported in the S1B subunit of SARS-CoV, which globally overlapped with uBCEL-S3, -S4 and -S5. Likewise, the remaining amino-terminal half of the S protein contained 39 previously validated SARS-CoV epitopes, which partially covered uBCEL-S1 and -S2, but not uBCEL-S7 from this study. In contrast, uBCEL-S1, -S2, -S8, and -S11 are found in regions with no overlap with epitopes previously described for SARS-CoV (Figure 8). The only two previously validated epitopes reported for E protein were also partially redundant with BCEH-E1. Five and six validated epitopes have been reported for the amino and carboxyl extremes of the M protein. The uBCEL-M1 presents redundancy with one of the later [33]. The N protein contained 40 previously validated epitopes, 20 in each half of the protein. The uBCEL-N1 and uBCEL-N8 are localized in highly redundant zones. Overall, the uBCELS predicted in this study show some degree of overlap with previously-validated SARS-CoV epitopes; however, redundancy is generally poor and partial, and there is no agreement with the structural criteria used in the present study.

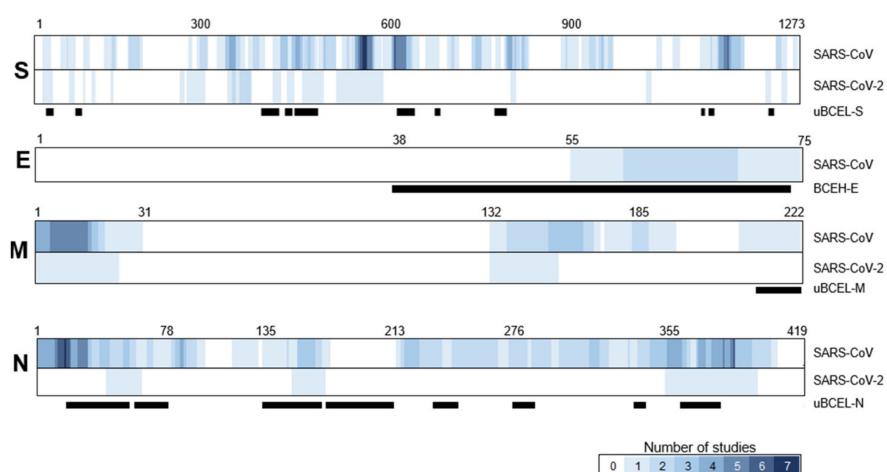


Figure 8. Location of previously validated linear B-cell epitopes in coronavirus virion proteins. Heatmaps indicate the bibliographic consensus of studies including this residue as part of a B-cell epitope in reported studies for SARS-CoV ($n = 43$) and SARS-CoV-2 ($n = 7$). Black bars indicate the position of uBCELS identified in this study. To the best of our knowledge, no B-cell epitopes have been reported for SARS-CoV-2 E protein. Key residues positions are indicated in each protein for identification.

3.8. Epitope Conservation in Bat Coronaviruses

Three coronavirus samples extracted from bats in China were recurrently found close to the SARS-CoV-2 clade in the phylogenetic trees generated for the four proteins. Bat-SL-CoVZXC21 (NCBI sample: MG772934.1, July 2015) and bat-SL-CoVZC45 (Sample: MG772933.1, February 2017) were isolated from *Rhinolophus sinicus* during the same study [55]. RaTG13, isolated from *Rhinolophus affinis* (Sample: MN996532.1, July 2013), is a well-known isolate that shares high global similarity to SARS-CoV-2 and is therefore considered a potential ancestor of the human lineage causing the current pandemic [50]. Thus, it is of interest to calculate the degree of conservation between the candidate uBCELS identified in this study and these sequences in order to assess the possibility of epitope switching in these bat coronaviruses. While the SARS-CoV reference sequences shared 76.9% identical residues and only six identical uBCELS compared to SARS-CoV-2, Bat-SL-CoVZXC21, and bat-SL-CoVZC45 reached 80% uBCEL residue identity and 11 identical uBCELS (Supplementary Data S1). Interestingly, RaTG13 showed 95.7% uBCEL residue identity and 14 uBCELS were a 100% match, with only a small divergence in the potential switched loops.

4. Discussion

In addition to discontinuous B-cell epitopes, viral linear epitopes located in protein loops have also proven useful in the biomedical field. In this study, we aimed to further exploit the linear epitope strategy by adding additional criteria to refine the selection pipeline, namely application of a consensus of epitope predictors, consideration of the antigen architecture, and the placement of each individual epitope into an evolutionary context.

A total of 20 uBCEPs were identified in the four SARS-CoV-2 virion proteins using this approach. In addition, a region containing a globular minifold in the E protein was also included in our approach given the prediction of a B-cell epitope in this section and its assumed folding completeness. The most divergent uBCEPs with respect to SARS-CoV were found in the S₁ subunit. The longest uBCEPs and the highest uBCEP protein coverage corresponded to the N protein, very likely because of its higher structural disorder.

The poor agreement observed between SARS-CoV-2 epitopes reported in the literature, and between previous studies and our uBCEP dataset, underscore the influence of the immunoinformatic strategy on the resulting epitope list. Differences in epitope identification selection criteria could radically affect the success of the downstream processes that employ the identified sequences.

Since the genomes of SARS-CoV and SARS-CoV-2 share 89% nucleotide identity [2], the wealth of information obtained for the former since 2003 may assist the development of antibody-based products for the current pandemic. Some recurrent, well-studied epitopes in SARS-CoV partially match with our uBCEP (see Table 1 and Supplementary Table S2). For example, in previous studies, convalescent sera with high neutralizing activity have been demonstrated to recognize epitopes with sequences related to uBCEP-S6/-S7, BCEH-E1, uBCEP-M1, and uBCEP-N1/-N3/-N8. In addition, uBCEP-S4 and -S5 are found in the S1B domain and include residues that directly interact with the ACE2 receptor. Several validated epitopes with similar sequences have been studied in SARS-CoV. Furthermore, the uBCEP-S9 and -S10 are equivalent to epitopes recognized by neutralizing antibodies that prevent viral entry into the cell [56]. Hence, antibodies against these uBCEPs may explicitly neutralize the SARS-CoV-2 attachment process. Finally, we have identified five novel epitopic regions within the S protein, whose inclusion in future assays offers a chance to explore novel antigenic features in the coronavirus.

Several prophylactic strategies, each with associated strengths and limitations, have been proposed to protect humans against viral infections [47]. These include the utilization of genetically-modified attenuated viruses, purified subunits (typically proteins), and nucleic acid molecules. Although limitations remain to be addressed, DNA-based platforms have enormous potential in development of vaccines for infectious diseases [57]. For instance, epitopic sections below the folding domain length range of approximately ≤ 40 residues, can be included in DNA vaccines. Importantly, we observed that BCEPs contain few and intermediate-affinity HLA-II epitopes (data not shown) and may therefore need to be combined with strong and promiscuous CD4+ T-cell epitopes that enhance their mild immunogenicity. Our approach intends to pave the way for exploiting this route by merging uBCEPs from the four virion proteins in tandem and circumventing the UPR in cells transfected with the DNA vaccine. The 437 residues in our epitope dataset permit a reduction of at least 4.5-fold of the virion proteome, thus facilitating the construction of chimeric proteins including only highly antigenic sequences. Antigen engineering involving the incorporation of protein loops into chimeric proteins to improve protection has already been applied in other viruses such as papillomavirus [58] and Zika [54].

The exploration of the linear epitope option in coronaviruses is additionally warranted due to the conformational re-arrangements observed in the S protein during infection [59], which may alter the stability of discontinuous epitopes, and the extraordinarily disordered nature of the N protein. The modular nature of our approach also permits the evaluation of synergistic effects between different epitopes in animal models. In addition, it may promote the identification of non-neutralizing immunodominant epitopes that divert the immune response towards enhancement of infectivity and

eosinophilia-related immunopathology by eliciting an unbalanced antiinflammatory cytokine and T1/T2 responses [60,61].

The strong sequence divergence found in five uBCELS of the S protein indicates that these are important drivers in coronavirus evolution and may play a role in the lack of immunoprotective compatibility observed between SARS-CoV and SARS-CoV-2 [62,63]. Epitope switching in loops leading to immunologically independent lineages has been observed in VIH and influenza virus [64,65]. We postulate that some of the uBCELS described here are also subjected to high selective pressure that promotes epitope switching, and that this aspect combines with glycan epitope masking and conformational changes [66] to hamper the memory response in natural animal reservoirs. Moreover, equivalent events may also take place in humans if the pandemic is prolonged over a long period of time. The incidence of residue changes in uBCELS observed in SARS-CoV-2 genomes sampled from humans to date is still low. However, the local and international dissemination of some lineages harboring up to two residue polymorphisms in uBCELS only five months after the original outbreak is of special concern. The progressive accumulation of such variants may result in immune evasion, i.e., serotypes, of the polyclonal antibody response. This would have serious consequences, such as reinfection of previously-immunized humans, vaccine escape, and false negatives in antibody-based rapid diagnostics. In this context, the existence of coronaviruses detected in bats showing intermediate epitopic features at different stages between SARS-CoV and SARS-CoV-2 is of great interest. Although virus sampling in these animals is infrequent, three bat-CoVs isolates embodied such middle links. In particular, equivalent uBCELS in RatG13—isolated as early as in July 2013—were surprisingly similar to those in SARS-CoV-2, indicating epitope migration between SARS-CoV and SARS-CoV-2 could have essentially occurred seven years ago, or earlier. Since bats are a natural reservoir for coronaviruses, epitope switch by mutation and recombination may be facilitated by the recursive attempts of coronaviruses to re-infect bats that have been previously immunized against former strains.

Our uBCEL list includes a landscape of constant and changing epitopes that are located both inside and outside of the conformational rearrangement zone, and include both glycosylated and non-glycosylated sequences in S protein. This gradient in conservation may help to elicit a balance between strong specific response against SARS-CoV-2 as well as cross-protection to future coronavirus variants. However, in the event that such cross-protection between present and as yet unidentified coronaviruses is ineffective, similar vaccines and specific antibodies for diagnostic tests and passive therapy may be promptly redesigned. Loop sequences with equivalent coordinates could be introduced into the same DNA framework immediately after the first sequenced genome of an emerging pathogenic coronavirus is available.

Although some redundancy has been observed between published epitopes and our uBCELS, our sequence combination is unique and the exact loop limits have been carefully delineated to avoid structural regions. We anticipate that the inclusion of partially structured epitopes in DNA molecules may compromise its success by inducing the UPR-related proteolysis and probably apoptosis. Such an outcome would negate the epitope integrity, presentation to naïve B-cells, and protective capacity. Our dataset follows a unified view at the service of strategies aimed at producing antibodies for different biomedical purposes. While the epitopes identified in this study will need to be experimentally tested, our characterized and rationalized catalog of epitopes is therefore of interest for vaccine developers and the general scientific community.

5. Conclusions

Vaccines against SARS-CoV-2 are needed to stopping deaths and the global economic drain caused by the current pandemic. Epitope-based nucleic acid vaccines, one of the most promising vaccine candidates, can trigger the unfolded protein response in host cells, thus reducing the amount of antigen available for immune stimulation and decreasing vaccine efficacy. To avoid this, all available methodologies for B-cell epitope prediction have been unified and a structural filter added in order to identify twenty loops in virion proteins enriched in B cell epitopes.

Mutational hotspots were observed in several of these regions with respect to the SARS-CoV 2003 pandemic virus. These are indicators of epitopic switch, which occurs to gradually evade the host immune response. This is strongly supported by the fact that some bat beta-coronavirus sequences are phylogenetically halfway between both human pandemics. Although unstructured epitopic zones are generally identical within SARS-CoV-2 human samples, up to two residue changes have been identified in some isolates. Exhaustive surveillance of epitopic mutations would therefore be highly recommended in order to predict future vaccine escape.

This novel epitope zone dataset is tailored to a range of prophylactic strategies that intend to elicit protective humoral responses, but could be hindered by protein misfolding. The results of this study may therefore have broad applications in complementing current initiatives aimed at developing immune-based therapeutics for SARS-CoV-2.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2076-393X/8/3/397/s1>, Supplementary Data S1: Conservation of B-cell epitopes in S, E, M, and N proteins across coronavirus species. Supplementary Table S1: References for databases and methods utilized in this study. Supplementary Table S2: Linear B-cell epitopes identified in spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins of SARS-CoV and SARS-CoV-2 that have been validated in vivo.

Author Contributions: Conceptualization, D.L., M.J.M. and A.J.M.-G.; methodology, A.C.-L., M.L.-S. and A.J.M.-G.; formal analysis, A.C.-L., M.L.-S. and A.J.M.-G.; data curation, A.C.-L.; writing, A.C.-L., M.L.-S., M.J.M. and A.J.M.-G.; writing—review and editing; D.L.; funding acquisition, D.L., M.J.M. and A.J.M.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Acción Estratégica en Salud from the ISCIII, grants MPY 380/18, MPY 388/18, and MPY 509/19. A.C.-L. is the recipient of a Comunidad de Madrid contract by the ISCIII. M.L.-S. is the recipient of a Sara Borrell contract by the ISCIII. A.J.M.-G. is the recipient of a Miguel Servet contract by the ISCIII.

Conflicts of Interest: M.J.M. is a founder and shareholder in the biotechnology company Vaxdyn, S.L. Vaxdyn played no role in the present study. No other competing interest is declared for the other co-authors.

Abbreviations

List of Non-Standard Abbreviations Utilized in This Work:

BCE	B-cell epitope
BCEH	B-cell epitope in an alpha-helix section
E	envelope protein
IEDB	immune epitope database
M	membrane protein
N	nucleocapsid protein
S	spike protein
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SDP	specificity determining position
TMH	transmembrane helix
uBCE	unstructured B-cell epitope
uBCEL	unstructured B-cell epitope-containing loop
UPR	unfolded protein response

References

- Wang, C.; Horby, P.W.; Hayden, F.G.; Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**, *395*, 470–473. [[CrossRef](#)]
- Chan, J.F.; Kok, K.H.; Zhu, Z.; Chu, H.; To, K.K.; Yuan, S.; Yuen, K.Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* **2020**, *9*, 221–236. [[CrossRef](#)] [[PubMed](#)]
- Cheng, Z.K.J.; Shan, J. Novel coronavirus: Where we are and what we know. *Infection* **2019**. [[CrossRef](#)]
- Tyrrell, D.A.; Bynoe, M.L. Cultivation of viruses from a high proportion of patients with colds. *Lancet* **1966**, *1*, 76–77. [[CrossRef](#)]

5. Cui, J.; Li, F.; Shi, Z.L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181–192. [[CrossRef](#)]
6. Peiris, J.S.; Guan, Y.; Yuen, K.Y. Severe acute respiratory syndrome. *Nat. Med.* **2004**, *10*, S88–S97. [[CrossRef](#)]
7. Assiri, A.; McGeer, A.; Perl, T.M.; Price, C.S.; Al Rabeeah, A.A.; Cummings, D.A.; Alabdullatif, Z.N.; Assad, M.; Almulhim, A.; Makhdoom, H.; et al. Hospital outbreak of Middle East respiratory syndrome coronavirus. *N. Engl. J. Med.* **2013**, *369*, 407–416. [[CrossRef](#)]
8. Chen, Y.; Liu, Q.Y.; Guo, D.Y. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* **2020**, *92*, 418–423. [[CrossRef](#)]
9. Walls, A.C.; Park, Y.J.; Tortorici, M.A.; Wall, A.; McGuire, A.T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181*, 281–292. [[CrossRef](#)]
10. He, Y.X.; Li, J.J.; Du, L.Y.; Yan, X.X.; Hu, G.G.; Zhou, Y.S.; Jiang, S.B. 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* **2006**, *24*, 5498–5508. [[CrossRef](#)]
11. Lien, S.P.; Shih, Y.P.; Chen, H.W.; Tsai, J.P.; Leng, C.H.; Lin, M.H.; Lin, L.H.; Liu, H.Y.; Chou, A.H.; Chang, Y.W.; et al. Identification of synthetic vaccine candidates against SARS CoV infection. *Biochem. Biophys. Res. Commun.* **2007**, *358*, 716–721. [[CrossRef](#)] [[PubMed](#)]
12. Hua, R.H.; Zhou, Y.J.; Wang, Y.F.; Hua, Y.Z.; Tong, G.Z. Identification of two antigenic epitopes on SARS-CoV spike protein. *Biochem. Biophys. Res. Commun.* **2004**, *319*, 929–935. [[CrossRef](#)]
13. Wang, X.H.; Xu, W.; Tong, D.Y.; Ni, J.; Gao, H.F.; Wang, Y.; Chu, Y.W.; Li, P.P.; Yang, X.M.; Xiong, S.D. A chimeric multi-epitope DNA vaccine elicited specific antibody response against severe acute respiratory syndrome-associated coronavirus which attenuated the virulence of SARS-CoV in vitro. *Immunol. Lett.* **2008**, *119*, 71–77. [[CrossRef](#)] [[PubMed](#)]
14. Yu, X.F.; Liang, L.H.; She, M.; Liao, X.L.; Gu, J.; Li, Y.H.; Han, Z.C. Production of a monoclonal antibody against SARS-CoV spike protein with single intrasplenic immunization of plasmid DNA. *Immunol. Lett.* **2005**, *100*, 177–181. [[CrossRef](#)] [[PubMed](#)]
15. Schoeman, D.; Fielding, B.C. Coronavirus envelope protein: Current knowledge. *Virol. J.* **2019**, *16*, 019–1182. [[CrossRef](#)]
16. Guo, J.P.; Petric, M.; Campbell, W.; McGeer, P.L. SARS corona virus peptides recognized by antibodies in the sera of convalescent cases. *Virology* **2004**, *324*, 251–256. [[CrossRef](#)] [[PubMed](#)]
17. Liu, J.; Sun, Y.; Qi, J.; Chu, F.; Wu, H.; Gao, F.; Li, T.; Yan, J.; Gao, G.F. The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *J. Infect. Dis.* **2010**, *202*, 1171–1180. [[CrossRef](#)]
18. Chow, S.C.S.; Ho, C.Y.S.; Tam, T.T.Y.; Wu, C.; Cheung, T.; Chan, P.K.S.; Ng, M.H.L.; Hui, P.K.; Ng, H.K.; Au, D.M.Y.; et al. Specific epitopes of the structural and hypothetical proteins elicit variable humoral responses in SARS patients. *J. Clin. Pathol.* **2006**, *59*, 468–476. [[CrossRef](#)]
19. Qian, C.; Qin, D.; Tang, Q.; Zeng, Y.; Tang, G.X.; Lu, C. Identification of a B-cell antigenic epitope at the N-terminus of SARS-CoV M protein and characterization of monoclonal antibody against the protein. *Virus Genes* **2006**, *33*, 147–156. [[CrossRef](#)]
20. He, Y.X.; Zhou, Y.S.; Siddiqui, P.; Niu, J.K.; Jiang, S.B. Identification of immunodominant epitopes on the membrane protein of the severe acute respiratory syndrome-associated coronavirus. *J. Clin. Microbiol.* **2005**, *43*, 3718–3726. [[CrossRef](#)]
21. Kannan, S.; Ali, P.S.S.; Sheeza, A.; Hemalatha, K. COVID-19 (Novel Coronavirus 2019)—Recent trends. *Eur. Rev. Med. Pharmacol. Sci.* **2020**, *24*, 2006–2011. [[PubMed](#)]
22. Cheung, Y.K.; Cheng, S.C.; Sin, F.W.; Chan, K.T.; Xie, Y. Induction of T-cell response by a DNA vaccine encoding a novel HLA-A*0201 severe acute respiratory syndrome coronavirus epitope. *Vaccine* **2007**, *25*, 6070–6077. [[CrossRef](#)] [[PubMed](#)]
23. Bussmann, B.M.; Reiche, S.; Jacob, L.H.; Braun, J.M.; Jassoy, C. Antigenic and cellular localisation analysis of the severe acute respiratory syndrome coronavirus nucleocapsid protein using monoclonal antibodies. *Virus Res.* **2006**, *122*, 119–126. [[CrossRef](#)]
24. Shin, G.C.; Chung, Y.S.; Kim, I.S.; Cho, H.W.; Kang, C. Preparation and characterization of a novel monoclonal antibody specific to severe acute respiratory syndrome-coronavirus nucleocapsid protein. *Virus Res.* **2006**, *122*, 109–118. [[CrossRef](#)] [[PubMed](#)]

25. Che, X.Y.; Hao, W.; Wang, Y.; Di, B.; Yin, K.; Xu, Y.C.; Feng, C.S.; Wan, Z.Y.; Cheng, V.C.; Yuen, K.Y. Nucleocapsid protein as early diagnostic marker for SARS. *Emerg. Infect. Dis.* **2004**, *10*, 1947–1949. [CrossRef] [PubMed]
26. Sun, P.; Guo, S.; Sun, J.; Tan, L.; Lu, C.; Ma, Z. Advances in In-silico B-cell Epitope Prediction. *Curr. Top. Med. Chem.* **2019**, *19*, 105–115. [CrossRef] [PubMed]
27. Sher, G.; Zhi, D.; Zhang, S. DRREP: Deep ridge regressed epitope predictor. *BMC Genom.* **2017**, *18*, 55–65. [CrossRef]
28. Van Regenmortel, M.H.V. Mapping Epitope Structure and Activity: From One-Dimensional Prediction to Four-Dimensional Description of Antigenic Specificity. *Methods* **1996**, *9*, 465–472. [CrossRef]
29. Carpentier, G.S.; Fleury, M.J.J.; Touzé, A.; Sadeyen, J.R.; Tourne, S.; Sizaret, P.Y.; Coursaget, P. Mutations on the FG Surface Loop of Human Papillomavirus Type 16 Major Capsid Protein Affect Recognition by Both Type-Specific Neutralizing Antibodies and Cross-Reactive Antibodies. *J. Med. Virol.* **2005**, *77*, 558–565. [CrossRef]
30. Qu, P.; Zhang, C.; Li, M.; Ma, W.; Xiong, P.; Liu, Q.; Zou, G.; Lavillette, D.; Yin, F.; Jin, X.; et al. A New Class of Broadly Neutralizing Antibodies That Target the Glycan Loop of Zika Virus Envelope Protein. *Cell Discov.* **2020**, *6*, 5. [CrossRef]
31. Xu, L.; Zheng, Q.; Li, S.; He, M.; Wu, Y.; Li, Y.; Zhu, R.; Yu, H.; Hong, Q.; Jiang, J.; et al. Atomic Structures of Coxsackievirus A6 and Its Complex With a Neutralizing Antibody. *Nat. Commun.* **2017**, *8*, 505. [CrossRef] [PubMed]
32. Potocnakova, L.; Bhide, M.; Pulzova, L.B. An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *J. Immunol. Res.* **2016**, *2016*, 6760830. [CrossRef] [PubMed]
33. Yu, H.; Jiang, L.F.; Fang, D.Y.; Yan, H.J.; Zhou, J.J.; Zhou, J.M.; Liang, Y.; Gao, Y.; Zhao, W.; Long, B.G. Selection of SARS-CoV-specific B cell epitopes by phage peptide library screening and evaluation of the immunological effect of epitope-based peptides on mice. *Virology* **2007**, *359*, 264–274. [CrossRef] [PubMed]
34. He, Y.X.; Li, J.J.; Heck, S.; Lustigman, S.; Jiang, S.B. Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: Implication for vaccine design. *J. Virol.* **2006**, *80*, 5757–5767. [CrossRef]
35. Hu, H.B.; Li, L.; Kao, R.Y.; Kou, B.B.; Wang, Z.G.; Zhang, L.; Zhang, H.Y.; Hao, Z.Y.; Tsui, W.H.; Ni, A.P.; et al. Screening and identification of linear B-cell epitopes and entry-blocking peptide of severe acute respiratory syndrome (SARS)-associated coronavirus using synthetic overlapping peptide library. *J. Comb. Chem.* **2005**, *7*, 648–656. [CrossRef]
36. Lu, W.; Wu, X.D.; De Shi, M.; Yang, R.F.; He, Y.Y.; Bian, C.; Shi, T.L.; Yang, S.; Zhu, X.L.; Jiang, W.H.; et al. Synthetic peptides derived from SARS coronavirus S protein with diagnostic and therapeutic potential. *FEBS Lett.* **2005**, *579*, 2130–2136. [CrossRef]
37. Rubinchik, E.; Chow, A.W. Recombinant expression and neutralizing activity of an MHC class II binding epitope of toxic shock syndrome toxin-1. *Vaccine* **2000**, *18*, 2312–2320. [CrossRef]
38. Qin, E.; Zhu, Q.; Yu, M.; Fan, B.; Chang, G.; Si, B.; Yang, B.; Peng, W.; Jiang, T.; Liu, B.; et al. A complete sequence and comparative analysis of a SARS-associated virus (Isolate BJ01). *Chin. Sci. Bull.* **2003**, *48*, 941–948. [CrossRef]
39. Vita, R.; Mahajan, S.; Overton, J.A.; Dhandha, S.K.; Martini, S.; Cantrell, J.R.; Wheeler, D.K.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **2019**, *47*, D339–D343. [CrossRef]
40. Yuan, M.; Wu, N.C.; Zhu, X.; Lee, C.D.; So, R.T.Y.; Lv, H.; Mok, C.K.P.; Wilson, I.A. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science* **2020**, *368*, 630–633. [CrossRef]
41. Kumar, S.; Maurya, V.K.; Prasad, A.K.; Bhatt, M.L.B.; Saxena, S.K. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). *Virusdisease* **2020**, *31*, 13–21. [CrossRef] [PubMed]
42. Tilocca, B.; Soggiu, A.; Musella, V.; Britti, D.; Sanguinetti, M.; Urbani, A.; Roncada, P. Molecular basis of COVID-19 relationships in different species: A one health perspective. *Microbes Infect.* **2020**, *22*, 218–220. [CrossRef]
43. Zheng, M.; Song, L. Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. *Cell. Mol. Immunol.* **2020**, *17*, 536–538. [CrossRef] [PubMed]

44. Baruah, V.; Bose, S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J. Med. Virol.* **2020**, *92*, 495–500. [CrossRef] [PubMed]
45. Robson, B. Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. *Comput. Biol. Med.* **2020**, *119*, 26. [CrossRef]
46. Grifoni, A.; Sidney, J.; Zhang, Y.; Scheuermann, R.H.; Peters, B.; Sette, A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **2020**, *12*, 30166–30169. [CrossRef]
47. Shang, W.; Yang, Y.; Rao, Y.; Rao, X. The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines. *Npj Vaccines* **2020**, *5*, 1–3. [CrossRef]
48. Ibrahim, I.M.; Abdelmalek, D.H.; Elfiky, A.A. GRP78: A cell’s response to stress. *Life Sci.* **2019**, *226*, 156–163. [CrossRef]
49. Marra, M.A.; Jones, S.J.; Astell, C.R.; Holt, R.A.; Brooks-Wilson, A.; Butterfield, Y.S.; Khattra, J.; Asano, J.K.; Barber, S.A.; Chan, S.Y.; et al. The Genome sequence of the SARS-associated coronavirus. *Science* **2003**, *300*, 1399–1404. [CrossRef]
50. Zhou, P.; Yang, X.L.; Wang, X.G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.R.; Zhu, Y.; Li, B.; Huang, C.L.; et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579*, 270–273. [CrossRef]
51. Chen, C.Y.; Chang, C.K.; Chang, Y.W.; Sue, S.C.; Bai, H.I.; Riangu, L.; Hsiao, C.D.; Huang, T.H. Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* **2007**, *368*, 1075–1086. [CrossRef] [PubMed]
52. Buus, S.; Rockberg, J.; Forsstrom, B.; Nilsson, P.; Uhlen, M.; Schafer-Nielsen, C. High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Mol. Cell. Proteom.* **2012**, *11*, 1790–1800. [CrossRef] [PubMed]
53. Kalinina, O.V.; Mironov, A.A.; Gelfand, M.S.; Rakhmaninova, A.B. Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. *Protein Sci.* **2004**, *13*, 443–456. [CrossRef]
54. Goo, L.; DeMaso, C.R.; Pelc, R.S.; Ledgerwood, J.E.; Graham, B.S.; Kuhn, R.J.; Pierson, T.C. The Zika virus envelope protein glycan loop regulates virion antigenicity. *Virology* **2018**, *515*, 191–202. [CrossRef] [PubMed]
55. Hu, D.; Zhu, C.; Ai, L.; He, T.; Wang, Y.; Ye, F.; Yang, L.; Ding, C.; Zhu, X.; Lv, R.; et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerg. Microbes Infect.* **2018**, *7*, 154. [CrossRef]
56. Ng, O.W.; Keng, C.T.; Leung, C.S.W.; Peiris, J.S.M.; Poon, L.L.M.; Tan, Y.J. Substitution at Aspartic Acid 1128 in the SARS Coronavirus Spike Glycoprotein Mediates Escape from a S2 Domain-Targeting Neutralizing Monoclonal Antibody. *PLoS ONE* **2014**, *9*. [CrossRef]
57. Rauch, S.; Jasny, E.; Schmidt, K.E.; Petsch, B. New Vaccine Technologies to Combat Outbreak Situations. *Front. Immunol.* **2018**, *9*, 1963. [CrossRef]
58. Li, Z.; Song, S.; He, M.; Wang, D.; Shi, J.; Liu, X.; Li, Y.; Chi, X.; Wei, S.; Yang, Y.; et al. Rational design of a triple-type human papillomavirus vaccine by compromising viral-type specificity. *Nat. Commun.* **2018**, *9*, 018–07199. [CrossRef]
59. Walls, A.C.; Tortorici, M.A.; Snijder, J.; Xiong, X.; Bosch, B.J.; Rey, F.A.; Veesler, D. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11157–11162. [CrossRef]
60. Du, L.; Tai, W.; Yang, Y.; Zhao, G.; Zhu, Q.; Sun, S.; Liu, C.; Tao, X.; Tseng, C.K.; Perlman, S.; et al. Introduction of neutralizing immunogenicity index to the rational design of MERS coronavirus subunit vaccines. *Nat. Commun.* **2016**, *7*, 13473. [CrossRef]
61. Enjuanes, L.; Zuniga, S.; Castano-Rodriguez, C.; Gutierrez-Alvarez, J.; Canton, J.; Sola, I. Molecular Basis of Coronavirus Virulence and Vaccine Development. *Adv. Virus Res.* **2016**, *96*, 245–286. [PubMed]
62. Wrapp, D.; Wang, N.S.; Corbett, K.S.; Goldsmith, J.A.; Hsieh, C.L.; Abiona, O.; Graham, B.S.; McLellan, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263. [CrossRef] [PubMed]

63. Ahmed, S.F.; Quadeer, A.A.; McKay, M.R. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* **2020**, *12*, 254. [[CrossRef](#)] [[PubMed](#)]
64. Patil, S.; Kumar, R.; Deshpande, S.; Samal, S.; Shrivastava, T.; Boliar, S.; Bansal, M.; Chaudhary, N.K.; Srikrishnan, A.K.; Murugavel, K.G.; et al. Conformational Epitope-Specific Broadly Neutralizing Plasma Antibodies Obtained from an HIV-1 Clade C-Infected Elite Neutralizer Mediate Autologous Virus Escape through Mutations in the V1 Loop. *J. Virol.* **2016**, *90*, 3446–3457. [[CrossRef](#)] [[PubMed](#)]
65. Plant, E.P.; Manukyan, H.; Sanchez, J.L.; Laassri, M.; Ye, Z. Immune Pressure on Polymorphous Influenza B Populations Results in Diverse Hemagglutinin Escape Mutants and Lineage Switching. *Vaccines* **2020**, *8*, 125. [[CrossRef](#)]
66. Walls, A.C.; Tortorici, M.A.; Frenz, B.; Snijder, J.; Li, W.; Rey, F.A.; DiMaio, F.; Bosch, B.J.; Veesler, D. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **2016**, *23*, 899–905. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).