An Immunoinformatics Approach for SARS-CoV-2 in Latam Populations and Multi-Epitope Vaccine Candidate Directed towards the World's Population

Andrés Felipe Cuspoca [1], Laura Lorena Díaz [1], Alvaro Fernando Acosta [1], Marcela Katherine Peñaloza [1], Yardany Rafael Méndez [1], Diana Carolina Clavijo [2] and Juvenal Yosa Reyes [3,*]

1   Grupo de Investigación en Epidemiología Clínica de Colombia (GRECO),
    Universidad Pedagógica y Tecnológica de Colombia, Tunja 150003, Colombia
2   Facultad de Ingeniería y Ciencias, Pontificia Universidad Javeriana Cali, Santiago de Cali 760031, Colombia
3   Laboratorio de Simulación Molecular, Facultad de Ciencias Básicas y Biomédicas, Universidad Simón Bolívar,
    Barranquilla 080002, Colombia
*   Correspondence: juvenal.yosa@unisimonbolivar.edu.co

Abstract: The coronavirus pandemic is a major public health crisis affecting global health systems with dire socioeconomic consequences, especially in vulnerable regions such as Latin America (LATAM). There is an urgent need for a vaccine to help control contagion, reduce mortality and alleviate social costs. In this study, we propose a rational multi-epitope candidate vaccine against SARS-CoV-2. Using bioinformatics, we constructed a library of potential vaccine peptides, based on the affinity of the most common major human histocompatibility complex (HLA) I and II molecules in the LATAM population to predict immunological complexes among antigenic, non-toxic and non-allergenic peptides extracted from the conserved regions of 92 proteomes. Although HLA-C, had the greatest antigenic peptide capacity from SARS-CoV-2, HLA-B and HLA-A, could be more relevant based on COVID-19 risk of infection in LATAM countries. We also used three-dimensional structures of SARS-CoV-2 proteins to identify potential regions for antibody production. The best HLA-I and II predictions (with increased coverage in common alleles and regions evoking B lymphocyte responses) were grouped into an optimized final multi-epitope construct containing the adjuvants Beta defensin-3, TpD, and PADRE, which are recognized for invoking a safe and specific immune response. Finally, we used Molecular Dynamics to identify the multi-epitope construct which may be a stable target for TLR-4/MD-2. This would prove to be safe and provide the physicochemical requirements for conducting experimental tests around the world.

Keywords: SARS-CoV-2; vaccine; LATAM; in silico

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a pathogen that emerged towards the end of 2019, primarily affects the respiratory tract. It is transmitted from person to person via respiratory droplets, aerosols containing viral particles, and direct contact of the mucosa with contaminated surfaces [1,2].

The first infected patient was identified in Wuhan, Hubei Province, China. The origin of the virus is thought to be the Wuhan seafood market, although some cases had no connection to this location. The virus spread rapidly through Wuhan and shortly thereafter to the rest of China's provinces [1,3]. By 20 February 2020, 19 countries had reported cases and mortalities caused by coronavirus disease 2019 (COVID-19). In March of 2020, the World Health Organization (WHO) declared SARS-CoV-2 the etiologic agent of the first pandemic caused by a coronavirus [3,4].

In Latin America (LATAM), COVID-19 was first reported in Sao Paulo, Brazil, on February 25, 2020 with the case of a 61-year-old male who has traveled to Italy [5]. Subsequently,

cases were reported in other LATAM countries, including Chile [6] and Colombia [7]. Haiti was the last LATAM country to report the arrival of COVID-19 on 19 March 2020. The arrival of COVID-19 in LATAM presents a great challenge to a healthcare infrastructure that is susceptible to problems, such as the lack of SARS-CoV-2 testing, personal protective equipment, and intensive care unit beds. Additionally, mass migration facilitates contact and spread across borders [8]. As the pandemic progressed, a lack of effective control measures accelerated infection and death rates in comparison to other European countries [9,10]. We believe that research institutions in LATAM should come together in the development of a universal vaccine for four fundamental reasons: (1) SARS-CoV-2 has caused high morbidity and mortality worldwide; (2) it is highly contagious; (3) countries, including those in LATAM, were not prepared for the pandemic, and (4) genomic variation could occur during the pandemic, and changes to antigenic sites in vaccine formulations may be required [5,11–13]. By the end of the first week of November 2020, about 60 million cases were seen worldwide, with 1.2 million deaths globally [14].

COVID-19 has a broad range of clinical manifestations in patients ranging from asymptomatic to acute respiratory distress syndrome that may lead to death. A greater risk of mortality in patients occurs in unison with comorbidities, such as arterial hypertension, chronic obstructive pulmonary disease, diabetes, and vascular diseases, especially cerebrovascular disease [4,15].

Studies of similar pathologies, such as severe acute respiratory syndrome (SARS), have shown that the independent expression of type 1 interferon (IFN-1)-stimulated genes and Toll-like receptor (TLR) 3 and 4 are associated with better outcomes in infected rats [16]. Also, individuals with a homozygous expression of the polymorphic variants of L-SIGN are known to have a better viral binding capacity, increased viral degradation, and diminished cell to cell infection [17]. The expression of the HLA-C*15:02 and HLA-DR*03:01 alleles has also been associated with viral clearance. These alleles facilitate viral antigen presentation and, consequently, the elimination of SARS-CoV mediated by T lymphocytes (TLs) CD8+/CD4+ and natural killer cells [18].

From these findings, we can infer that the preservation of IFN-1 production and some alleles of HLA-I and HLA-II may be related to asymptomatic conditions and mild symptoms in COVID-19 patients [18]. In contrast, changes to IFN-1-producing signal pathways (such as polymorphisms or mutations) that compromise a patient's innate immunity, and differential expression of sex-dependent angiotensin-converting enzyme 2 (ACE2) receptor, may be associated with non-modifiable risk factors [19].

1.1. Virology of SARS-CoV-2

SARS-CoV-2 belongs to the Coronaviridae family and is part of the β group of coronaviruses. It is a 29.9 kb, positive-sense, single-stranded, enveloped RNA virus. It is similar to the etiologic agents that cause SARS and Middle East Respiratory Syndrome (MERS), which share 79.5% and 50% of their identity with SARS-CoV-2, respectively [3,20]. Coronavirus genomes are composed of 6–11 open reading frames (ORFs) [21], with the first ORF (ORF1a/b) containing two-thirds of the viral RNA. This ORF translates the pp1a and pp1ab polyproteins and encodes 16 non-structural proteins (NSPs). The other ORFs encode structural and accessory proteins. The genome contains accessory genes: two between the spike surface glycoprotein (SP) and small envelope protein genes (ORF3a and 3b); four between the matrix protein (MG) and nucleocapsid protein (NP) genes (6, 7a, 7b, 8), 9a, and 9b in the NP gene; and 10 after the NP gene [3,22,23]. The genome structure is shown in (Figure S1).

1.2. Medication for COVID-19

Currently, only Dexamethasone and Remdesivir have been shown to modify the natural course of the disease. The "RECOVERY" randomized clinical trial (RCT) was carried out in England. In this study, oral or intravenous Dexamethasone 6 mg/day was associated with a reduction in the mortality of patients requiring assisted ventilation and

patients who were in the second week of clinical diagnosis [24]. Another study, "ACTT1", involved 10 countries in a multicenter RCT. The trial showed that the nucleotide analogue prodrug Remdesivir, which used a loading dose of 200 mg and a maintenance dose of 100 mg/day, significantly reduced the time it took for patients to recover from COVID-19. Additionally, the need for supplemental oxygen after 10 days of intravenous treatment diminished [25]. However, the results of the WHO multicenter RCT "SOLIDARITY" study on the evaluation of the effect of "off label" drugs, which included the follow-up of about 11,330 adults from 4 continents, 30 countries and 405 hospitals, did not find Remdesivir to be associated with substantial prevention of in-hospital mortality, benefiting only a small fraction of patients when using a loading dose and standard maintenance [26]. Currently, the Food and Drug Administration has approved the use of Remdesivir and Dexamethasone for COVID-19, although the latter is only a temporary approval [25,27]. Other therapies have not shown improvements in mortality or recovery in patients with COVID-19. In an RCT carried out in China, patients with severe or life-threatening COVID-19 symptoms were studied. The findings showed that convalescent plasma did not result in clinical improvement after 28 days of follow-up, demonstrating no superiority to standard medical management [28]. In India, these findings were subsequently confirmed in a national multicenter RCT titled "PLACID" [29]. This is consistent with the "TSUNAMI" RCT, recently published from the Italian population with COVID-19, where it was not evidenced that a high antibody titer present in convalescent plasma will result in a reduction in mortality at 30 days, nor less need for invasive mechanical ventilation, indicating only a marginal effect in the absence of acute respiratory distress [30]. Other "off label" drugs are being evaluated worldwide to assess their usefulness against COVID-19. Tocilizumab, an IL-6 receptor antagonist, is being tested in an RCT (EudraCT: No. 2020-001408-41) in Germany to evaluate its efficacy and safety in patients with severe COVID-19 pneumonia. In the phase III RCT results, the use of Canakinumab, an IL-1β inhibitor, evidence to reduce the days of the hospitalization and mortality, in patients who do not require invasive ventilatory assistance, but do require supplemental O2 [31]. On the other hand, the use of Sarilumab, an IL-6 receptor antagonist, in patients with severity criteria, no benefit was observed compared to placebo [32]. Another phase II RCTs involving Thalidomide (NCT04273529), and monoclonal antibodies such as Anakinra (NCT04603742), Gimsilumab (NCT04351243), and Ruxolitinib are ongoing (NCT04359290).

The authors hypothesize that it could slow the progression of pneumonia and inflammation induced by SARS-CoV-2 [33]. On the other hand, promising results have been observed using mesenchymal cells derived from human umbilical cords [34] in patients with moderate and severe COVID in China. This has resulted in complete resolution of pneumonia within two weeks of infection without mortality and serious adverse reactions. It is currently in a non-randomized, controlled trial (NCT04288102).

1.3. Vaccine for SARS-CoV-2

A vaccine will be the most cost-effective strategy for preventing infection and reducing COVID-19-related morbidity and mortality [35]. According to a WHO report, vaccine studies for COVID-19 are using different vaccine strategies, such as non-replicating viral vectors, RNA, inactivated, viral-like particles, protein Subunit, peptides, and DNA approaches. To date, there are more than 100 vaccine candidates in preclinical trials [36]. Vaccine strategies and their phases are summarized in (Figure S2). In both SARS-CoV and SARS-CoV-2, several candidates structural proteins have been studied as vaccine targets. SP has been widely studied due to their capacity to induce neutralizing antibodies (NAb) that prevent the virus from binding and fusing with ACE2 receptor, as well as activating TLs. Various vaccine models have been proposed using the complete protein, SP binding domain, virus-like particles, DNA, or viral vectors [37,38]. NP was found to be expressed in large amounts during infection. NP is highly immunogenic and is a potential vaccine target. It develops a memory response from TLs that remain up to 11 years after SARS-CoV infection, detecting at the same time TLs of a specific memory, in addition to NP towards

other structural proteins such as SP and MG [39]. NP has recently been identified as an important virulence trait in the clinical outcome of COVID-19. In patients with moderate and severe COVID-19, sC5b-9, C4d, and NP are associated with respiratory failure (PAFI < 40 kPa) and the need for oxygen therapy. In murine models, this could be explained because NP is capable of dimerizing, activating, and cleaving MASP2 [40], a serine protease that can induce complement activation [41]. NP, MBL, MASP-2, C4 alpha, and C3 or C5b-9 have been observed in the lung tissue of patients with fatal COVID-19, which could indicate new therapeutic targets [40]. Previously, in a murine model infected by SARS-CoV and their principal structural proteins, NP alone resulted in a dysregulated inflammatory response in addition to pulmonary infiltrate of neutrophils, eosinophils, and lymphocytes. This led to a thickening of the pulmonary epithelium and severe pneumonia when using NP or SARS-CoV but not SP [42]. Therefore, the full-length protein NP as a vaccine target for SARS-CoV-2 could be considered less safe when compared to SP, since they share a homology greater than 90% [43]. Consequently, if it is used as a vaccine target, shorter length approaches should be chosen, since immunogenic peptides of NP have been found capable of evoking an immune response and inhibiting viral replication in influenza A and SARS-COV viruses [43–47].

Other experimental studies about structural proteins on SARS-CoV have identified the SP and MG as targets to induce NAb in the serum of patients [48]. Furthermore, in SARS-CoV-2, IgG-like NAb, which targets the receptor-binding domain (RBD) of SP, has been observed in patients with severe COVID-19 and convalescent patients [49,50]. NAb of NP has been related to the number of specific TLs which produce IFN-γ [50]. In SARS-CoV, NP, MG, and E have been shown not to produce NAb; however, they are potential antigens for antiviral cytotoxic T cells (CTL) [51]. Taking into account the above, SP, NP, and MG proteins are the main targets for stimulating the creation of antibodies through B lymphocytes (BLs) and TLs response when using specific peptides and incomplete proteins. Theoretically, individuals without previous infection by SARS or COVID-19 have shown specific TLs capable of recognizing peptides from SARS-CoV-2 polyprotein 1 a/b that could stop the viral cycle at an early stage, thus avoiding an expansive immune effect [52]. Although the impact of these findings is unknown, studies derived from other coronaviruses could indicate greater protection from a worse clinical outcome given by less severity of the disease [53]. However, a rational vaccine approach could include: (1) neutralization based on structural proteins, (2) early detection of replication-related proteins, and (3) accessory proteins involved in the formation of the mature virion, as well as protection against the residual damage in NP.

In LATAM, the outbreak has had negative sociocultural, economic, and political effects. Without a doubt, everyone will be affected by further outbreaks, especially vulnerable populations. The construction of a vaccine is therefore imperative for reducing health costs and morbimortality associated with COVID-19 [5,11–13].

Using an immunoinformatics approach, we propose a multi-epitope peptide vaccine model based on peptide binding properties extracted from conserved regions of SARS-CoV-2 proteomes. Moreover, we provide considerations for binding these peptides with HLA-I and II alleles most frequently found in LATAM (Figure 1).
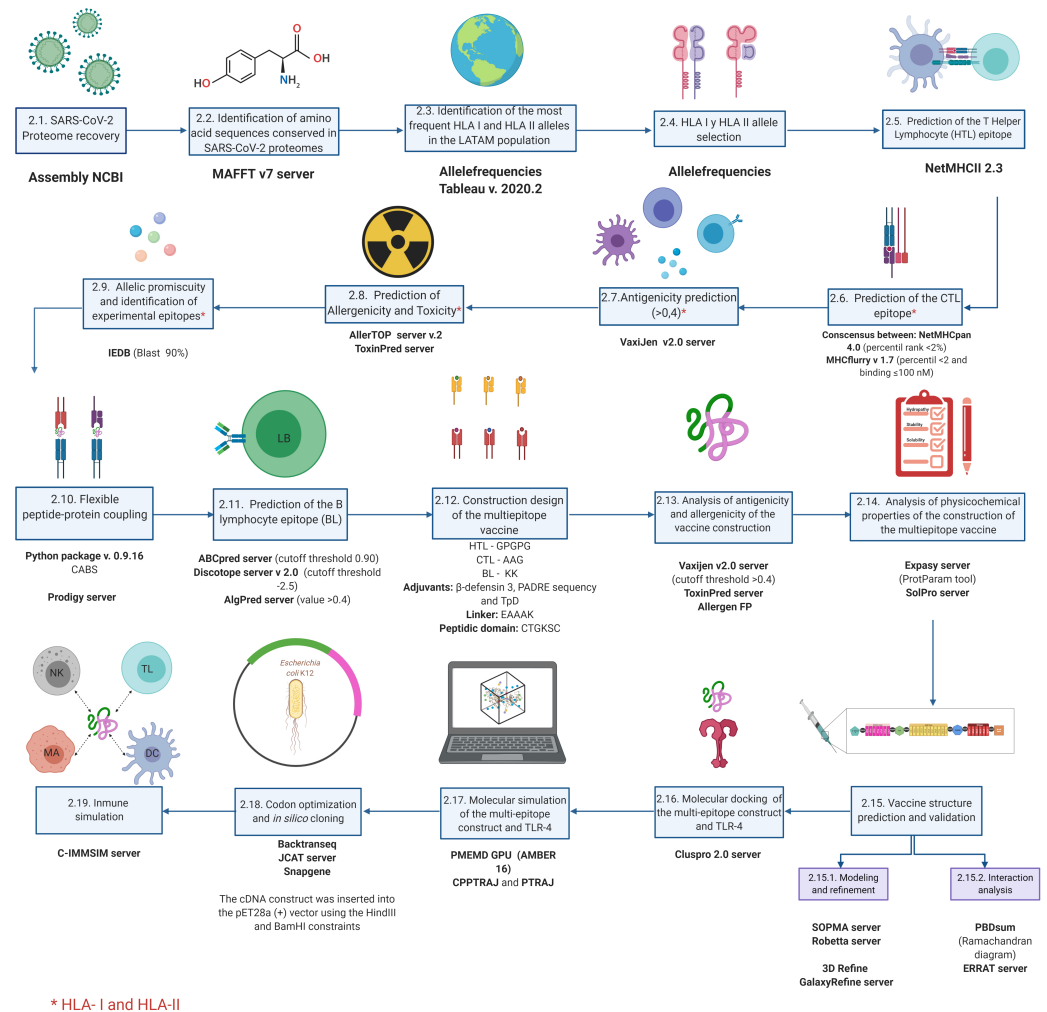
Figure 1. Vaccine development pipeline: the design stages of the multi-epitope vaccine proposal are shown. Each stage is important for meeting the objective of identifying peptides capable of generating an immune response, taking into account the most frequent alleles in Latin America. The important servers and cutoff values used in the process are specified.

2. Materials and Methods

2.1. SARS-CoV-2 Proteome Recovery

The proteomes of 92 SARS-CoV-2 strains were downloaded from the Assembly database available at the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/assembly, accessed on 27 April 2020). The metadata was extracted manually. Origin, isolation type, and sequencing date were identified. The interactions between SARS-CoV-2 proteins and the other factors studied were illustrated using the freely available web resource BioRender.com.

2.2. Identification of Amino Acid Sequences Conserved in SARS-CoV-2 Proteomes

To exclude the potential non-synonymous substitutions in the amino acid sequences from the recovered SARS-CoV-2 proteomes, we performed multiple alignments using the MAFFT server v7.0 (https://mafft.cbrc.jp/alignment/server/, accessed on 29 April 2020) [54], with the genome NC_04551 as a reference. This contained structural and accessory proteins, including the transcript ORF1a/b. The resulting preserved sequences formed non-redundant sequences which were used for further analysis.

2.3. Identification of the Most Frequent HLA-I and HLA-II Alleles in LATAM Population

The allele frequencies for each HLA-I (HLA-A, HLA- B and HLA-C) and HLA-II (HLA-DRB) allele for the Central and South American regions were downloaded using a specialist search tool (http://www.allelefrequencies.net/default.asp, accessed on 10 May 2020). The phenotype percentage (PF) was calculated using Equation (1). The most frequent phenotypes were separated. PF values greater than the Me were grouped by locus and country. Tableau v2020.2 [55] was used to create a geographic representation of the most frequent alleles.

$$PF = 1 - (1 - AlleleFrequency)^2 \qquad (1)$$

2.4. HLA-I and HLA-II Allele Selection

Only the alleles corresponding to the DRB loci were identified for HLA-II analysis due to the higher precision of these predictions to identify individual training and cross-training with other HLA-II molecules in algorithms based on artificial neural networks (such as NetMHCII and NetMHCIIpan). In general, HLA-I algorithms had a greater number of trained molecules and produced more accurate results than those of HLA-II. DRB is highly diverse in nature: >3000 isolates have been identified in humans [56]. This variability seems to be caused by the rapid and diverse evolutionary response of extracellular pathogens [57].

The HLA-DQ loci were not considered due to their association with autoimmunity and the instability of the complexes with their own epitopes. Additionally, the HLA-DP loci were not taken into account due to their less frequent associations with pathogens, which primarily contribute to tolerance and the proper function of innate virus response [58].

For HLA-I, the individual alleles most frequently found in LATAM at loci A, B, and C were used instead of identifying the promiscuous alleles more likely to identify SARS-CoV-2 proteins because the latter approach may lead to reduced precision when searching for an effective vaccine. The HLA-I group of genes is also isolated more frequently than those in DRB, with the B locus being the most diverse.

2.5. Prediction of T Helper Lymphocytes (HTL) Epitope

To predict TLs and CD4+ epitopes, we used conserved sequences of SARS-CoV-2 structural proteins with at least 15 mers. The algorithm NetMHCII v2.3 was used. This approach is based on training individual molecules from complex experiments, which gives greater accuracy [59]. The molecules available to make an agglutination prediction were as follows: DRB1_0101,DRB1_0103 DRB1_0301,DRB1_0401,DRB1_0402,DRB1_0403, DRB1_0404, DRB1_0405, DRB1_0701, DRB1_0801, DRB1_0802, DRB1_0901, DRB1_1001, DRB1_1101, DRB1_1201, DRB1_1301, DRB1_1302, DRB1_1501 and DRB1_1602. For each allele, all predictions were grouped and then filtered by <2% rank to identify probable binders that had a greater number of predictions with <IC50 affinities. The predictions that remained after filtering were considered to be potential HLA-II epitopes capable of being recognized by HTL CD4+.

2.6. Prediction of the CTL Epitope

To break through the SARS-CoV-2 viral assembly at an early stage of infection, we used structural accessories, and NSPs to predict potential epitopes for CTLs. We predicted peptides related to HLA molecules most common in LATAM, using the sequences of proteomes with at least 9 mers.

The two algorithms with the best experimental correlations were used, NetMHCpan and MHCflurry [60,61]. These approaches used cross-trained and individually trained neural networks. These approaches are based on experimental data from the Immune Epitope Database and Analysis Resource (IEBD) database, integration of eluted peptides, and other ligands identified by mass spectrometry.

A portable version of NetMHCpan 4.0 was used following request to the "Health Tech" area of the Technical University of Denmark (https://services.healthtech.d-tu.dk/

cgi-bin/sw_request, accessed on 3 May 2020). MHCflurry v1.6.1 was downloaded from the author's github repository (https://github.com/openvax/mhcflurry/releases, accessed on 3 May 2020). The results were filtered using a strong binding prediction (<2% rank) based on the author's recommendations and using a methodology based on binding affinity. In MHCflurry, the sequences were filtered with an affinity identity percentile <2 and cutoff values for predicting bond strength of up to 100 nM.

The peptide sequences resulting from the algorithms were tested for their immunogenicity using an immunogenicity tool that uses the position of the residues in the HLA molecule cleft and characteristics such as their basic nature and size to predict the interaction with the CD8+ CTL receptor and the initiation of an immunogenic response. Epitopes capable of generating an immune response by CTLs were grouped by their positive score. Those with positive values were considered candidates for further evaluation of antigenicity, allergenicity, and toxicity [62].

2.7. Antigenicity Prediction

The intrinsic characteristics of antigens, such as the protein nature, structure, physicochemical and extrinsic properties, are known to be related to immune response and are largely regulated by HLA, self-tolerance and host genetics [63]. The Vaxijen 2.0 server (http://www.ddg-harmfac.net/vaxijen/VaxiJen/VaxiJen.html, accessed on 19 May 2020) was used to identify non-redundant epitopes predicted to bind strongly to HLA-I and II molecules with antigenic potential. This server uses an antigen reference database and compares the physicochemical properties of the amino acids by cross-covariance. It converts the sequences of non-redundant epitopes that result in strong agglutinations into uniform vectors [64]. This process discriminates between probable and unlikely antigens. Peptides with strong agglutinations and an antigenicity threshold of >0.4 were considered to be potential immunogens.

2.8. Prediction of Allergenicity and Toxicity

To rule out probable allergic and toxic reactions caused by the interaction of peptides, we used AllerTOP v. 2.0. (https://www.ddg-pharmfac.net/AllerTOP/, accessed on 19 May 2020) [65] and ToxinPred (http://crdd.osdd.net/raghava/toxinpred/, accessed on 19 May 2020) [66]. AllerTOP employs a similar approach to Vaxijen 2.0: it uses an auto-covariance transformation to normalize the alignment of peptides with immunogenic potential and includes an automatic and manual pull of cured allergenic and non-allergenic proteins. Probable allergens are determined by several automatic machine learning techniques. These methods have a sensitivity and specificity of 0.87 and 0.90, respectively, which is achieved using descriptors of the allergy-related characteristics of individual amino acids [61]. ToxinPred uses a machine vector support technique to discriminate, via amino acid composition analysis and a quantitative matrix, the peptides with immunogenic potential and toxic probability. The position and frequency were analyzed along with other characteristics of the amino acids that are most abundant in toxic peptides to obtain accuracy close to 97%. Non-toxic and non-allergenic peptides with immunogenic potential were considered to be potential SARS-CoV-2 vaccine targets.

2.9. Allelic Promiscuity and Identification of Experimental Epitopes

Matrices containing potential vaccine epitopes of HLA-I and HLA-II molecules were created to identify the promiscuity of each of the predictions with vaccine potential. They were characterized as follows: peptides conserved in proteomes of 9 or 15 mer, immunogenic characteristics, non-toxic and non-allergenic, and a strong affinity to the most common alleles found in LATAM. These matrices were grouped by allelic HLA class to identify possible non-redundant combinations capable of covering all the HLA molecules tested. Peptides had equal predictions, the highest score from Vaxijen was used to decide the best peptide candidate. The groups with the lowest number of peptides were considered to be optimal for experimental validation as part of the rational multi-epitope construct.

Given that experimentally validated peptides in closely related viruses such as SARS-CoV could also be immunogenic targets, the peptides comprising the multi-epitope construct were subject to a 90% BLAST search in the Immune Epitope Database Analysis Resource (IEDB; https://www.iedb.org/, accessed on 3 November 2020).

2.10. Flexible Peptide-Protein Docking

We estimated the conformation of complexes between peptides with vaccine potential and the HLA molecules found most frequently in LATAM. These must interact as a flexible and stable anchor that allows interaction with CD4+ and CD8+ TLs receptors [67]. The CABS coupling algorithm was used to perform global molecular coupling between proteins and peptides. This allows full flexibility of the peptide and receptor backbone [68]. A standalone python package of CABS v 0.9.16 was downloaded from a repository (https://bitbucket.org/lcbio/cabsdock/downloads/, accessed on 4 July 2020).

The HLA molecules with experimental resolution found in RCSB PDB were used as the receptors. We identified the corresponding chains, according to the HLA class (α and β-2 globulin for HLA-I; α and β for HLA-II) while predicting the secondary structure of the peptides by PSIPRED through the RPBS web portal (https://bioserv.rpbs.univ-paris-diderot.fr/index.html, accessed on 4 July 2020). The default parameters were used for all other parameters. The energy calculations of the complexes were calculated using the Prodigy server (https://bianca.science.uu.nl/prodigy/, accessed on 4 July 2020) [69].

2.11. Prediction of the BL Epitope

To identify potential continuous epitopes capable of stimulating a BL response, we used the artificial neural network-based ABCpred tool [70]. The cutoff threshold was set to 0.90, which allowed for a greater specificity to be selected. To predict potential discontinuous BL epitopes, Discotope v2.0 [71] (server version) (http://www.cbs.dtu.dk/services/DiscoTope/, accessed on 20 July 2020) was used. The cutoff threshold and specificity were set to -2.5 and 80%, respectively.

This prediction included both SP and NP, which were extracted from the RCSB PDB with the following identifiers: 6lzg, 6m0j, 6vw1, 6w41, 6yla, 6yor, 6csb, 6vxx, 6vyb, 6m3m, 6vyo and 6wkp. Predictions resulting from the algorithms were considered to be potential epitopes. The peptide with the highest ABCpred score was also chosen with a partial or total presence in the experimental sequences identified as SARS-CoV immunological targets in IEBD with a 90% BLAST alignment. Also, it was located in potentially immunogenic regions, either because they were shown in Discotope-provided areas from various experimental structures or because they were contained in a specific protein domain. The antigenic peptides with values >0.4 in Vaxijen that were non-toxic and non-allergenic according to AlgPred [72] (http://crdd.osdd.net/raghava/algpred/submission.html, accessed on 25 July 2020) were chosen as potential BL epitopes and were added to the multi-epitope construct.

2.12. Recuperation of Validated Epitopes and Identification of Post-Transcriptional Modifications

To analyze the possible post-transcriptional modifications of potential promiscuous vaccine peptides (PPVPs), we used the NetNGlyc 1.0 server [73] (http://www.cbs.dtu.dk/services/NetNGlyc/, accessed on 3 November 2020) with the Wuhan-Hu-1 strain being used as a reference for the accessory proteins and NP. Regarding SP, the structure PDB:6VSB was used for non-structural proteins, which are replicated in the cytoplasm but have some luminal domains that are prone to N-glycosylation. UNIPROT (https://www.uniprot.org, accessed on 3 November 2020) was used to pinpoint their cellular location.

To identify potential immunogenic regions as well as recent validations of the PPVPs, we downloaded the epitope database available on 3 November 2020. It consisted of 822 validated epitopes for TLs and 330 for BLs from SARS-CoV-2 assays. Additionally, we used supplementary assays as targets for the HLA molecule that comes from immunitrack (https://www.immunitrack.com/free-coronavirus-report-for-download/, accessed on

3 November 2020), which consists of a library of peptides tested as binders of various HLA-I and II alleles, taking as a threshold reference stability greater than 30, according to authors' recommendations.

2.13. Multi-Epitope Vaccine Construct Design

To create a multi-epitope vaccine, potential HTL, CTL, and BL epitopes were linked using GPGPG, AAG and KK linkers, respectively. For a better immunogenic response, four adjuvants were added using the EAAAK linker, *β*-defensin 3, Tetanus toxin peptide and a Diphtheria toxin peptide linked by a cathepsin cleavage site (TpD), and a universal T helper epitope (PADRE). A peptide domain (CTGKSC) targeted by M cells was also added to the C terminal, promoting transcytosis between enterocytes and antigenic uptake at the intestinal level [74].

2.14. Analysis of Antigenicity and Allergenicity of the Vaccine Construct

To define whether the final design of the multi-epitope proposal was safe and viable, its antigenic capacity was considered by predicting antigenicity with Vaxijen values with a threshold of >0.4. Its toxicity was predicted using ToxinPred and its allergenicity using Allergen FP [75] (http://ddg-pharmfac.net/AllergenFP/, accessed on 1 December 2020).

2.15. Analysis of the Physicochemical Properties of the Multi-Epitope Vaccine Construct

To analyze the construct's physicochemical properties, the Expasy server and the ProtParam tool [76] were used to determine optimal recognition by the immune system. Negative values for the grand average of hydropathy (GRAVY) [77] and stability are especially important for adequate antigen presentation, as well as determining solubility, which was performed using SolPro. Other parameters related to production and expression were also considered, including the aliphatic index (related to the construct thermostability).

2.16. Vaccine Structure Prediction and Validation

The secondary structure of the multi-epitope design was predicted using SOPMA [78] and PSI-PRED [79]. The tertiary structure was predicted using the Robetta server based on homology [80]. An evaluation of the structure quality was carried out using the Ramachandran diagram in the PDBsum [81] and by the ERRAT server [82]. Refinement was performed using 3D Refine [83] and then GalaxyRefine [84].

2.17. Molecular Docking of the Multi-Epitope Construct and TLR-4

The TLR-4 receptor was selected as the ideal immunological target of the multi-epitope construct since it is capable of inducing INF production and expressing itself in the cell membrane of dendritic cells. We studied the resulting interactions in the stable formation of the TLR-4/ Myeloid Differentiation Factor 2 (MD-2) complex and the multi-epitope construct using molecular docking, performed using the Cluspro server 2.0 [85] (https://cluspro.bu.edu/login.php, accessed on 10 December 2020). The TLR-4/MD-2 hetero-tetramer was used as a receptor (obtained from PDB RCSB database; ID: 3FXI), and the refined multi-epitope construct was used as a ligand. The residues at the binding interface of the resulting complex were analyzed by PDBsum and plotted by UCSF Chimera v1.14 [86].

2.18. Molecular Dynamics Simulation of the Multi-Epitope Construct and TLR-4 Complex

Coordinates of the best model of the multi-epitope construct and TLR-4 were used to perform molecular dynamics analysis. Minimization and molecular dynamic protocols were performed with AMBER 16 [87]. ff14SB force field parameters were used for the amino acid residues [88]. The complex was subjected to unrestricted molecular dynamic simulations for all atoms in an explicit solvent using the PMEMD GPU version algorithm in Amber16 [87].

The Leap module integrated within Amber16 was used to add missing hydrogen atoms and add Cl^- ions for neutralization. The systems were immersed in an orthorhombic box using the TIP3P [89] water model. The long-range electrostatic interactions were calculated using the particle mesh Ewald method [90], with a direct space and a vdW cutoff of 12 Å//. An initial minimization was applied to the solute using a potential of 500 kcal mol^{-1} Å//^{2} for 10,000 steps using the steepest descent algorithm, followed by 10,000 steps with the conjugate gradient method. Subsequently, 10,000 unrestricted minimization steps were simulated using a conjugate gradient algorithm.

The heating protocol was carried out with a gradual increase in temperature from 0 to 310.15 K using a harmonic restriction of 5 kcal mol^{-1} Å//^{2} applied to the solute. A Langevine thermostat with a collision frequency of 1 ps^{-1} was used with the canonical assembly (NVT). The complex was equilibrated at 310.15 K in an NPT assembly for 10 ns without restriction using the Berendsen barostat to maintain the pressure at 1 bar. The SHAKE [91] algorithm was used to restrict the bonds of all hydrogen atoms. A 2 fs time-step was used with the precision model SPFP [92] in the molecular dynamic simulation.

Finally, 68 ns of production was simulated in an NPT assembly with a target pressure of 1 bar and a pressure coupling constant of 2 ps. Production trajectories were analyzed each 2 ps along the whole simulation using CPPTRAJ and PTRAJ [93].

2.19. Codon Optimization and In Silico Cloning

To propose a realistic scenario for peptide vaccine cloning, in silico analyses were conducted to identify the best options for the expression and isolation of the multi-epitope construct. The *Escherichia coli* (*E. coli*) K 12 expression system was selected since it is inexpensive to grow, relatively easy to manipulate genetically, and generally produces high levels of recombinant proteins. Also, it has an optimized lineage for overexpression of recombinant proteins [94]. To identify the best cloning strategy, the complete protein sequence of the multi-epitope construct was first converted into cDNA using Backtranseq reverse translation. The resulting cDNA sequence was further optimized by adapting the most frequently used codons of *E. coli* K 12 to enhance protein expression using the JCAT server [95]. To achieve adequate protein purification, the plasmid vector pET-28a(+) was chosen because of the possibility of labelling polyhistidine towards the N or C terminals of the multi-epitope construct. Therefore, a complete protein was obtained by avoiding possible truncated proteins.

Once the appropriate scheme for cloning was identified, SnapGene v5.1.2 was used to provide enhanced flexibility for displaying and annotating sequences. For this, the HindIII and BamHI restriction sites were used. A cut was made that enabled us to retake a closed structure of the plasmid vector with the appropriate position of the optimized genetic sequence of the construct.

2.20. Immune Simulation

The immunogenic behavior of the multi-epitope construct was simulated using the C-IMMSIM server (https://kraken.iac.rm.cnr.it/C-IMMSIM/, accessed on 19 December 2020). This agent-based computer model handles a diverse number of cells representing innate and acquired immunity. By following a set of rules obtained at an experimental level, interactions with the vaccine construct are capable of simulating behaviors that may suggest the probable generation of immune memory. This is achieved by combining the mesoscopic scale of the immune system using three compartments: the bone marrow, thymus, and lymphatic organs. In addition, it uses deep learning tools and molecular level techniques to predict the interaction of the construct and its affinity from the matrices of some HLA molecules. The algorithm can also identify probable linear BLs epitopes from physicochemical parameters. The minimum inter-dose time for current vaccines is no more than 4 weeks. For this reason, simulation values were adjusted with three injections separated by 1, 84, and 168 time-steps, resulting in <4 weeks between doses. The simulation was completed in up to 200 time-steps, with other predetermined values.

To test the response capacity from the interaction between the multi-epitope construction and the immune system, a viral challenge was performed one year after the start of the extended vaccine scheme and was simulated beyond day 460.

3. Results

3.1. Recovery of SARS-CoV-2 Proteomes

Proteomes from 92 isolates of SARS-CoV-2 were recovered in FASTA format using the GenBank database as a reference. Metadata related to the collection date, city, host, and source of isolation were also downloaded. Isolates were collected by nasopharyngeal (n = 31) and oropharyngeal (n = 12) swabs, and to a lesser extent by bronchoalveolar lavage (n = 11). Sequencing was performed from December 2019 (the Wuhan-Hu-1 strain) to 11 March 2020 (the SARS-CoV-2/human/USA/PC00101P/2020 isolate). Most sequences came from the United States of America (n = 51) and China (n = 27). Only one isolate came from LATAM (Brazil, February 2020). The entire metadata set of proteomes used, as well as access identification, is shown in (Table S1).

3.2. A Vaccine for SARS-CoV-2 Must Take into Account Non-Structural Proteins

"The spike protein acts like an Early Trojan horse."

SP plays a fundamental role in SARS-CoV-2 infection because it mediates the virus entry into host cells through its S1 domain, which binds to the ACE2 receptor making the subsequent cleavage of the S2 domain.

The increased glycosylation of SP allows it to evade the adaptive immune response and protect epitopes from recognition and antibody neutralization. According to a recent study, antibodies derived from exposure to SP could be related to the clinical outcome by targeting surfactant proteins related to alveolar surfactant [96]. Through viral proteins and RNA receptors, the immune response is the first line of defense against SARS-CoV-2 infection, producing cytokines and IFN-1 [97]. Therefore promoting viral clearance by activating a systemic antiviral state.

As with SARS-CoV, SARS-CoV-2 has proteins encoded by the coronavirus genome that are capable of interfering with the innate immune response subsequently an early phase of infection. This affects various signaling pathways important in maintaining immunity [98] (Figure S3). Opportunely detecting and stopping the pathway interferences could be a more complete approach to vaccination as opposed to only neutralizing RBD or other SP domains. SARS-CoV-2 evidently, has shown that early translated proteins have functions beyond viral replication. They are also capable of deeply and negatively modulating pathways related to IFN, including its synthesis [99].

For example, NSP3 possesses papain-like protease (PLP) domains. Making it is one of the two proteases that participate in the cleavage of NSPs from polyprotein 1 a/b. PLP in SARS-CoV-2 has a more evident dual de-ISGylation capacity than deubiquitination [100]. This impacts pathways up-regulated by IFN, including viral peptide presentation and processing [101].

Although NSP3 is barely evidenced when compared to structural proteins in SARS-CoV and SARS-CoV-2 [102], it is in cohesion with accessory proteins, participating in the assembly of the mature virion [102].

NSP3, and to a lesser extent NSP2 and NSP12, have been identified as the NSPs with the most translation in cell cultures. Other NSPs are associated with the altered transcription and translation of proteins in the host, including IFN [99]. This is achieved by disrupting messenger RNA (NSP16) splicing that selectively blocks translation in ribosomes. This only allows the leading sequence of the viral subgenomic RNA to continue (NSP1) and interrupts the passage of proteins to the membrane. This includes secretory proteins such as IFN, cytokines, and HLA by affecting the SRP complex (NSP8-NSP9, NSP8, NSP9).

Therefore, identifying the mechanisms of action, characteristics of structural NSPs, and accessory transcripts within the process of recognition, entrance, invasion, and replication, is important for finding potential vaccine targets. When given a set of proteins and knowing their mechanism of action inside the cell, identifying the NSPs may be an early approach to avoiding extensive immune involvement. NSPs produced at an early stage and necessary for replication should be considered. Structural protein production is also important since it is essential in the SARS-CoV-2 life cycle.
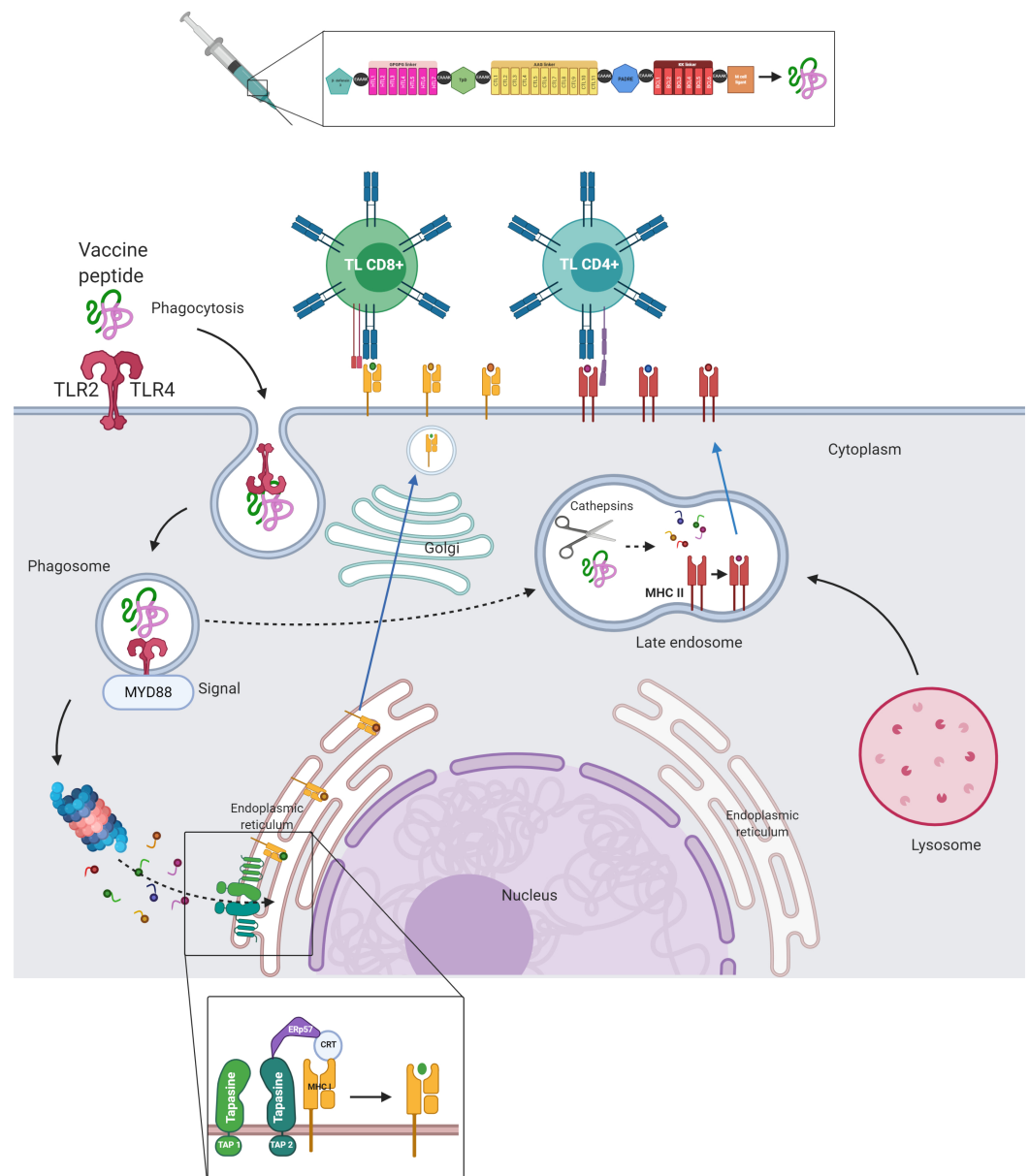

Figure 2. Antigenic presentation from the proposed vaccine: The assembly of peptides with vaccine potential are recognised by cell membrane receptors capable of recognising patterns associated with pathogens, such as TLR2 and 4 from dendritic cells. After recognition, the construct is phagocytized by the cell together with TLR, allowing its interaction with MyD88 and the maturation of the phagosome. From the phagosome, the peptide can take two routes: the first route is towards the proteasome, where the peptide is degraded, internalised in the ER and assembled with HLA-I molecules; the second route involves the internalization of the peptide in the late endosome, where it is assembled with HLA-II molecules and subsequently presented on the cell membrane of the LT [103–105].

The frequency and alignment of 92 proteomes were considered to identify conserved regions. This was based on the evidence of the effect of neutralizing antibodies towards structural proteins and the activation/recognition of TL towards structural, non-structural, and accessory proteins from SARS and COVID-19 convalescent patients. Our findings are shown in (Table S2). The conserved sequence blocks were used in their entirety for the prediction of conserved epitopes. These characteristics are included in our vaccine proposal summarized in (Figure 2).

3.3. Most Frequently Identified HLA-I and HLA-II Alleles in LATAM

HLA systems in humans are located in the most polymorphic region of the genome, related to the continuous selection when interacting with extracellular and intracellular pathogens. Furthermore, this diversity corresponds to the rapid adaptation to environmental change, and actually, this has become a hallmark of the first migrations of various human populations [106].

When identifying vaccine targets for SARS-CoV-2 and other emerging pathogens of probable zoonotic origin, it may be possible to generate an immunological response memory. While this would only work in certain populations, given that it is restricted to HLA molecules, it could occur because of the diverse affinities that HLA-I and II molecules have for different sets of peptides that are processed from SARS-CoV-2.

Identifying people and populations with greater susceptibility to COVID-19 could help construct effective healthcare focused on patients who are more susceptible and have an increased risk of mortality. Theoretically, HLA capable of recognizing a greater range of SARS-CoV-2 peptides has been found through computational algorithms, which have been validated in clinical settings, and are associated with better results, especially in heterozygous individuals [107].

Although we lack a precise medicine model that identifies HLA and peptide binding capacity from the SARS-CoV-2 proteome, global initiatives like HLA COVID-19 has been launched to find relevant translational strategies. Additionally, other authors have proposed routine HLA typing in patients with COVID-19 [108].

We identified 168 of the most frequent HLA-I alleles in 17 countries. The HLA-A*02:01 allele was found at an above-average frequency in Argentina, Brazil, Chile, Colombia, Cuba, Ecuador, Nicaragua, Peru, and Venezuela. The alleles A*24:02 and B*40:02 were identified above the 80th percentile; they rank first in frequency in most countries. Some countries had missing data and small sample sizes, e.g., Guatemala (where only the B*53:01 allele was found above the mean) and Trinidad and Tobago (where the only available allele, C*16:02, was used for complementary analyses).

As shown in (Table S3), the frequencies found between the HLA-I loci were A (n = 126), B (n = 264), and C (n = 64); those alleles above the mean frequencies were A (n = 48), B (n = 91) and C (n = 27).

For HLA-II, none of the countries shared an allele higher than the Median (Me) per country. DRB1*03:01 allele was found with greater frequency in at least 14 countries, except for 3: Mexico, Venezuela, and Paraguay. The HLA-II alleles above the 90th percentile were correlated with most of the allelic groups recovered, except for DRB1*10 and DRB1*12 allelic groups, which were not represented by alleles. DRB1 * 14: 02 was the most frequent allele and was isolated mainly from the Colombian, Brazil, and Argentina populations. The allelic groups and the most frequent alleles found in 18 countries are attached in (Table S4). All DRB1 alleles listed by NetMHC 3.2, spanning 18 countries, had above average frequencies. (Figures S4 and S5) shows the distribution of the main alleles and allelic groups by country according to HLA-I and HLA-II.

3.4. Potential Epitopes of TLs, CD4+, CD8+, and BLs are Contained in the Conserved Sequences of 92 SARS-CoV-2 Proteomes

3.4.1. HLA-II

To identify and select the potential HLA-II epitopes, we used 19 HLA-II molecules from the available DRB gene alleles to perform an agglutination prediction using NetMHCII 2.3. We excluded the conserved sequences that were <15 mers from SP, NP, and MG proteins. In total, 199 peptides with strong binding capacities and antigens were identified. The proteins with the highest number of agglutinations was SP (n = 114), followed by MG (n = 62) and NP (n = 23). The origin of predictions from the conserved sequence and the 199 antigenic peptides are included in (Table S5).

Table S6 summarizes the conserved sequences present in the SP binding domain of interest. DRB1*01:01, DRB1*04:05, DRB1*10:01, and DRB1*16:02 alleles were found to be promiscuous in this domain (predicted to be strong binders). The allele DRB1*16:02 is found more frequently in LATAM; therefore, it may be related to clinical outcomes and deserves to be studied in more detail.

The 15 predicted mers, with antigenic and strong binding characteristics, were evaluated for allergenic and toxic traits using AllerTOP and ToxinPred, respectively.

3.4.2. HLA-I

To achieve a more precise identification of strong binders from the 167 most frequent HLA-I molecules in LATAM, we used two algorithms based on artificial neural networks identified as the most accurate by IEBD (as of 15 March 2019).

A set of 944 peptides were classified as strong binders and antigens. These were the result of a consensus from the two algorithms. The proteins with the highest number of predictions in order of frequency were as follows: ORF1 transcript (n = 573), SP (n = 163), MG (n = 51), ORF-3a transcript (n = 43), and NP (n = 34). The contribution made by each conserved sequence to these proteins and the binders are included in (Table S7).

A prediction was not possible for the HLA-B*51:10 allele because it was unavailable in the list. In addition to the antigenicity and strong binding characteristics of these groups, we also considered those with vaccine and immunogenic potential, non-allergenic traits, and non-toxic characteristics on our list. Adaptive immunity protection is now known to change the natural course of the disease by neutralizing SARS-CoV-2 [109].

3.4.3. BLs

The most recent experimental structures of the SP and NP proteins found in RCSB PDB (https://www.rcsb.org/, accessed on 14 July 2020) with the following IDs were used: 6lzg, 6m0j, 6vw1, 6w41, 6yla, 6yor, 6csb, 6vxx, 6vyb, 6m3m, 6vyo, and 6wkp. These structures have special characteristics, such as binding to ACE2 receptor, an antibody extracted from a convalescent SARS patient named CR3022 [110], and conformational pre-fusion in the open and closed state. Analyses were performed using the Discotope 2.0 server adjusted to 80% specificity.

Using the specific SP or NP chains, we implemented the prediction algorithm on all proteins based on their available 3D structure. The predictions, characteristics of the complexes, intervals, and frequency with which probable epitopes were identified in each structure are shown in (Table S8). For SP, we have indicated three possible continuous epitopes and one discontinuous region that may be important due to the extracellular access. The areas between residues 443-450, 487-494, 496-506, and discontinuous 454-459-460-469-471 seem to be more frequently related to probable interactions with BL. A principal feature of these predictions is the absence of glycosylation in the residues.

For NP, we identified five regions of possible continuous epitopes and two discontinuous ones for the RNA binding domain. On average, the intervals are longer than SP. The regions comprise the intervals in the residues: 59–64, 91–106, 120–130, 136–148, 150–156 and discontinuous between different intervals in the residues 66–82, 115–130, 163–171. The

NP and SP with linear and discontinuous epitopes zones described above are illustrated in (Figure S6).

The linear epitope intervals of SP and NP described in the previous structural approach were taken into account, to identify longer sequences (>15 mers), which would be capable of evoking a response mediated by the BL receptor. For this approach, we used the ABCpred server and adjusted specificity to 90%.

Other filters were used to identify two vaccine peptide candidates for BLs, within of SP and NP proteins from SARS-CoV-2, that are presented in (Table S9). The filters included predicting epitopes in the functional domains using RDB for SP or the RNA binding site for NP, antigenicity, non-toxicity, non-allergenicity, and the identification of conserved sequences in SARS-CoV with experimental immunological evidence available at IEBD. After the filters were applied, the best options were selected to choose the linear epitopes candidates for BL to be incorporated into the multi-epitope construction. These sequences are summarized in (Table S10).

3.5. SARS-CoV-2 MG and SP Contain Conserved Sequences Capable of Interacting with a Large Number of HLA-II Molecules

For HLA-II molecules, a Me of 8 agglutinations was found in the 18 molecules analyzed. The five main alleles that were found to be more capable of recognizing peptides in at least 15 mers were: DRB1*04: 02 (n = 50), DRB1*09: 01 (n = 35), DRB1*08: 01-DRB1*16: 02 (n = 33), DRB1*04: 03 (n = 30), which only appear in their entirety in Chile and Brazil. The DRB1*16: 02 and DRB1*04: 03 alleles were more frequent in the Chile, Brazil, and Colombia populations respectively, which could be related to a better clinical outcome. Although Argentina had these alleles, the frequency in its population is low compared to DRB1*08: 02, which resulted in a lower ability to recognize peptides (n = 13). It should be noted that this allele was found more frequently in other countries such as Brazil, Colombia, and Peru, which could be related to an unfavorable clinical outcome. The alleles with the highest promiscuity in HLA-II molecules were SFRLFARTRSMWSFN (n = 7), a peptide that comes from the MG protein of SARS-CoV-2 and is partially conserved in SARS-CoV. QSIIAYTMSLGAENS (n = 4) and VLSFELLHAPATVCG (n = 4) from the SARS-CoV 2 SP protein followed. The agglutinations between the 199 antigenic peptides from the SP, NP, and MG proteins against HLA-II are shown in (Table S5).

3.6. Most Frequent HLA-1 in LATAM and Recognition of Peptides of Conserved SARS-CoV-2 Regions

The peptide with the highest promiscuity in HLA-I molecules was MPYFFTLLL (n = 66), which was found in the ORF1 transcript obtained from the SARS-CoV-2 proteome. Other peptides included FAMQMAYRF (n = 58), found in SP, and FLLNKEMYL (n = 42), which was located in the ORF1 transcript.

The alleles with the highest promiscuity for all the conserved proteins in SARS-CoV-2 proteomes (restricted to 9 mers) were preferentially found in the C locus, with a Me of 82 agglutinations and 18.2 σ. The molecules were HLA-C*08:03 (n = 98) with the highest agglutination, and HLA-C*04:01 (n = 42) with the lowest agglutination. Locus A follows with a Me of 47 agglutinations and 10 σ. With the molecules HLA-A*34:02, HLA-A*68:02 (n = 62), and HLA-A*36:01 (n = 22) with the highest and lowest agglutination, respectively.

Finally, the B locus with an Me of 40 agglutinations and 15.5 σ. With the molecules HLA-B*35:17, HLA-B*35:20, HLA-B*35:30 (n = 73), and HLA-B*27:05 (n = 8) with the highest agglutination and lowest agglutination, respectively.

In (Table S7), the total number of antigenic alleles and peptides with a strong binding affinity for HLA-I peptides obtained from the entire SARS proteome-CoV-2 are shown. For a more in-depth analysis of the HLA-I predictions, the alleles are classified from highest to lowest according to their theoretical ability to recognize HLA-I peptides. The proportions refer to allele frequency and locus found in LATAM countries (Tables S11 and S12).

For Locus A, the BRA-CHL-COL-CRI-VEN countries have populations that carry HLA-A alleles with a greater ability to recognize HLA-I peptides, while ARG-CUB-ECU-MTQ-NIC-PER has the lower capacity (Table S12-Locus A). For Locus B, we found that no country had a greater frequency for HLA-B alleles. Rather, LATAM showed a greater ability to recognize HLA-I peptides. The ARG-BRA-CHL-COL-CRI-CUB-NIC-PER countries presented a lower capacity to recognize HLA-B alleles, especially in VEN and CRI. (Table S12-Locus B).

Regarding Locus C, we had the least amount of data on HLA per country. Nevertheless, it showed a greater consistency, which resulted in a greater ability to recognize HLA-I peptides in BRA-COL-CRI-NIC-PER-TTO. In the case of CHL, HLA-C alleles had a lower capacity to recognize HLA-I peptides. (Table S12-Locus C).

There could be more susceptibility to mortality and morbidity due to SARS-CoV-2, especially in countries that share more than one locus with less capacity to recognize HLA-I peptides. Such is this case with ARG-CHL-CUB-NIC. The alleles most frequently found in LATAM included HLA-A*02:01, HLA-A*24:02 and, HLA-B*40:02. Based on their capacity to recognize peptides derived from SARS-CoV-2, these alleles were ranked 81st, 119th and, 154th, respectively.

3.7. Vaccine Candidates, Post-Translational Modifications, and Experimentally Validated Peptides

Matrices with vaccine potential were constructed to identify peptides associated with the most frequent HLA-I and HLA-II molecules in LATAM. The least number of peptides found to cover the most frequent LATAM alleles is summarized in Table 1.

Table 1. Potential promiscuous vaccine peptides for the most frequent HLA-I and HLA-II alleles in LATAM. Position: Refers to the first and last amino acid residue in the range occupied by the peptide in the source protein. Source protein: Refers to the proteins identified from the proteomes obtained from NCBI. Vaxijen: Antigenicity described by a threshold greater than 0.4. AllerTOP: Prediction of whether the peptide is a probable allergen. HLA: Experimental database referenced in methodology as Immunitrack, in which the stable coupling of the peptide with at least one HLA molecule is evidenced. IEBD: Refers to the identification of peptides at the base of IEBD specifying if the peptide is found as a ligand of HLA, BL, or TL or if it is contained in an immunogenic region validated experimentally by any assay. * Peptide is conserved in SARS-CoV-2 and SARS-CoV.

Number of Peptide	Peptide	Positions	Protein of Origin	Vaxijen	AllerTOP	HLA	IEBD
1	AAAYYVGYLQPRTFL	262–276	SP	0.48	Non-Allergen	n/a	Validated region
2	DDSEPVLKGVKLHYT *	1259–1273	SP	1.18	Non-Allergen	n/a	Validated peptide—Assay in BLs
3	LVIGAVILRGHLRIA	138–152	MG	0.88	Non-Allergen	n/a	n/a
4	QSIIAYTMSLGAENS	690–704	SP	0.57	Non-Allergen	n/a	Validated region
5	QSLLIVNNATNVVIK	115–129	SP	0.43	Non-Allergen	n/a	Validated region
6	SFRLFARTRSMWSFN	99–113	MG	0.80	Non-Allergen	n/a	n/a
7	VLSFELLHAPATVCG	512–526	SP	0.48	Non-Allergen	n/a	Validated region
8	CISTKHFYW	3147–3155	ORF1-NSP4	1.90	Non-Allergen	n/a	n/a
9	FAMQMAYRF	898–906	SP	1.03	Non-Allergen	n/a	Validated region
10	FLLNKEMYL *	3183–3191	ORF1-NSP4	0.44	Non-Allergen	HLA-A*02:01	Validated peptide—Assay in TLs.
11	GYKSVNITF	835–843	ORF1-NSP3	2.37	Non-Allergen	n/a	n/a
12	ITLCFTLKR	110–118	ORF7a	2.02	Non-Allergen	n/a	n/a
13	KRAKVTSAM	4022–4030	ORF1-NSP8	0.76	Non-Allergen	n/a	n/a
14	KVKYLYFIK *	4225–4233	ORF1-NSP9	1.06	Non-Allergen	HLA-A*11:01	Validated peptide—Assay in TLs.
15	LEMELTPVV	1012–1020	ORF1-NSP3	1.97	Non-Allergen	n/a	n/a
16	MPYFFTLLL	2169–2177	ORF1-NSP3	0.49	Non-Allergen	n/a	n/a
17	VMYASAVVL	3684–3692	ORF1-NSP6	0.48	Non-Allergen	n/a	n/a
18	WTAGAAAYY	258–266	SP	0.63	Non-Allergen	HLA-A*01:01	Validated region

These PPVPs share characteristics for the construction of multi-epitopes, including being antigenic, non-toxic and, non-allergenic. The first 7 peptides come from structural proteins that are directed towards HLA-II. The remaining 11 come from NSP, except for P9 and P12 that come from SP and ORF7a, respectively, and are directed towards HLA-I. From the SARS-CoV-2 dataset available in IEBD, the "Experimental evidence available" column highlights P10 as recently validated in three patients convalescing from COVID-19 (https://www.iedb.org/assay/12156798, accessed on 3 November 2020). Furthermore, peptides P1, P4, P5, P7, and P9 were harboured in linear epitopes of TLs validated experimentally. P10, P14, and P18 have recently been validated in other HLA restricted studies

(https://www.immunitrack.com/free-coronavirus-report-for-download/, accessed on 3 November 2020), Since SARS-CoV-2 can share conserved regions with SARS-CoV and takes longer to study, we investigated whether these PPVPs have been experimentally tested on TLs, HLA, or BLs ligands. With a Blast of 90% in IEBD, peptides P2, P10, and P14 are also conserved in SARS-CoV and have experimental validation. P2 was both HLA and BLs validated. recognized linear epitopes. So far, no specific immunogenic regions or experimental assays have been identified that validate T, B, or HLA cell ligands in the experimental information recovered. On the other hand, P3, P6, and P12 were only partially found in immunogenic regions in experimental SARS-CoV-2 assays. Since they have immunogenic characteristics indicated in the other experimentally validated peptides, these may be PPVPs that could delimit broader regions. Regarding the post-translational modifications, only P5 was found with an N-glycosylation evidenced experimentally in the PDB 6VSB crystallized structure. However, this epitope was completely harboured in a linear epitope in IEBD (test code 8160608), which is capable of being recognized by TLs from convalescent COVID-19 patients. Taking into account that P5 is found in an experimentally validated immunogenic region, it is tentative to think that it is capable of being presented by HLA. As a result of this, it was entered into the construct. In the SP of SARS-CoV-2, we did not identify glycan shields with densities that limit probable antibody recognition [111]. However, its interaction with innate immunity remains unknown. With these modifications, HLA may be able to recognize and present the peptides derived from antigenic processes when they are captured by dendritic cells [112]. These have been most frequently studied in cancer, evidenced in T and BLs primers, which are in some cases even more immunogenic [113–115]. Because they are naturally recognized as epitopes, their in-vitro recognition can be included in complementary studies. It quickly identifies the immunogenic regions susceptible to being glycosylated and therefore optimizes the potential vaccine formulations. Finally, in (Tables S5 and S7), the HLA-I and II alleles are shown to be strong binders of the PPVPs.

3.8. Flexible Peptide-Protein Coupling Signals Favourable Energy and Anchorage Residues to HLA Molecules

To estimate the conformation of the complexes, we used the following allele experimental structures as receptors: DRB1_0401, DRB1_1101, HLA-B*35:01 and HLA-C*07:02 (IDs in RCSB-PDB: 5NI9, 6CPN, 1XH3 and 5VGE, respectively). As peptide ligands, we used SFRLFARTRSMWSFN and FAMQMAYRF.

The best models were those contained in clusters with the highest density, according to the average root-mean-square deviation (RMSD) obtained from each simulation and their adequate position in the HLA cleft. The results are shown in Figure 3.

A map of the contacts between peptides and HLA-I molecules is shown with residues found to more frequently interact in the clusters (functioning as an anchor and generally located in the first and last residues). They also display interactions with β-2 globulin that stabilize the α1 and α2 chains, where the binding groove is formed, and facilitate peptide bonding [67].

In HLA-II, the regions that most frequently interact as anchors are FARTRSMWS for DRB1_0401 and FRLFARTRS for DRB1_1101 (although they differ from the core sequence of the prediction). The residues 4, 5 and 7 most frequently interact, giving a greater number of anchors than those in HLA-I; this results in flanking regions with small deviations in the peptide backbone and differing HLA cleft accommodation.
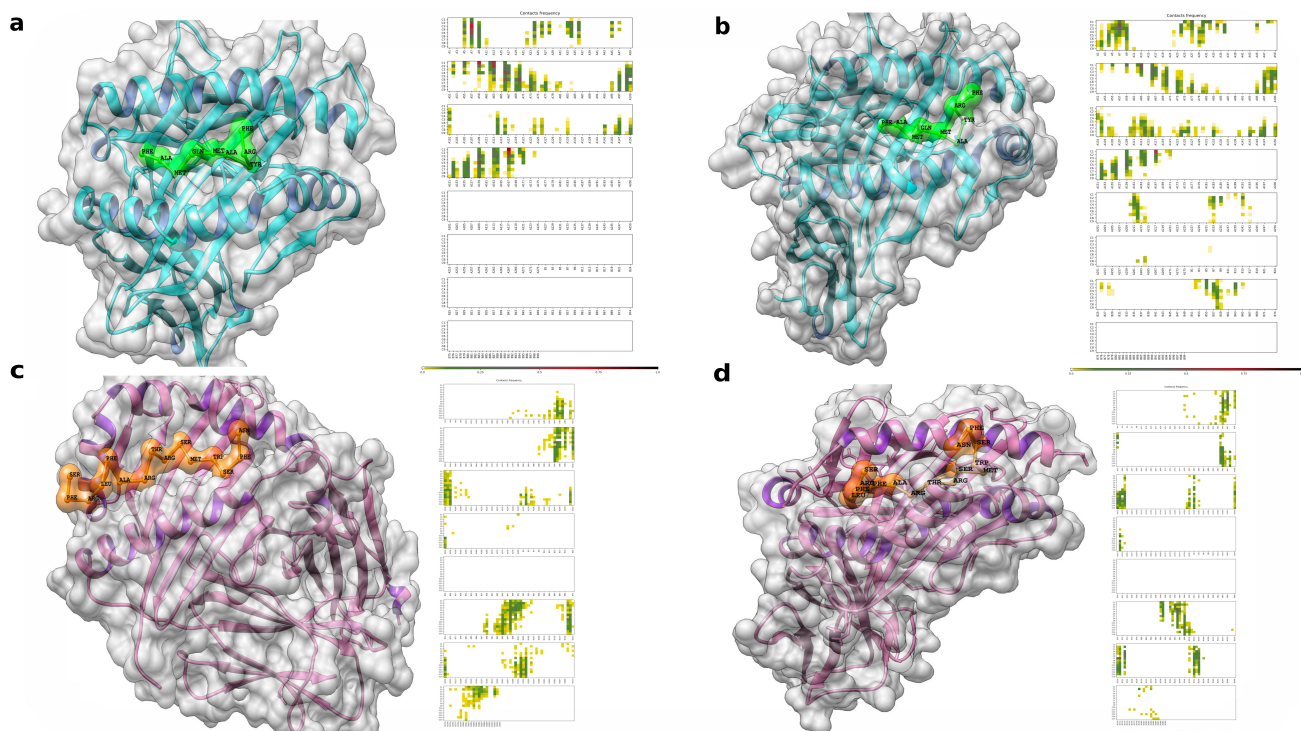
Figure 3. Estimation of flexible molecular protein-peptide coupling: In the complexes (upper panels) are two non-redundant HLA-I molecules most frequently found in the Latin American population coupled with a peptide with vaccine potential from SP, namely FAMQMAYRF. (**a**), Cluster 1, density of 239, ΔG of −5.2. (**b**), Cluster 1, density of 355, ΔG of −4.9. In the complexes (lower panels) are two non-redundant HLA-II molecules most frequently found in the Latin American population coupled with a peptide with vaccine potential from MG, namely SFRLFARTRSMWSFN. (**c**), Cluster 1, density of 138, ΔG of −4.6. (**d**), Cluster 2, density of 168, ΔG of −4.4.

3.9. Molecular Docking, Interactions with Heterodimer TLR-4/MD-2

In the N terminal, β-defensins have a range of immune responses related to cell maturation that mediate innate immunity, such as dendritic cells, TLs, and antiviral activities [116,117]. Another adjuvant used was the universal memory TLs helper peptide TpD, an auxiliary peptide that can aid memory generation as a target of TLs CD4+ [118]. The adjuvant Pan DR T helper epitope PADRE was also attached to the construct. It is relevant in multi-epitope constructs given its ability to potently stimulate the innate and humoral immune systems through high and specific IgG titer generation. It can also overcome barriers, indicated by the high diversity among HLA molecules; hence, it reaches a larger population and is safe [119,120]. Towards the C terminal, a peptide domain, CTGKSC, capable of interacting with M cells and mediating up to an 8-fold increase in intestinal absorption, was added [74]. CTGKSC has been formulated in oral multi-epitope vaccine candidates [74], acting as a transporter towards the M cells of the epithelium associated with the follicle in the Peyer's patches of the intestine in humans, where it can promote a specific immune response.

The linkers used as spacers between HTL GPGPG and AYY epitopes are suitable and facilitate antigen presentation by directly interacting with transport and assembly mediators to HLA molecules [121]. The di-lysine KK that separates epitopes from BLs is located close to the C terminal. Among the adjuvants used were the EAAAK linkers: efficient separators between the domains present in the multi-epitope construct [122]. The following linkers were used in the vaccine, which contained 510 amino acids: 6 EAAK, 6 GPGPG, 11 AAY, and 5 KK. The proposed construct order is presented in Figure 4.
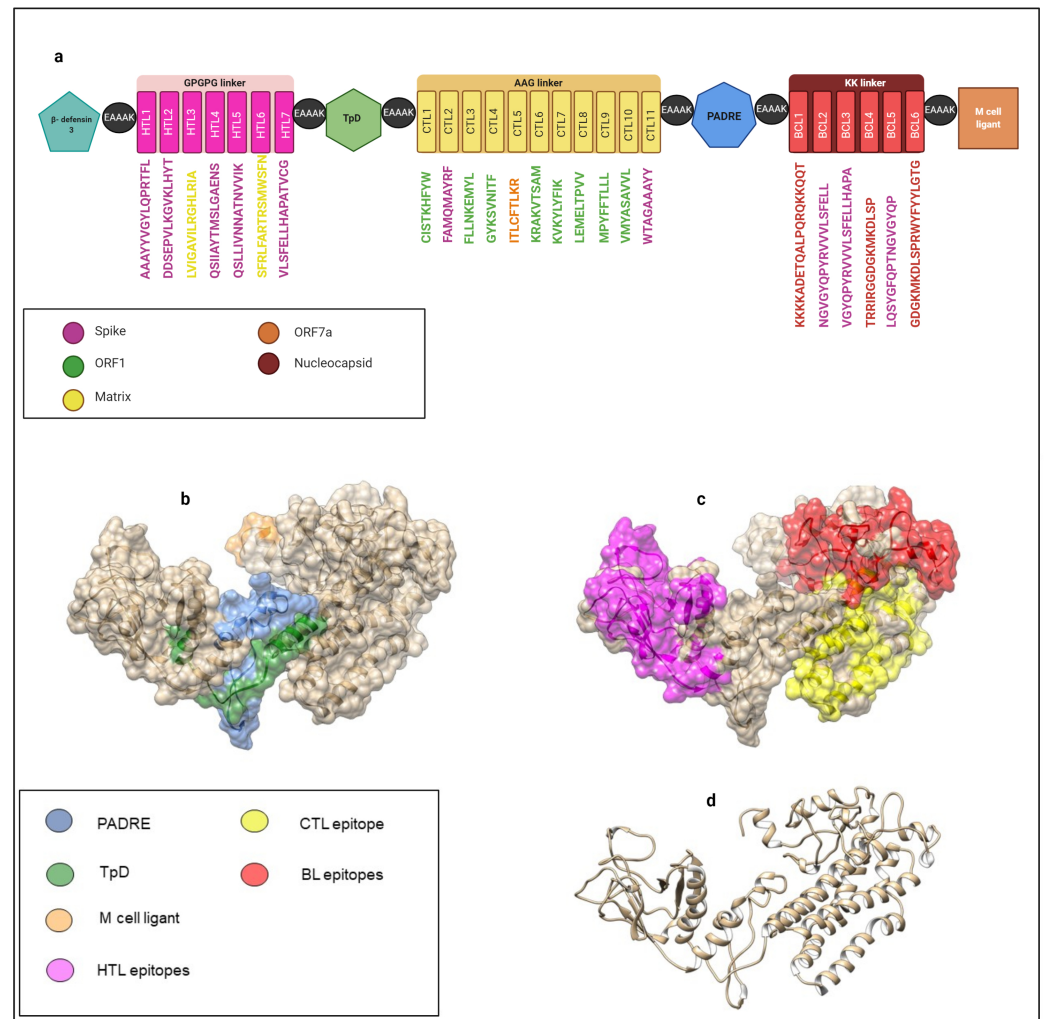
Figure 4. Construction of the Proposed Multi-Epitope Vaccine. Structure of the multi-epitope construct: three-dimensional diagrammatic structure of the multi-epitope construct showing (**a**), the separators, origin of the peptides and adjuvants used; (**b**), the location of the adjuvants PADRE, TpD and M cell ligand; and (**c**), the location of the epitope's cytotoxic T lymphocytes (CTLs), T helper lymphocytes (HTLs) and B lymphocytes (BLs). (**d**), Ribbon diagram of the multiepitope construct.

3.10. The Physicochemical Properties of the Multi-Epitope Construct Are Consistent with the Requirements for Generating an Immune Response in an Experimental Model

To generate a safe, stable, and capable construct able to evoke an immune response; the antigenic, non-allergenic, and non-toxic properties of the multi-epitope construct were established using Vaxijen 2.0, Allergen FP 1.0, and ToxinPred. The physicochemical characteristics of the multi-epitope construct, along with special consideration for the final CTGKSC peptide, are shown in Table 2.

Table 2. Physicochemical characteristics of the multi-epitope construct. The default thresholds for ToxinPred, Vaxijen > 0.4, and allergen FP were chosen to be the best balance between specificity, sensitivity, and precision while taking account of linear and non-linear reasons in the results. Besides, the physicochemical characteristics, construct, antigenic, non-allergenic, non-toxic, soluble, and stable characteristics are demonstrated suggesting that interaction with the immune system can allow and maintain recognition by the innate immune system, as well as by BL.

Peptide C Terminal	CTGKSC+	CTGKSC-
Vaxijen V2.0	0.6419	0.6455
AllerToP V2.0	Non-Allergenic	Non-Allergenic
Allergen FP V1.0	Non-Allergenic	Non-Allergenic
ToxinPred	Non-Toxic	Non-Toxic
	Physicochemical Parameters	
GRAVY	-0.038	-0.033
Molecular Weight	55.67 kDa	54.62 kDa
Stability	33.76 (<40)	33.47 (<40)
Solubility	72%	65%
Half-Life (reticulocytes)	30 h	30 h
Aliphatic index	81.81	82.62
Size (amino acids)	510	499

In addition to generating a safe construct, it was necessary to identify a thermostable multi-epitope construct, indicated by the aliphatic index for laboratory testing. Solubility and thermostability are associated with adequate overexpression in (*E. coli*) which is the bacteria most commonly used to produce recombinant proteins [123].

Given the parameters from the primary structure of the multi-epitope construct, adequate production in vitro can be inferred because of overexpression in *E. coli* and observed safety in immunological studies.

3.11. Three-Dimensional Structure Modeling and Validation of the Multi-Epitope Vaccine Construct

The predicted structural conformation of our construct can be correlated with functional annotation and other multi-epitope constructs. The secondary structure was analyzed using the SOPMA and PSIPRED servers, which revealed the presence of 43% α-helix, 21% β-foil, 30% coils, and 6% β- in the vaccine construction (Figures S7 and S8). A reliable approximation of the three-dimensional structure of the construct can be used for further in-depth studies. Molecular dynamics can be used to analyze coupling stability with immunological targets in innate recognition of viral elements as membrane TLR receptors with the production capacity of IFN-1. Therefore, we approximated the tertiary structure using the Robetta server, based on homology models, and we used the Ramachandran diagram and ERRAT to validate the quality of this approach. The results are summarized in (Figure S9).

By using the ClusPro 2.0 server to simulate molecular docking between the vaccine construct and TLR-4, 30 models were generated (Table S13). These were classified by cluster size according to their representative position. The lowest energy -1192.6 (Figure S10) was found in the sixth cluster with 22 members. The first cluster contained 35 members, which indicates an acceptable probability for the native pose of the complex. The cluster with a balanced adjustment was chosen given that it was closer than adjustments 2 and 3, which were related to a majority of hydrophobic and electrostatic interfaces [124].

This model positions the N terminal of the multi-epitope construct and larger interface area towards the concave side of TLR-4, where its ectodomain forms mostly hydrogen and salt bridges. The free convex side is broken into the N terminal of TLR-4, which indicates the formation of hydrogen and salt bridges resulting from interactions with C-terminal of the multi-epitope construct.

Also, interactions are present in the internal part and run towards MD-2; these are mostly hydrogen and saline bridges. The latter is maintained primarily by the ARG 157 and ARG 159 residues of the SFRLFARTRSMWSFN peptide. These are illustrated in (Figures S10–S12).

3.12. Molecular Dynamics

The root mean square deviation RMSD for the vaccine-receptor complex was evaluated during the complete simulation time (Figure 5a). The RMSD plot showed a considerable increase until it stabilized at around 20 ns, with an average of about 0.8 nm for the rest of the simulation time. These changes indicate that the vaccine attempts to find the best position based on its receptor. After 20 ns from the beginning of the simulation, it remained stable, which is an indicator that the vaccine reaches the best conformation to form a stable complex. Also, the root-mean-square-fluctuation (RMSF) of the vaccine and its receptor was evaluated. Residues from the vaccine start at GLY-1483 and finalize at residue CYS-1992 (Figure 5b). The RMSF showed a stable conformation from residue GLY-1483 to LYS1683 from the vaccine segment, which are part of the *β*-defensin 3 to TpD of the vaccine construct, Figure 4a. This stabilization is made due to the salt bridges formed in the residue ASP-1424:ARG1639, ASP-756:ARG-1641, GLU-1180:ARG-1524, GLU-1183:ARG1520, and GLU-1556:ARG930, which showed regular contact during the entire simulation (Figure S13). On the other hand, a segment of the vaccine, which corresponds to residue PHE-1684 to CYS-1992 (Figure 5b), showed a higher RMSF, displaying a pronounced motion compared with the GLY-1483 to LYS1683 segment, while the TLR-4 receptor remained stable. Apart from this motion, the complex remained stable during the 68 ns of simulation.

The radius of gyration (Rg) was calculated to determine the compactness of the vaccine-receptor complex system during the simulation (Figure 5c). It did not show a significant change. The complex exhibits a compact folded structure during the simulation, with an average Rg of approximately 4.8 nm. This folded complex structure is stabilized on average by around 9 H-bonds (Figure 5d) and 5 salt bridges, which were described above. Thus, the Vaccine-TLR4 complex showed a stable conformation during the 68 ns of simulation.

3.13. Optimization of cDNA from the Vaccine Construct for Optimal Expression of the Vaccine Product

For insertion of the vaccine construct into a plasmid vector, the CTGKSC+ protein sequence of 510 amino acids was inversely translated to a cDNA of 1530 nucleotides in length. The host system of expression varies, and the cDNA must adapt according to the use of the host codon. For optimal expression of the multi-epitope construct in the *E. coli* K 12 host, the resulting cDNA was codon-optimized according to the JCAT server. Furthermore, optimization of the rho-independent transcription terminator and prokaryotic ribosomal binding sites in the middle of the cDNA sequence were avoided to generate optimal and complete protein expression. To insert the construct into the pET28a(+) cloning vector, the BamHI and HindIII cleavage sites were also avoided. The results are shown in (Figure S14).
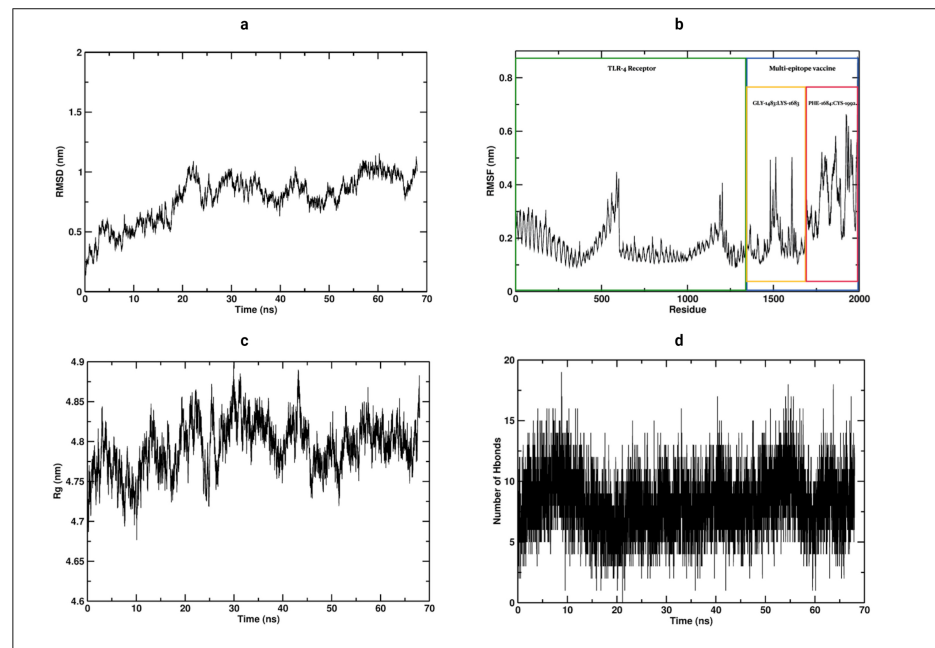
Figure 5. Analysis of trajectory. (**a**), Root mean square deviation (RMSD) Cα atoms. (**b**), Root mean square fluctuation for Cα atoms (RMSF), TLR-4 receptor initialize at amino acid 1 and goes to amino acid number 1482, and the vaccine from amino acid number 1483 to 1992. (**c**), Rg plot; vaccine construct is stable in its compact form during the simulation time. (**d**), Changes in the number of hydrogen bonds between the TLR-4 receptor and multi-epitope vaccine molecule during MD simulation.

3.14. Expression of the Multi-Epitope Vaccine Construct in E. coli K 12 by In Silico Cloning

The *E. coli* K 12 strain was selected as the cloning organism since multiple-epitope vaccines are expressed and purified more easily in this bacterium. For this purpose, the expression vector pET28a(+) was used and excised with the restriction enzymes BamHI and HindIII. The optimized cDNA was then inserted near the ribosome binding site using Snapgene (Figure S14).

3.15. Immune System Simulation

To identify the immunogenic profile of the vaccine construct, we used the C-IMMSIM immune server. As shown in Figure 6, the secondary and tertiary responses showed greater global responses and the presence of memory T and BLs. This may have been due to the cumulative effect of cells at the serum level that possesses a memory profile exceeding the total in injection 3. This is also associated with decreased antigenic concentrations and normal immunoglobulin levels. Cell-specific lineages were also stimulated, with the general activation of CD4+ and CD8+ TLs. Also, HTL-1 cell-based immunity was predominant. This is associated with the production of the cytokines and interleukins IFN-1, TNF-β and IL-2, as well as the activation of professional antigen-presenting cells.
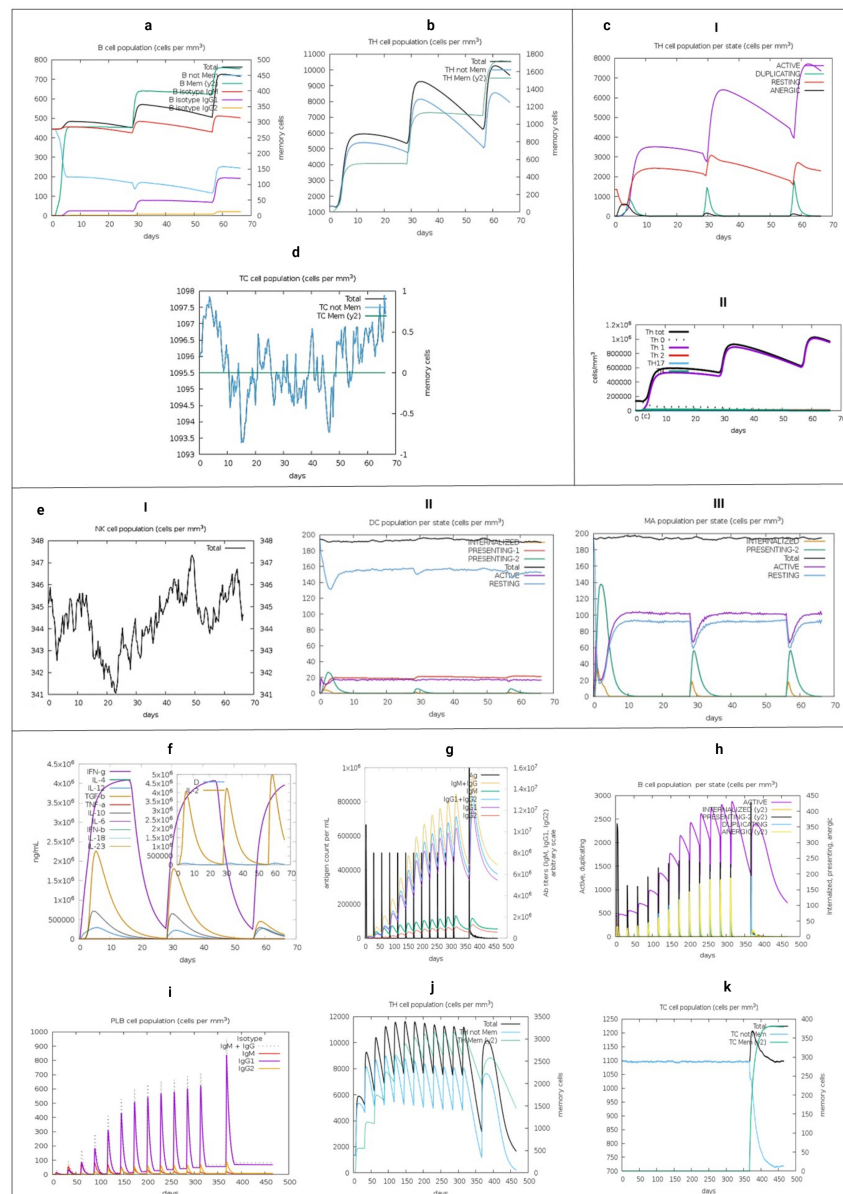
Figure 6. Immune simulation using the multi-epitope construct. (**a**), BLs, (**b**), CD4+ TLs and (**d**), CD8+ TLs were simulated, presenting a cumulative effect towards the third injection on the 56th day of simulation. This suggests the early presence of TLs memory and a change towards the immunoglobulin isotype, immunoglobulin G (IgG), being more predominant. While the antigenic stimulus lasts, there is a polarization towards a certain type of response (**c**), HTL-1, which is consistent with (**e**), the active antigenic presentation of professional antigen-presenting cells. This could be partly stimulated by other non-presenters, as well as the production of interleukins, such as (**f**), IFN γ, TGF-β and IL-2. An infection challenge, composed of a virus responding to the sequence of SARS-CoV-2 proteins covered by the multi-epitope construct, was simulated on day 366. (**g**), An indifference in immunoglobulin M (IgM) production and increased IgG response suggests favourable conditions for viral antibody clearance in the BLs. (**h**), The duplication and antigenic presentation of BLs correspond to the stimulated response by the simulated virus, which lasts beyond the viral challenge. (**i**), An IgG isotype shows a substantial response to the viral challenge, with subtype IgG1 most notably responding. (**j**), The CD4+ TLs population is globally stimulated; the memory response results in consolidation, which increases cell numbers and is still available over 100 days after the viral challenge. (**k**), The population of memory CD8+ cells is stimulated by the viral challenge, acting directly on viral clearance.

These results are consistent with the construct having a natural immunogenic capacity, which was applied without Lipopolysaccharide (LPS) in the simulation. A more extensive immune simulation corroborates the memory formation indicated in this brief approach. It was conducted until day 311 with 12 injections, in a step of time beyond 460 days, adjusting the intervals that did not surpass 4 weeks, 12, 94, 178, 262, 346, 430, 514, 598, 682, 766, 850, 934 (Figure S15).

4. Discussion

In this study, we used a reverse engineering approach to design a multi-epitope vaccine for SARS-CoV-2 based on the identification of PPVPs (Table 2) from the conserved regions of proteins in 92 SARS-CoV-2 proteomes. Besides, we added potential epitopes for BLs, adjuvants, and linkers seeking a specific immune response (Figure 5). By using predictors with a higher experimental correlation against HLA-I, we can describe possible considerations in LATAM populations according to the locus and most frequent alleles found (Figures S4 and S5). Our approach differs from other in silico vaccine models proposed for SARS-CoV-2 because we considered the total of structural, non-structural, and accessory proteins to identify potential vaccine candidates. This not only aims to neutralize the input [125] or most immunogenic proteins [126], but also those related to replication and immune modulation. This offers broader protection without making a large protein since we used the least number of peptides capable of covering all HLA alleles most frequently found in 18 LATAM countries. In Table 1, P1 to P7, which targets HTL, aims to stimulate LT and activate BL. P8 to P18 are PPVPs directed towards CTL that result from the cleavage of polyprotein 1a/b. P8 and P18 from NSP did not result in luminal subcellular locations resistant to post-translational modifications. In addition, P10, P14, and P18 obtained through the similar algorithms used in this study were already experimentally validated.

Other authors have previously identified peptides from SARS-CoV-2 using similar bioinformatic tools similar to ours and have classified possible vaccine targets such as P9, P7, P5, P18 (in SP), and P12 (in ORF7) [127–132]. The multi-epitope construct followed by PPVPs in the linear epitopes of SP and NP with the capability of interacting directly with the LB receptor is presented in (Table S10). In SP, linear epitopes were taken by consensus, occupying the RBD and lacking N-glycosylations, thereby attempting to disrupt ACE2 coupling by stimulating specific antibodies with neutralizing capacity. However, it has been identified that specific antibodies against RBD are not the only way to neutralize SARS-CoV-2 [133]. The experimental evidence available in the IEBD column of Table 2 indicates peptides with experimental validation in both SARS-CoV and SARS-CoV-2, far from RBD such as P2. This was complemented with experimental tests on BLs and HLA. Another peptide, P9, shares similar characteristics but is only partially conserved, delimiting only a potential immune region. These indicate that other regions are capable of stimulating LB, LT, and HLA shared by SARS-CoV-2 and SARS-CoV, far away from RBD, which means that P2 could generate antibodies towards SP outside of RBD. Recently, 61 monoclonal antibodies isolated in serum from patients with severe COVID-19 requiring mechanical ventilation have been identified; 9 antibodies exhibited a high neutralizing capacity against SARS-Cov-2 in vitro, 5 were directed at RBD, 3 against the N-terminal domain, and 2 against indeterminate regions of [133]. Other authors have confirmed the neutralizing capacity of directed antibodies towards the N-terminal when isolating it from convalescent COVID-19 patients [134]. Our study consists entirely of bioinformatic analysis, but some PPVPs resulted from experimental evidence. Some PPVPs that were not found in the experimental databases consulted were P3, P6, P8, P11, P12, P13, P15, P16, and P17. It is important to gather information on these PPVPs as potential vaccine targets because they share common immunogenic characteristics.

HTL with antigenic properties primarily comes from SP (n = 114), MG (n = 62), and NP (n = 23). In comparison, the results for CTL were ORF1a/b (n = 573), followed by SP (n = 63), and NP (n = 34). Most of these epitopes originated in the ORF1a/b transcript or

from MG and SP. Compared to other proteins like accessories or other structural proteins, these could harbour a greater number of conserved immunogenic regions because they were found in predictions with greater agglutinations against LATAM alleles.

In this study, we corroborated NP as a relevant vaccine target. We found areas in NP with predictions of linear and discontinuous epitopes longer than those of SP, resulting in a greater chance of antibody production. We also highlighted the immunogenicity of SP and identified a greater number of strong agglutination predictions in association with more frequent HLA-II molecules in LATAM. We emphasize that the majority were found outside the receptor-binding domain. In SP, an interval between residues 492 and 524 was found to contain several PPVPs associated with HLA-I and II molecules. This was also found in the most frequent phenotypes in LATAM, e.g., HLA-B*16:02. This is an important immunogenic region that is also associated with the linear epitopes of BLs. However, it is important to keep in mind that using the complete SP subunit could be more effective in a vaccine formulation. It should be noted that mutations in this immunodominant region could imply changes in antigenic recognition with loss of protective capacity [135]. Our multi-epitope construct offers a specific immune response with greater attention to stimulating innate immunity, especially against relevant proteins in the phase of infection. Reflecting on whether it can offer cellular protection, even if antigenic drift or natural selection or by any SARS-CoV-2 vaccine itself throughout the pandemic or after, it may exceed the protection offered by neutralizing antibodies against the RBD. Protection that is not expected at least directly, when using the recombinant SP protein present in current commercial formulations obtained from the Wuhan-Hu-1 sequence [136]. Multi-epitope constructs, including the one proposed here, if required, can be periodically updated specifically covering new conserved B or TLs epitopes that emerge in SARS-CoV-2 variants, avoiding a delayed elimination of infected cells. Since by not using all the SP protein, other epitopes of SP present in most variants can be raised, without falling into the effect of the original antigenic sin [136–138]. Although we studied the most frequent alleles in LATAM, we only have one SARS-CoV-2 proteome from Brazil in the sample of proteomes (this can be considered as a drawback of the study, because of the limited amount of data to date), that were selected for the in silico analysis described in the pipeline (Figure 1—Pipeline) and available in (Table S1). We used a conservative approach by eliminating the non-synonymous amino acid substitutions presented in the 92 proteomes collected between December 2019 and 11 March 2020, from other regions of the world, mainly Asia. These blocks of conserved non-redundant amino acids were used for immunoinformatics analysis to avoid reformulations of the multi-epitope construct. This is due to the uncertainty about certain variants that could become more frequent over time and occupy immunogenic regions of SARS-CoV-2 throughout the ongoing pandemic. According to GISAID statistics [139], a study in the US by Wang et al. [140], and in Uruguay by Elizondo et al. [141], for the last week of October 2020, there was no indication of an aa-changing SNP with a clear trend in PPVPs and for LB epitopes (Table S10) in the vaccine construct. Although the US and Uruguay had introductions of SARS-CoV-2 from different geographical areas of the old world [141,142], there is evidence of similarity between the aa-changing SNP with the global ranking exposed by GISAID. The approach used remains current and utilizes data from around the world despite coming from analysis with a reduced number of proteomes in an early stage of the pandemic [139]. As well as being the result of the analysis of a group of the most frequent HLA alleles in LATAM. We added catalogued safe adjuvants to increase immunogenicity and overcome the HLA polymorphic barrier, which is typically an obstacle to the development of epitope-based vaccines [143]. Some HLAs and supertypes have been associated with a lack of response, e.g., the oral rotavirus vaccine in infants [144]. PADRE has been shown to improve the immune response in human papillomavirus vaccines by generating a robust CD8+ response and in hepatitis B virus vaccines by improving the presentation of epitopes and thereby generating specific CTLs [145,146]. Human β-defensin 3, which activates human monocytes dependent on TLR1/2 [147], is considered an endogenous adjuvant as it induces the maturation of

Langerhans cells to dendritic cells and stimulates these cells to induce strong proliferation and IFN-γ production by CD4+ TLs [148]. This approach was previously used to construct bovine herpes 1 DNA vaccine, in which it increased the production of IFN-γ-dependent LT CD8+ [149]. We also used TpD as an adjuvant, which has previously been studied extensively in animal models and peripheral blood samples. This resulted in increased TLs and the robust generation of antibodies [118]. CTGKSC theoretically increases the water solubility of the vaccine proposal (Table 2) due to its amphiphilic properties [74] and it could better deliver the multi-epitope construct to TLR-4/MD-2 in the M cells of the Peyer's patches follicles using polymeric [74,150,151]. With this alternative delivery and validation of the vaccine proposal, a greater reach could be achieved to include remote populations with difficult access to medical assistance.

The final construct was catalogued as antigenic, non-allergenic, and non-toxic, which establishes it as a safe and powerful vaccine candidate against SARS-CoV-2 (Table 1). Furthermore, the adjuvants provided in the multiepitope construct allows for a synergistic effect with other types of adjuvants. Proposals for multi-epitope peptide vaccines similar to that of this study have been proposed for various infectious etiologies (DEN-2 [152], Hendra [153], Nipah [154], *Pseudomonas aeruginosa* [77], *Klebsiella pneumoniae* [155], *Plasmodium* spp. [156], and non-infectious (Kaposi's sarcoma [157]). For in vivo models, multi-epitope vaccines for influenza virus type A [158], Ebola virus [159], HPV-16 [160], and Uropathogenic *E. coli* [161] have been tested and have shown to be capable of stimulating the innate and adaptive immunity to ensure specificity and safety.

In SARS-CoV-2, the validation of immunogenic peptides obtained from structural, non-structural, and accessory proteins was carried out. This was based on a bioinformatic analysis, similar to ours, which demonstrated the activation of TL in the peripheral blood of convalescent Covid-19 patients at the University Hospital Tuebingen [162]. From these findings, the authors selected 8 peptides capable of interacting and activating TL-CD4, TL-CD8, and producing IFN-γ. These were incorporated together with the XS15 adjuvant, emulsified in MONTANIDE™ ISA 51 VG, which targets TLR1/2. This approach is currently in Phase 1 trials (NCT04546841) [163]. Another study in the preclinical phase focused on optimizing a vaccine for TL-CD8 using artificial intelligence and computational algorithms like the aforementioned study. The authors also used conserved regions of structural, nonstructural, and accessory proteins for the binding prediction to the HLA-A*02:01 allele, one of the most common worldwide. Using an in vivo methodology, the authors utilized an HLA-A*02:01 transgenic mouse, which indicated an activation and production of memory TL to 55 peptides coupled to the PADRE sequence together with the Montaigne adjuvant. In patients with a history of symptomatic, moderate, and severe COVID-19, there was at least one and a maximum of six months without evidence of infection. These patients were shown to activate peripheral blood monocytes towards these peptides [164].

Using an immune simulation with a heterozygous HLA adjustment, represented by some of the most frequently recovered LATAM alleles, we confirmed the immunogenic nature of the vaccine constructs. Without the extra adjuvant LPS, the vaccine constructs promoted an immune profile that indicated adequate antigenic presentation and the subsequent generation of immune memory in groups of T and BLs. Also, there was a polarized response towards HLT-1 and the stimulation of several immunoglobulins (IgG1+IgG2, IgM and, IgG+IgM) following the first injection. There was a robust response to the third injection, with evidence of the production of IFN-1 and IL-2 cytokines related to CD8+ lineage expansion. Comparative simulations using the C-IMMSIM algorithm have been conducted with experimentally validated peptides. Correlations with in vitro studies were found [165], which suggests that in vitro experiments with our construct may potentially evoke similar cellular behaviours.

These predictions are consistent with the analysis of flexible protein-peptide docking, which provides evidence of the adequate anchorage of the side chains of some residues of two promiscuous peptides located in the vicinity of the single groove of different HLAs. This results in adequate accommodation of the core sequence in HLA-I and II molecules. It

is associated with effective immune interaction, and therefore sufficient presentation of CD4+ and CD8+ TLs, that when is activated, trigger an immune response [67].

Our in-depth secondary structure analysis showed the predominance of α-helices, while the tertiary structure showed an optimal spatial arrangement of amino acids. The modelled structure was further improved which increased its general quality. When we performed a molecular coupling of TLR-4/MD-2 using ClusPro v2, the resultant interfaces were generally based on hydrogen and salt bridges supported primarily by TLR4. This fixes the multi-epitope construct towards the concave and terminal C region of its ectodomain, identifying a probable non-canonical interaction. At this location, some P6 residues of MG have more contact near the more hydrophobic region of MD-2, where coupling with LPS has been reported [166]. Therefore, this interaction could be stable and safe since it shows the hydrophobic pocket of empty MD-2 which is known to interact with traces of LPS and is associated with a fatal toxic immune response [167]. Other protein interactions that occupy TLR-4 could be safety-relevant. In COVID-19 an immunopathological process has been described that at least theoretically could be explained in part from the interaction of SP with TLR-4, increasing the production of inflammatory cytokines secondary to the formation of interfaces composed of antigenic residues of SP, interacting with external residues of TLR-4 [168]. In our multi-epitope construct, although we use antigenic peptides that come from SP, none of these antigenic residues is part of the interfaces described.

In nature, other non-canonical interactions have been described as in the case of the HIV Tat protein that results in an interface between the N-terminal and TLR-4/MD-2, showing greater stability in the resulting complex [169]. Also, other studies on larger and smaller viral proteins, including Tat [170,171], have identified an inhibitory competition of the binding site towards MD-2 by lipid A of LPS from *Rhodobacter sphaeroides*. This could affect the activation by disrupting the formation of a stable complex with TLR-4. However, the MD-2 interaction may not always be necessary for receptor activation. For example, in *E. coli*, the adhesive subunit FimH can evoke a more immunogenic profile against LPS by targeting TLR-4, which up-regulates the expression of HLA-I and II molecules [172]. Therefore, the stable and low entropy formation between the interfaces of the tetramer with potential ligands, including multi-epitope constructs, could activate it [173,174]. The position and situation that our multi-epitope construct occupies has been found in other vaccine models constructed using different methods. Therefore, it seems to be relevant to the stabilization of the multi-epitope construct [165]. In addition to the immunogenicity and stability identified in our molecular dynamics analysis, the physicochemical analysis of the construct revealed other desirable characteristics related to a safer profile. In particular, the construct was found to be nontoxic and nonallergenic. It was also thermally stable, which suggests that it would be suitable for the overexpression in the bacterial model *E. coli* K 12. A successful in silico cloning procedure is a cost-effective option that can be extrapolated to laboratories in LATAM countries. Based on the algorithms used in this study, some considerations regarding susceptibility and the expected clinical result should be used discreetly. We identified the HLA-C Locus as the highest loading capacity of antigenic peptides that are predicted to strongly binding towards the most frequent HLA-I alleles in LATAM from conserved regions of SARS-CoV-2 (Table S11). However, the HLA-C promoter lacks binding sites to the transcription factor NF-*κ β* [175], which is upregulated in the antiviral state. Therefore, it may not represent an accurate picture of an active infection. Although the C locus can present antigenic peptides and activate CTL, this locus is more in tune with the effector function of natural killers [176], which could regain importance in the activation of NK in COVID-19. However, a greater presence of antigenic peptides is necessary for its membrane expression, and its high capacity to recognize peptides from SARS-CoV-2 could be associated with less thymic selection. Interestingly, the HLA-B*35:01 allele, which was found in one of the HLA molecules capable of recognizing a greater number of peptides from SARS-CoV-2, was found more frequently in BRA-CHL-COL-CUB -NIC-PER (Table S11). It was recently recognized in a cohort of Spanish patients with moderate COVID-19, which did not have a fatal outcome [107]. According to the

country of origin, when compared to Locus C and A, locus B did not represent a group of alleles in LATAM with a higher carrying capacity of peptides. As a result, despite the country of origin, certain people will have a better outcome from SARS-CoV-2. It would be interesting to review BRA-CHL-COL-CRI-VEN in the future. Especially VEN, since it was found with alleles belonging to Locus A, which can recognize many peptides from SARS-CoV-2. Countries such as ARG-CUB-ECU-NIC have a lower capacity to recognize alleles from locus A. Additional attributes of HLA molecules found in HLA-B*35:01 include stability in acidic pH and alternate antigenic presentation [177] and allelic heterozygosity in patients with COVID-19, which could bring us closer to a precise treatment. This type of approach can better characterize populations at risk and allow a deeper understanding of the immunopathology caused by SARS-CoV-2. A collaborative initiative on the study of HLA and COVID-19 is currently available at (http://www.hlacovid19.org, accessed on 19 December 2020).

5. Conclusions

SARS-CoV-2 undoubtedly poses a substantial social challenge for all citizens of the world. Not only are its effects devastating to public health, but they have also set back the progress made over decades in other fields like economics, equity regarding social determinants in health and social management of other endemics outbreaks, especially in regions with emerging economies. By using bioinformatics tools and considering the immunological profiles of vulnerable and diverse regions in LATAM, the use of peptide binding predictors such as SARS-CoV-2 and HLA-I can be considered. This can be used as part of a precise treatment plan focused on characterizing people and populations with a greater risk of COVID-19 infection.

A vaccine is the best option for limiting the medium- and long-term effects of COVID-19 on the global population. We created a candidate multi-epitope vaccine directed towards the world population, with an immunogenic capacity that surpassed the safety standards required to begin complementary experimental studies. Furthermore, we try to analyze the binding affinity of potential antigenic peptides resulting from our pipeline of most prevalent HLA molecules in LATAM as part of a regional immunoinformatics approach. Several PPVPs that resulted from this methodology were found using experimental validation and demonstrating others that share similar immunogenic characteristics and should be validated.

We hope to quickly validate the vaccine candidate and take action against the ongoing pandemic.

Table S7: Results of the potential binding between the most frequent HLA-I epitopes in LATAM. Table S8: Results of the linear and discontinuous BLs epitopes predictions from the experimental proteins SP and NP. Table S9: ABCpred server with lineal epitopes and immunogenic characteristics. Table S10: Characteristics of the pair of BL epitope candidates. Table S11: HLA-I molecules used in this study with the theoretical ability to bind different non-redundant peptides from conserved SARS-CoV-2 sequences. Table S12: Proportions of agglutination predictions found according to locus and country. Table S13: Cluspro model results.

Author Contributions: A.F.C. Conceptualization, data curation, formal analysis, methodology, project administration, data validation, bioinformatics computations, Writing—original draft, Writing—review and editing. L.L.D. Data curation, Writing—review and editing—Image design. A.F.A. Data curation, Writing—review and editing. M.K.P. Data curation, Writing—review and editing. Y.R.M. Conceptualization, data curation, formal analysis, methodology, project administration, data validation, bioinformatics computations, Writing—original draft, Writing—review and editing. D.C.C. Molecular simulation analysis, Writing—review and editing J.Y.R. Conceptualization, methodology, project administration, Molecular simulations, Writing—original draft, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The proteomes of 92 SARS-CoV-2 strains were downloaded from the Assembly database available at the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/assembly, accessed on 27 April 2020).

Conflicts of Interest: The authors declare that they have no competing interest.

References

1. Prompetchara, E.; Ketloy, C.; Palaga, T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pac. J. Allergy Immunol.* **2020**, *38*, 1–9. [PubMed]
2. Li, G.; Fan, Y.; Lai, Y.; Han, T.; Li, Z.; Zhou, P.; Pan, P.; Wang, W.; Hu, D.; Liu, X.; et al. Coronavirus infections and immune responses. *J. Med. Virol.* **2020**, *92*, 424–432. [CrossRef] [PubMed]
3. Jin, Y.; Yang, H.; Ji, W.; Wu, W.; Chen, S.; Zhang, W.; Duan, G. Virology, epidemiology, pathogenesis, and control of COVID-19. *Viruses* **2020**, *12*, 372. [CrossRef] [PubMed]
4. Adhikari, S.P.; Meng, S.; Wu, Y.J.; Mao, Y.P.; Ye, R.X.; Wang, Q.Z.; Sun, C.; Sylvia, S.; Rozelle, S.; Raat, H.; et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: A scoping review. *Infect. Dis. Poverty* **2020**, *9*, 29. [CrossRef]
5. Rodriguez-Morales, A.J.; Gallego, V.; Escalera-Antezana, J.P.; Méndez, C.A.; Zambrano, L.I.; Franco-Paredes, C.; Suárez, J.A.; Rodriguez-Enciso, H.D.; Balbin-Ramon, G.J.; Savio-Larriera, E.; et al. COVID-19 in Latin America: The implications of the first confirmed case in Brazil. *Travel Med. Infect. Dis.* **2020**, *35*, 101613. [CrossRef]
6. Rodriguez-Morales, A.J.; Rodriguez-Morales, A.G.; Méndez, C.A.; Hernández-Botero, S. Tracing new clinical manifestations in patients with Covid-19 in chile and its potential relationship with the SARS-CoV-2 divergence. *Curr. Trop. Med. Rep.* **2020**, 1–4. [CrossRef]
7. Millán-Oñate, J.; Millan, W.; Mendoza, L.A.; Sánchez, C.G.; Fernandez-Suarez, H.; Bonilla-Aldana, D.K.; Rodríguez-Morales, A.J. Successful recovery of COVID-19 pneumonia in a patient from Colombia after receiving chloroquine and clarithromycin. *Ann. Clin. Microbiol. Antimicrob.* **2020**, *19*, 16. [CrossRef]
8. The, L. The unfolding migrant crisis in Latin America. *Lancet* **2019**, *394*, 1966.
9. Mellan, T.A.; Hoeltgebaum, H.H.; Mishra, S.; Whittaker, C.; Schnekenberg, R.P.; Gandy, A.; Unwin, H.J.T.; Vollmer, M.A.; Coupland, H.; Hawryluk, I.; et al. Report 21: Estimating COVID-19 cases and reproduction number in Brazil. *medRxiv* **2020**. [CrossRef]
10. Ochoa, Y.C.; Sanchez, D.E.R.; Peñaloza, M.; Motta, H.F.C.; Méndez-Fandiño, Y.R. Effective Reproductive Number estimation for initial stage of COVID-19 pandemic in Latin American Countries. *Int. J. Infect. Dis.* **2020**. [CrossRef]
11. Ong, E.; Wong, M.U.; Huffman, A.; He, Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *BioRxiv* **2020**. [CrossRef]
12. Renu, K.; Prasanna, P.L.; Gopalakrishnan, A.V. Coronaviruses pathogenesis, comorbidities and multi-organ damage—A review. *Life Sci.* **2020**, 117839. [CrossRef]

13. Sánchez-Duque, J.; Arce-Villalobos, L.; Rodríguez-Morales, A. Enfermedad por coronavirus 2019 (COVID-19) en América Latina: Papel de la atención primaria en la preparación y respuesta. *Aten. Primaria Soc. Esp. Med. Fam. Comunitaria* **2020**, *52*, 369–372. [CrossRef]

14. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [CrossRef]

15. Wang, B.; Li, R.; Lu, Z.; Huang, Y. Does comorbidity increase the risk of patients with COVID-19: Evidence from meta-analysis. *Aging* **2020**, *12*, 6049. [CrossRef]

16. Totura, A.L.; Baric, R.S. SARS coronavirus pathogenesis: Host innate immune responses and viral antagonism of interferon. *Curr. Opin. Virol.* **2012**, *2*, 264–275. [CrossRef]

17. Chan, V.S.F.; Chan, K.Y.K.; Chen, Y.; Poon, L.L.M.; Cheung, A.N.Y.; Zheng, B.; Chan, K.H.; Mak, W.; Ngan, H.Y.S.; Xu, X.; et al. Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nat. Genet.* **2006**, *38*, 38–46. [CrossRef]

18. Wang, S.F.; Chen, K.H.; Chen, M.; Li, W.Y.; Chen, Y.J.; Tsao, C.H.; Yen, M.y.; Huang, J.C.; Chen, Y.M.A. Human-leukocyte antigen class I Cw 1502 and class II DR 0301 genotypes are associated with resistance to severe acute respiratory syndrome (SARS) infection. *Viral Immunol.* **2011**, *24*, 421–426. [CrossRef]

19. Tay, M.Z.; Poh, C.M.; Rénia, L.; MacAry, P.A.; Ng, L.F.P. The trinity of COVID-19: Immunity, inflammation and intervention. *Nat. Rev. Immunol.* **2020**, *20*, 363–374. [CrossRef]

20. Guo, Y.R.; Cao, Q.D.; Hong, Z.S.; Tan, Y.Y.; Chen, S.D.; Jin, H.J.; Tan, K.S.; Wang, D.Y.; Yan, Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—An update on the status. *Mil. Med. Res.* **2020**, *7*, 11. [CrossRef]

21. Song, Z.; Xu, Y.; Bao, L.; Zhang, L.; Yu, P.; Qu, Y.; Zhu, H.; Zhao, W.; Han, Y.; Qin, C. From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* **2019**, *11*, 59. [CrossRef] [PubMed]

22. Wu, A.; Peng, Y.; Huang, B.; Ding, X.; Wang, X.; Niu, P.; Meng, J.; Zhu, Z.; Zhang, Z.; Wang, J.; et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* **2020**, *27*, 325–328. [CrossRef] [PubMed]

23. Narayanan, K.; Huang, C.; Makino, S. SARS coronavirus accessory proteins. *Virus Res.* **2008**, *133*, 113–121. [CrossRef] [PubMed]

24. The RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19—Preliminary report. *N. Engl. J. Med.* **2020**, doi:10.1056/NEJMoa2021436. [CrossRef]

25. Beigel, J.H.; Tomashek, K.M.; Dodd, L.E.; Mehta, A.K.; Zingman, B.S.; Kalil, A.C.; Hohmann, E.; Chu, H.Y.; Luetkemeyer, A.; Kline, S.; et al. Remdesivir for the treatment of Covid-19. *N. Engl. J. Med.* **2020**, *383*, 1813–1826. [CrossRef] [PubMed]

26. Consortium, W.S.T. Repurposed antiviral drugs for COVID-19—Interim WHO SOLIDARITY trial results. *N. Engl. J. Med.* **2021**, *384*, 497–511. [CrossRef] [PubMed]

27. Food and Drug Administration. *Temporary Policy for Compounding of Certain Drugs for Hospitalized Patients by Outsourcing Facilities during the COVID-19 Public Health Emergency: Guidance for Industry*; Food and Drug Administration: Silver Spring, MD, USA, 2020.

28. Li, L.; Zhang, W.; Hu, Y.; Tong, X.; Zheng, S.; Yang, J.; Kong, Y.; Ren, L.; Wei, Q.; Mei, H.; et al. Effect of Convalescent Plasma Therapy on Time to Clinical Improvement in Patients With Severe and Life-threatening COVID-19: A Randomized Clinical Trial. *JAMA* **2020**, *324*, 460–470. [CrossRef]

29. Agarwal, A.; Mukherjee, A.; Kumar, G.; Chatterjee, P.; Bhatnagar, T.; Malhotra, P. Convalescent plasma in the management of moderate covid-19 in adults in India: Open label phase II multicentre randomised controlled trial (PLACID Trial). *BMJ* **2020**, *371*, m4232. [CrossRef]

30. Italian Medicine Agency. *COVID-19: Tsunami Study, Plasma does not Reduce the Risk of Respiratory Damage or Death*; Italian Medicine Agency: Rome, Italy, 2020.

31. Landi, L.; Ravaglia, C.; Russo, E.; Cataleta, P.; Fusari, M.; Boschi, A.; Giannarelli, D.; Facondini, F.; Valentini, I.; Panzini, I.; et al. Blockage of interleukin-1β with canakinumab in patients with Covid-19. *Sci. Rep.* **2020**, *10*, 21775. [CrossRef]

32. Lescure, F.X.; Honda, H.; Fowler, R.A.; Lazar, J.S.; Shi, G.; Wung, P.; Patel, N.; Hagino, O.; Bazzalo, I.J.; Casas, M.M.; et al. Sarilumab in patients admitted to hospital with severe or critical COVID-19: A randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Respir. Med.* **2021**. [CrossRef]

33. Rilinger, J.; Kern, W.V.; Duerschmied, D.; Supady, A.; Bode, C.; Staudacher, D.L.; Wengenmayer, T. A prospective, randomised, double blind placebo-controlled trial to evaluate the efficacy and safety of tocilizumab in patients with severe COVID-19 pneumonia (TOC-COVID): A structured summary of a study protocol for a randomised controlled trial. *Trials* **2020**, *21*, 470. [CrossRef]

34. Meng, F.; Xu, R.; Wang, S.; Xu, Z.; Zhang, C.; Li, Y.; Yang, T.; Shi, L.; Fu, J.; Jiang, T.; et al. Human umbilical cord-derived mesenchymal stem cell therapy in patients with COVID-19: a phase 1 clinical trial. *Signal Transduct. Target. Ther.* **2020**, *5*, 172. [CrossRef]

35. Ahn, D.G.; Shin, H.J.; Kim, M.H.; Lee, S.; Kim, H.S.; Myoung, J.; Kim, B.T.; Kim, S.J. Current status of epidemiology, diagnosis, therapeutics, and vaccines for novel coronavirus disease 2019 (COVID-19). *J. Microbiol. Biotechnol.* **2020**, *30*, 313–324. [CrossRef]

36. World Health Organization. *DRAFT landscape of COVID-19 candidate vaccines*; World Health Organization: Geneva, Switzerland, 2020.

37. Dhama, K.; Sharun, K.; Tiwari, R.; Dadar, M.; Malik, Y.S.; Singh, K.P.; Chaicumpa, W. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. *Hum. Vaccines Immunother.* **2020**, *16*, 1232–1238. [CrossRef]

38. Le, T.T.; Andreadakis, Z.; Kumar, A.; Roman, R.G.; Tollefsen, S.; Saville, M.; Mayhew, S. The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov.* **2020**, *19*, 305–306. [CrossRef]

39. Ng, O.W.; Chia, A.; Tan, A.T.; Jadi, R.S.; Leong, H.N.; Bertoletti, A.; Tan, Y.J. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* **2016**, *34*, 2008–2014. [CrossRef]

40. Gao, T.; Hu, M.; Zhang, X.; Li, H.; Zhu, L.; Liu, H.; Dong, Q.; Zhang, Z.; Wang, Z.; Hu, Y.; et al. Highly pathogenic coronavirus N protein aggravates lung injury by MASP-2-mediated complement over-activation. *MedRxiv* **2020**. [CrossRef]

41. Beltrame, M.H.; Boldt, A.B.; Catarino, S.J.; Mendes, H.C.; Boschmann, S.E.; Goeldner, I.; Messias-Reason, I. MBL-associated serine proteases (MASPs) and infectious diseases. *Mol. Immunol.* **2015**, *67*, 85–100. [CrossRef]

42. Yasui, F.; Kai, C.; Kitabatake, M.; Inoue, S.; Yoneda, M.; Yokochi, S.; Kase, R.; Sekiguchi, S.; Morita, K.; Hishima, T.; et al. Prior immunization with severe acute respiratory syndrome (SARS)-associated coronavirus (SARS-CoV) nucleocapsid protein causes severe pneumonia in mice infected with SARS-CoV. *J. Immunol.* **2008**, *181*, 6337–6348. [CrossRef]

43. Tilocca, B.; Soggiu, A.; Sanguinetti, M.; Musella, V.; Britti, D.; Bonizzi, L.; Urbani, A.; Roncada, P. Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect.* **2020**, *22*, 188–194. [CrossRef]

44. Wang, J.; Wen, J.; Li, J.; Yin, J.; Zhu, Q.; Wang, H.; Yang, Y.; Qin, E.; You, B.; Li, W.; et al. Assessment of immunoreactive synthetic peptides from the structural proteins of severe acute respiratory syndrome coronavirus. *Clin. Chem.* **2003**, *49*, 1989–1996. [CrossRef]

45. Jiang, H.; Xu, Y.; Li, L.; Weng, L.; Wang, Q.; Zhang, S.; Jia, B.; Hu, H.; He, Y.; Jacob, Y.; et al. Inhibition of influenza virus replication by constrained peptides targeting nucleoprotein. *Antivir. Chem. Chemother.* **2011**, *22*, 119–130. [CrossRef]

46. Wei, J.; Li, Z.; Yang, Y.; Ma, G.; Su, Z.; Zhang, S. An Apoferritin–Hemagglutinin Conjugate Vaccine with Encapsulated Nucleoprotein Antigen Peptide from Influenza Virus Confers Enhanced Cross Protection. *Bioconjugate Chem.* **2020**, *31*, 1948–1959. [CrossRef]

47. Oliveira, S.C.; de Magalhães, M.T.; Homan, E.J. Immunoinformatic Analysis of SARS-CoV-2 Nucleocapsid Protein and Identification of COVID-19 Vaccine Targets. *Front. Immunol.* **2020**, *11*, 2758. [CrossRef] [PubMed]

48. Liang, M.-F.; Du, R.-L.; Liu, J.-Z.; Li, C.; Zhang, Q.-F.; Han, L.-L.; Yu, J.-S.; Duan, S.-M.; Wang, X.-F.; Wu, K.-X.; et al. SARS patients-derived human recombinant antibodies to S and M proteins efficiently neutralize SARS-coronavirus infectivity. *Biomed. Environ. Sci.* **2005**, *18*, 363. [PubMed]

49. Iyer, A.S.; Jones, F.K.; Nodoushani, A.; Kelly, M.; Becker, M.; Slater, D.; Mills, R.; Teng, E.; Kamruzzaman, M.; Garcia-Beltran, W.F.; et al. Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Sci. Immunol.* **2020**, *5*, eabe0367. [CrossRef] [PubMed]

50. Ni, L.; Ye, F.; Cheng, M.L.; Feng, Y.; Deng, Y.Q.; Zhao, H.; Wei, P.; Ge, J.; Gou, M.; Li, X.; et al. Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. *Immunity* **2020**, *52*, 971–977.e3. [CrossRef]

51. Buchholz, U.J.; Bukreyev, A.; Yang, L.; Lamirande, E.W.; Murphy, B.R.; Subbarao, K.; Collins, P.L. Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9804–9809. [CrossRef]

52. Le Bert, N.; Tan, A.T.; Kunasegaran, K.; Tham, C.Y.; Hafezi, M.; Chia, A.; Chng, M.H.Y.; Lin, M.; Tan, N.; Linster, M.; et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **2020**, *584*, 457–462. [CrossRef]

53. Sariol, A.; Perlman, S. Lessons for COVID-19 immunity from other coronavirus infections. *Immunity* **2020**, *53*, 248–263. [CrossRef]

54. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinform.* **2019**, *20*, 1160–1166. [CrossRef]

55. Tableu. Academic Programs | Tableau Software, 2020. Available online: https://www.tableau.com/community/academic (accessed on 27 March 2021).

56. Robinson, J.; Halliwell, J.A.; Hayhurst, J.D.; Flicek, P.; Parham, P.; Marsh, S.G.E. The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Res.* **2015**, *43*, D423–D431. [CrossRef]

57. Manczinger, M.; Boross, G.; Kemény, L.; Müller, V.; Lenz, T.L.; Papp, B.; Pál, C. Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS Biol.* **2019**, *17*, e3000131. [CrossRef]

58. Niehrs, A.; Altfeld, M. Regulation of NK-cell function by HLA class II. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 55. [CrossRef]

59. Jensen, K.K.; Andreatta, M.; Marcatili, P.; Buus, S.; Greenbaum, J.A.; Yan, Z.; Sette, A.; Peters, B.; Nielsen, M. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **2018**, *154*, 394–406. [CrossRef]

60. Jurtz, V.; Paul, S.; Andreatta, M.; Marcatili, P.; Peters, B.; Nielsen, M. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **2017**, *199*, 3360–3368. [CrossRef]

61. O'Donnell, T.J.; Rubinsteyn, A.; Bonsack, M.; Riemer, A.B.; Laserson, U.; Hammerbacher, J. MHCflurry: Open-source class I MHC binding affinity prediction. *Cell Syst.* **2018**, *7*, 129–132. [CrossRef]

62. Calis, J.J.A.; Maybeno, M.; Greenbaum, J.A.; Weiskopf, D.; De Silva, A.D.; Sette, A.; Keşmir, C.; Peters, B. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput. Biol.* **2013**, *9*, e1003266. [CrossRef]

63. Berzofsky, J.A. Intrinsic and Extrinsic Factors in Protein Antigenic Structure. *Science* **1985**, *229*, 932–940. [CrossRef]

64. Doytchinova, I.A.; Flower, D.R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **2007**, *8*, 4. [CrossRef]

65. Dimitrov, I.; Bangov, I.; Flower, D.R.; Doytchinova, I. AllerTOP v. 2—A server for in silico prediction of allergens. *J. Mol. Model.* **2014**, *20*, 2278. [CrossRef]

66. Gupta, S.; Kapoor, P.; Chaudhary, K.; Gautam, A.; Kumar, R.; Open Source Drug Discovery Consortium; Raghava, G.P.S. In Silico approach for predicting toxicity of peptides and proteins. *PLoS ONE* **2013**, *8*, e73957. [CrossRef]

67. Liu, J.; Gao, G.F. Major histocompatibility complex: Interaction with peptides. *eLS* **2011**. [CrossRef]

68. Kurcinski, M.; Pawel Ciemny, M.; Oleniecki, T.; Kuriata, A.; Badaczewska-Dawid, A.E.; Kolinski, A.; Kmiecik, S. CABS- dock standalone: A toolbox for flexible protein-peptide docking. *Bioinformatics* **2019**, *35*, 4170–4172. [CrossRef]

69. Xue, L.C.; Rodrigues, J.P.; Kastritis, P.L.; Bonvin, A.M.; Vangone, A. PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **2016**, *32*, 3676–3678. [CrossRef] [PubMed]

70. Saha, S.; Raghava, G.P.S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins Struct. Funct. Bioinform.* **2006**, *65*, 40–48. [CrossRef] [PubMed]

71. Kringelum, J.V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS Comput. Biol.* **2012**, *8*, e1002829. [CrossRef] [PubMed]

72. Saha, S.; Raghava, G.P.S. AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* **2006**, *34*, W202–W209. [CrossRef] [PubMed]

73. Gupta, R.; Jung, E.; Brunak, S. Prediction of N-glycosylation sites in human proteins. *Preparation* **2004**, *46*, 203–206.

74. Fievez, V.; Plapied, L.; Plaideau, C.; Legendre, D.; des Rieux, A.; Pourcelle, V.; Freichels, H.; Jérôme, C.; Marchand, J.; Préat, V.; et al. In Vitro identification of targeting ligands of human M cells by phage display. *Int. J. Pharm.* **2010**, *394*, 35–42. [CrossRef]

75. Dimitrov, I.; Naneva, L.; Doytchinova, I.; Bangov, I. AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics* **2014**, *30*, 846–851. [CrossRef]

76. Walker, J.M. *The Proteomics Protocols Handbook*; Springer: Berlin/Heidelberg, Germany, 2005.

77. Solanki, V.; Tiwari, M.; Tiwari, V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. *Sci. Rep.* **2019**, *9*, 5240. [CrossRef]

78. Geourjon, C.; Deleage, G. SOPM: A self-optimized method for protein secondary structure prediction. *Protein Eng. Des. Sel.* **1994**, *7*, 157–164. [CrossRef]

79. Buchan, D.W.A.; Jones, D.T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **2019**, *47*, W402–W407. [CrossRef]

80. Song, Y.; DiMaio, F.; Wang, R.Y.R.; Kim, D.; Miles, C.; Brunette, T.J.; Thompson, J.; Baker, D. High-resolution comparative modeling with RosettaCM. *Structure* **2013**, *21*, 1735–1742. [CrossRef]

81. Laskowski, R.A.; Jabłońska, J.; Pravda, L.; Vařeková, R.S.; Thornton, J.M. PDBsum: Structural summaries of PDB entries. *Protein Sci.* **2018**, *27*, 129–134. [CrossRef]

82. Colovos, C.; Yeates, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **1993**, *2*, 1511–1519. [CrossRef]

83. Bhattacharya, D.; Nowotny, J.; Cao, R.; Cheng, J. 3Drefine: An interactive web server for efficient protein structure refinement. *Nucleic Acids Res.* **2016**, *44*, W406–W409. [CrossRef]

84. Heo, L.; Park, H.; Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* **2013**, *41*, W384–W388. [CrossRef]

85. Kozakov, D.; Hall, D.R.; Xia, B.; Porter, K.A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein-protein docking. *Nat. Protoc.* **2017**, *12*, 255. [CrossRef]

86. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]

87. Case, D.A.; Darden, T.A.; Cheatham, T.E., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.C.; Zhang, W.; Merz, K.M.; et al. *AmberTools 16*; University of California: San Francisco, CA, USA, 2016.

88. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef] [PubMed]

89. Jorgensen, W.L. I. Am. Chem. Sot. 106 (1984) 6638. WL Jorgensen, J. Chandrasekhar, JD Madura, RW Impey and ML Klein. *J. Chem. Phys.* **1983**, *79*, 926. [CrossRef]

90. Harvey, M.J.; Giupponi, G.; Fabritiis, G.D. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639. [CrossRef]

91. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]

92. Le Grand, S.; Götz, A.W.; Walker, R.C. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380. [CrossRef]

93. Roe, D.R.; Cheatham, T.E., III. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [CrossRef]

94. Rosano, G.L.; Ceccarelli, E.A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **2014**, *5*, 172. [CrossRef]

95. Grote, A.; Hiller, K.; Scheer, M.; Münch, R.; Nörtemann, B.; Hempel, D.C.; Jahn, D. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **2005**, *33*, W526–W531. [CrossRef]

96.   Kanduc, D.; Shoenfeld, Y. On the molecular determinants the SARS-CoV-2 attack. *Clin. Immunol.* **2020**, *215*, 108426. [CrossRef]

97.   Olejnik, J.; Hume, A.J.; Mühlberger, E. Toll-like receptor 4 in acute viral infection: Too much of a good thing. *PLoS Pathog.* **2018**, *14*, e1007390. [CrossRef]

98.   DeDiego, M.L.; Nieto-Torres, J.L.; Jimenez-Guardeño, J.M.; Regla-Nava, J.A.; Castaño-Rodriguez, C.; Fernandez-Delgado, R.; Usera, F.; Enjuanes, L. Coronavirus virulence genes with main focus on SARS-CoV envelope gene. *Virus Res.* **2014**, *194*. [CrossRef]

99.   Banerjee, A.K.; Blanco, M.R.; Bruce, E.A.; Honson, D.D.; Chen, L.M.; Chow, A.; Bhat, P.; Ollikainen, N.; Quinodoz, S.A.; Loney, C.; et al. SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell* **2020**, *183*, 1325–1339.e21. [CrossRef]

100. Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.; Bhattacharya, A.; Schulz, L.; Widera, M.; Mehdipour, A.R.; Tascher, G.; et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020**, *587*, 657–662. [CrossRef]

101. Durfee, L.A.; Lyon, N.; Seo, K.; Huibregtse, J.M. The ISG15 conjugation system broadly targets newly synthesized proteins: Implications for the antiviral function of ISG15. *Mol. Cell* **2010**, *38*, 722–732. [CrossRef]

102. Neuman, B.W.; Joseph, J.S.; Saikatendu, K.S.; Serrano, P.; Chatterjee, A.; Johnson, M.A.; Liao, L.; Klaus, J.P.; Yates, J.R.; Wüthrich, K.; et al. Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* **2008**, *82*, 5279–5294. [CrossRef]

103. Blum, J.S.; Wearsch, P.A.; Cresswell, P. Pathways of antigen processing. *Annu. Rev. Immunol.* **2013**, *31*, 443–473. [CrossRef]

104. Blander, J.M. Coupling Toll-like receptor signaling with phagocytosis: potentiation of antigen presentation. *Trends Immunol.* **2007**, *28*, 19–25. [CrossRef]

105. Hegde, N.R.; Chevalier, M.S.; Johnson, D.C. Viral inhibition of MHC class II antigen presentation. *Trends Immunol.* **2003**, *24*, 278–285. [CrossRef]

106. Vina, M.A.F.; Hollenbach, J.A.; Lyke, K.E.; Sztein, M.B.; Maiers, M.; Klitz, W.; Cano, P.; Mack, S.; Single, R.; Brautbar, C.; et al. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. *Philos. Trans. R. Soc. Biol. Sci.* **2012**, *367*, 820–829. [CrossRef]

107. Iturrieta-Zuazo, I.; Rita, C.G.; García-Soidán, A.; de Malet Pintos-Fonseca, A.; Alonso-Alarcón, N.; Pariente-Rodríguez, R.; Tejeda-Velarde, A.; Serrano-Villar, S.; Castañer-Alabau, J.L.; Nieto-Gañán, I. Possible role of HLA class-I genotype in SARS-CoV-2 infection and progression: A pilot study in a cohort of Covid-19 Spanish patients. *Clin. Immunol.* **2020**, *219*, 108572. [CrossRef] [PubMed]

108. Pisanti, S.; Deelen, J.; Gallina, A.M.; Caputo, M.; Citro, M.; Abate, M.; Sacchi, N.; Vecchione, C.; Martinelli, R. Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of Covid-19. *J. Transl. Med.* **2020**, *18*, 352. [CrossRef] [PubMed]

109. Addetia, A.; Crawford, K.H.; Dingens, A.; Zhu, H.; Roychoudhury, P.; Huang, M.; Jerome, K.R.; Bloom, J.D.; Greninger, A. Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with high attack rate. *medRxiv* **2020**. [CrossRef] [PubMed]

110. Ter Meulen, J.; Van Den Brink, E.N.; Poon, L.L.M.; Marissen, W.E.; Leung, C.S.W.; Cox, F.; Cheung, C.Y.; Bakker, A.Q.; Bogaards, J.A.; Van Deventer, E.; et al. Human monoclonal antibody combination against SARS coronavirus: Synergy and coverage of escape mutants. *PLoS Med.* **2006**, *3*, e237. [CrossRef]

111. Watanabe, Y.; Allen, J.D.; Wrapp, D.; McLellan, J.S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369*, 330–333. [CrossRef]

112. Wolfert, M.A.; Boons, G.J. Adaptive immune activation: Glycosylation does matter. *Nat. Chem. Biol.* **2013**, *9*, 776–784. [CrossRef]

113. Lipinski, T.; Fitieh, A.; St Pierre, J.; Ostergaard, H.L.; Bundle, D.R.; Touret, N. Enhanced immunogenicity of a tricomponent mannan tetanus toxoid conjugate vaccine targeted to dendritic cells via Dectin-1 by incorporating *β* -glucan. *J. Immunol.* **2013**, *190*, 4116–4128. [CrossRef]

114. Hanisch, F.G.; Ninkovic, T. Immunology of O-glycosylated proteins: approaches to the design of a MUC1 glycopeptide-based tumor vaccine. *Curr. Protein Pept. Sci.* **2006**, *7*, 307–315. [CrossRef]

115. Roulois, D.; Grégoire, M.; Fonteneau, J.F. MUC1-specific cytotoxic T lymphocytes in cancer therapy: Induction and challenge. *BioMed Res. Int.* **2013**, *2013*, 871936. [CrossRef] [PubMed]

116. Howell, M.D.; Streib, J.E.; Leung, D.Y.M. Antiviral activity of human beta-defensin 3 against vaccinia virus. *J. Allergy Clin. Immunol.* **2007**, *119*, 1022–1025. [CrossRef] [PubMed]

117. Chen, X.; Zaro, J.L.; Shen, W.C. Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* **2013**, *65*, 1357–1369. [CrossRef] [PubMed]

118. Fraser, C.C.; Altreuter, D.H.; Ilyinskii, P.; Pittet, L.; LaMothe, R.A.; Keegan, M.; Johnston, L.; Kishimoto, T.K. Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. *Vaccine* **2014**, *32*, 2896–2903. [CrossRef] [PubMed]

119. Fast, E.; Chen, B. Potential T-cell and B-cell Epitopes of 2019-nCoV. *bioRxiv* **2020**. [CrossRef]

120. Ip, P.P.; Boerma, A.; Regts, J.; Meijerhof, T.; Wilschut, J.; Nijman, H.W.; Daemen, T. Alphavirus-based vaccines encoding nonstructural proteins of hepatitis C virus induce robust and protective T-cell responses. *Mol. Ther.* **2014**, *22*, 881–890. [CrossRef]

121. Dorosti, H.; Eslami, M.; Negahdaripour, M.; Ghoshoon, M.B.; Gholami, A.; Heidari, R.; Dehshahri, A.; Erfani, N.; Nezafat, N.; Ghasemi, Y. Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. *J. Biomol. Struct. Dyn.* **2019**, *37*, 3524–3535. [CrossRef]

122. Arai, R.; Ueda, H.; Kitayama, A.; Kamiya, N.; Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* **2001**, *14*, 529–532. [CrossRef]

123. Idicula-Thomas, S.; Balaji, P.V. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. *Protein Sci.* **2005**, *14*, 582–592. [CrossRef]

124. Vajda, S.; Yueh, C.; Beglov, D.; Bohnuud, T.; Mottarella, S.E.; Xia, B.; Hall, D.R.; Kozakov, D. New additions to the C lus P ro server motivated by CAPRI. *Proteins Struct. Funct. Bioinform.* **2017**, *85*, 435–444. [CrossRef]

125. Tamalika, K.; Utkarsh, N.; Srijita, B.; Debashrito, D.; Filippo, C.; Mueller, D.M.; Srivastava, A.P. A candidate multi-epitope vaccine against SARS-CoV-2. *Sci. Rep.* **2020**, *10*, 10895.

126. Enayatkhani, M.; Hasaniazad, M.; Faezi, S.; Guklani, H.; Davoodian, P.; Ahmadi, N.; Einakian, M.A.; Karmostaji, A.; Ahmadi, K. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *J. Biomol. Struct. Dyn.* **2020**, *39*, 2857–2872. [CrossRef]

127. Almofti, Y.A.; Abd-elrahman, K.A.; Gassmallah, S.A.E.; Salih, M.A. Multi Epitopes Vaccine Prediction against Severe Acute Respiratory Syndrome (SARS) Coronavirus Using Immunoinformatics Approaches. *Am. J. Microbiol. Res.* **2018**, *6*, 94–114. [CrossRef]

128. Rakib, A.; Sami, S.A.; Mimi, N.J.; Chowdhury, M.M.; Eva, T.A.; Nainu, F.; Paul, A.; Shahriar, A.; Tareq, A.M.; Emon, N.U.; et al. Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. *Comput. Biol. Med.* **2020**, 103967. [CrossRef] [PubMed]

129. Abraham Peele, K.; Srihansa, T.; Krupanidhi, S.; Ayyagari, V.S.; Venkateswarulu, T.C. Design of multi-epitope vaccine candidate against SARS-CoV-2: A In-Silico study. *J. Biomol. Struct. Dyn.* **2020**, 1–10. [CrossRef]

130. Su, Q.d.; Yi, Y.; Zou, Y.n.; Jia, Z.y.; Qiu, F.; Wang, F.; Yin, W.j.; Zhou, W.t.; Zhang, S.; Yu, P.c.; et al. The biological characteristics of SARS-CoV-2 Spike protein Pro330-Leu650. *Vaccine* **2020**, *38*, 5071–5075. [CrossRef]

131. Mitra, D.; Shekhar, N.; Pandey, J.; Jain, A.; Swaroop, S. Multi-epitope based peptide vaccine design against SARS-CoV-2 using its spike protein. *bioRxiv* **2020**. [CrossRef]

132. Joshi, A.; Joshi, B.C.; Mannan, M.A.u.; Kaushik, V. Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. *Inform. Med. Unlocked* **2020**, *19*, 100338. [CrossRef]

133. Liu, L.; Wang, P.; Nair, M.S.; Yu, J.; Rapp, M.; Wang, Q.; Luo, Y.; Chan, J.F.W.; Sahi, V.; Figueroa, A.; et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **2020**, *584*, 450–456. [CrossRef]

134. Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **2020**, *369*, 650–655. [CrossRef]

135. Wang, Y.D.; Sin, W.Y.F.; Xu, G.B.; Yang, H.H.; Wong, T.y.; Pang, X.W.; He, X.Y.; Zhang, H.G.; Ng, J.N.L.; Cheng, C.S.S.; et al. T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. *J. Virol.* **2004**, *78*, 5612–5618. [CrossRef]

136. Williams, T.C.; Burgers, W.A. SARS-CoV-2 evolution and vaccines: cause for concern? *Lancet Respir. Med.* **2021**, *9*, 333–335. [CrossRef]

137. Chakradhar, S. Updated, augmented vaccines compete with original antigenic sin. *Nat. Med.* **2015**, *21*, 540–541. [CrossRef] [PubMed]

138. Tavasolian, F.; Rashidi, M.; Hatam, G.R.; Jeddi, M.; Hosseini, A.Z.; Mosawi, S.H.; Abdollahi, E.; Inman, R.D. HLA, Immune Response, and Susceptibility to COVID-19. *Front. Immunol.* **2021**, *11*, 3581. [CrossRef] [PubMed]

139. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46. [CrossRef] [PubMed]

140. Wang, R.; Chen, J.; Gao, K.; Hozumi, Y.; Yin, C.; Wei, G.W. Characterizing SARS-CoV-2 mutations in the United States. *arXiv* **2020**, arXiv:2007.12692.

141. Elizondo, V.; Harkins, G.W.; Mabvakure, B.; Smidt, S.; Zappile, P.; Marier, C.; Maurano, M.; Perez, V.; Mazza, N.; Beloso, C.; et al. SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. *Medrxiv* **2020**, *10*, 51–65.

142. Koyama, T.; Platt, D.; Parida, L. Variant analysis of SARS-CoV-2 genomes. *Bull. World Health Organ.* **2020**, *98*, 495. [CrossRef]

143. Zhao, L.; Zhang, M.; Cong, H. Advances in the study of HLA-restricted epitope vaccines. *Hum. Vaccines Immunother.* **2013**, *9*, 2566–2577. [CrossRef]

144. Liu, Y.; Guo, T.; Yu, Q.; Zhang, H.; Du, J.; Zhang, Y.; Xia, S.; Yang, H.; Li, Q. Association of human leukocyte antigen alleles and supertypes with immunogenicity of oral rotavirus vaccine given to infants in China. *Medicine* **2018**, *97*, e12706. [CrossRef]

145. Wu, C.Y.; Monie, A.; Pang, X.; Hung, C.F.; Wu, T.C. Improving therapeutic HPV peptide-based vaccine potency by enhancing CD4+ T help and dendritic cell activation. *J. Biomed. Sci.* **2010**, *17*, 88. [CrossRef]

146. Wang, H.; Su, X.; Zhang, P.; Liang, J.; Wei, H.; Wan, M.; Wu, X.; Yu, Y.; Wang, L. Recombinant heat shock protein 65 carrying PADRE and HBV epitopes activates dendritic cells and elicits HBV-specific CTL responses. *Vaccine* **2011**, *29*, 2328–2335. [CrossRef]

147. Funderburg, N.; Lederman, M.M.; Feng, Z.; Drage, M.G.; Jadlowsky, J.; Harding, C.V.; Weinberg, A.; Sieg, S.F. Human *β*-defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18631–18635. [CrossRef]

148. Ferris, L.K.; Mburu, Y.K.; Mathers, A.R.; Fluharty, E.R.; Larregina, A.T.; Ferris, R.L.; Falo, L.D., Jr. Human beta-defensin 3 induces maturation of human Langerhans cell-like dendritic cells: an antimicrobial peptide that functions as an endogenous adjuvant. *J. Investig. Dermatol.* **2013**, *133*, 460–468. [CrossRef]

149. Mackenzie-Dyck, S.; Kovacs-Nolan, J.; Snider, M.; Babiuk, L.A.; van Drunen Littel-van den Hurk, S. Inclusion of the bovine neutrophil beta-defensin 3 with glycoprotein D of bovine herpesvirus 1 in a DNA vaccine modulates immune responses of mice and cattle. *Clin. Vaccine Immunol.* **2014**, *21*, 463–477. [CrossRef]

150. Pridgen, E.M.; Alexis, F.; Farokhzad, O.C. Polymeric nanoparticle drug delivery technologies for oral delivery applications. *Expert Opin. Drug Deliv.* **2015**, *12*, 1459–1473. [CrossRef]

151. Wang, M.; Gao, Z.; Zhang, Z.; Pan, L.; Zhang, Y. Roles of M cells in infection and mucosal vaccines. *Hum. Vaccines Immunother.* **2014**, *10*, 3544–3551. [CrossRef]

152. Ali, M.; Pandey, R.K.; Khatoon, N.; Narula, A.; Mishra, A.; Prajapati, V.K. Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Sci. Rep.* **2017**, *7*, 9232. [CrossRef]

153. Kamthania, M.; Srivastava, S.; Desai, M.; Jain, A.; Shrivastav, A.; Sharma, D. Immunoinformatics Approach to Design T-cell Epitope-Based Vaccine Against Hendra Virus. *Int. J. Pept. Res. Ther.* **2019**, *25*, 1627–1637. [CrossRef]

154. Ojha, R.; Pareek, A.; Pandey, R.K.; Prusty, D.; Prajapati, V.K. Strategic development of a next-generation multi-epitope vaccine to prevent Nipah virus zoonotic infection. *ACS Omega* **2019**, *4*, 13069–13079. [CrossRef]

155. Dar, H.A.; Zaheer, T.; Shehroz, M.; Ullah, N.; Naz, K.; Muhammad, S.A.; Zhang, T.; Ali, A. Immunoinformatics-Aided Design and Evaluation of a Potential Multi-Epitope Vaccine against Klebsiella Pneumoniae. *Vaccines* **2019**, *7*, 88. [CrossRef]

156. Pandey, R.K.; Bhatt, T.K.; Prajapati, V.K. Novel immunoinformatics approaches to design multi-epitope subunit vaccine for malaria by investigating anopheles salivary protein. *Sci. Rep.* **2018**, *8*, 1125. [CrossRef]

157. Chauhan, V.; Rungta, T.; Goyal, K.; Singh, M.P. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach. *Sci. Rep.* **2019**, *9*, 2517. [CrossRef] [PubMed]

158. Romeli, S.; Hassan, S.S.; Yap, W.B. Multi-Epitope Peptide-Based and Vaccinia-Based Universal Influenza Vaccine Candidates Subjected to Clinical Trials. *Malays. J. Med. Sci. MJMS* **2020**, *27*, 10.

159. Bazhan, S.I.; Antonets, D.V.; Karpenko, L.I.; Oreshkova, S.F.; Kaplina, O.N.; Starostina, E.V.; Dudko, S.G.; Fedotova, S.A.; Ilyichev, A.A. In silico designed ebola virus T-cell multi-epitope DNA vaccine constructions are immunogenic in mice. *Vaccines* **2019**, *7*, 34. [CrossRef] [PubMed]

160. De Oliveira, L.M.F.; Morale, M.G.; Chaves, A.A.M.; Cavalher, A.M.; Lopes, A.S.; de Oliveira Diniz, M.; Schanoski, A.S.; de Melo, R.L.; de Souza Ferreira, L.C.; de Oliveira, M.L.S.; et al. Design, immune responses and anti-tumor potential of an HPV16 E6E7 multi-epitope vaccine. *PLoS ONE* **2015**, *10*, e0138686. [CrossRef]

161. Hasanzadeh, S.; Habibi, M.; Shokrgozar, M.A.; Cohan, R.A.; Ahmadi, K.; Karam, M.R.A.; Bouzari, S. In Silico analysis and in vivo assessment of a novel epitope-based vaccine candidate against uropathogenic Escherichia coli. *Sci. Rep.* **2020**, *10*, 16258. [CrossRef]

162. Nelde, A.; Bilich, T.; Heitmann, J.S.; Maringer, Y.; Salih, H.R.; Roerden, M.; Lübke, M.; Bauer, J.; Rieth, J.; Wacker, M.; et al. SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. *Nat. Immunol.* **2020**, *22*, 1–12. [CrossRef]

163. Ascarateil, S.; Puget, A.; Koziol, M.E. Safety data of Montanide ISA 51 VG and Montanide ISA 720 VG, two adjuvants dedicated to human therapeutic vaccines. *J. Immunother. Cancer* **2015**, *3*. [CrossRef]

164. Gauttier, V.; Morello, A.; Girault, I.; Mary, C.; Belarif, L.; Desselle, A.; Wilhelm, E.; Bourquard, T.; Pengam, S.; Teppaz, G.; et al. Tissue-resident memory CD8 T-cell responses elicited by a single injection of a multi-target COVID-19 vaccine. *bioRxiv* **2020**. [CrossRef]

165. Khan, M.A.A.; Ami, J.Q.; Faisal, K.; Chowdhury, R.; Ghosh, P.; Hossain, F.; Abd El Wahed, A.; Mondal, D. An immunoinformatic approach driven by experimental proteomics: in silico design of a subunit candidate vaccine targeting secretory proteins of Leishmania donovani amastigotes. *Parasites Vectors* **2020**, *13*, 196. [CrossRef]

166. Maeshima, N.; Fernandez, R. Recognition of lipid A variants by the TLR4-MD-2 receptor complex. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 3. [CrossRef]

167. Park, B.S.; Lee, J.O. Recognition of lipopolysaccharide pattern by TLR4 complexes. *Exp. Mol. Med.* **2013**, *45*, e66. [CrossRef] [PubMed]

168. Choudhury, A.; Mukherjee, S. In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs. *J. Med. Virol.* **2020**, *92*, 2105–2113. [CrossRef]

169. Haij, N.B.; Leghmari, K.; Planès, R.; Thieblemont, N.; Bahraoui, E. HIV-1 Tat protein binds to TLR4-MD2 and signals to induce TNF-*α* and IL-10. *Retrovirology* **2013**, *10*, 123. [CrossRef] [PubMed]

170. Modhiran, N.; Watterson, D.; Muller, D.A.; Panetta, A.K.; Sester, D.P.; Liu, L.; Hume, D.A.; Stacey, K.J.; Young, P.R. Dengue virus NS1 protein activates cells via Toll-like receptor 4 and disrupts endothelial cell monolayer integrity. *Sci. Transl. Med.* **2015**, *7*, 304ra142. [CrossRef]

171. Rallabhandi, P.; Phillips, R.L.; Boukhvalova, M.S.; Pletneva, L.M.; Shirey, K.A.; Gioannini, T.L.; Weiss, J.P.; Chow, J.C.; Hawkins, L.D.; Vogel, S.N.; et al. Respiratory syncytial virus fusion protein-induced toll-like receptor 4 (TLR4) signaling is inhibited by the TLR4 antagonists Rhodobacter sphaeroides lipopolysaccharide and eritoran (E5564) and requires direct interaction with MD-2. *MBio* **2012**, *3*, e00218-12. [CrossRef]

172. Zhang, W.; Xu, L.; Park, H.B.; Hwang, J.; Kwak, M.; Lee, P.C.; Liang, G.; Zhang, X.; Xu, J.; Jin, J.O. Escherichia coli adhesion portion FimH functions as an adjuvant for cancer immunotherapy. *Nat. Commun.* **2020**, *11*, 1187. [CrossRef]

173. Piccinini, A.; Midwood, K. DAMPening inflammation by modulating TLR signalling. *Mediat. Inflamm.* **2010**, *2010*, 672395. [CrossRef]
174. Wang, Y.; Su, L.; Morin, M.D.; Jones, B.T.; Whitby, L.R.; Surakattula, M.M.; Huang, H.; Shi, H.; Choi, J.H.; Wang, K.w.; et al. TLR4/MD-2 activation by a synthetic agonist with no similarity to LPS. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E884–E893. [CrossRef]
175. Anderson, S.K. Molecular evolution of elements controlling HLA-C expression: Adaptation to a role as a killer-cell immunoglobulin-like receptor ligand regulating natural killer cell function. *Hla* **2018**, *92*, 271–278. [CrossRef]
176. Blais, M.E.; Dong, T.; Rowland-Jones, S. HLA-C as a mediator of natural killer and T-cell activation: spectator or key player? *Immunology* **2011**, *133*, 1–7. [CrossRef]
177. Tey, S.K.; Khanna, R. Autophagy mediates transporter associated with antigen processing-independent presentation of viral epitopes through MHC class I pathway. *Blood J. Am. Soc. Hematol.* **2012**, *120*, 994–1004. [CrossRef] [PubMed]