A Multi-Epitope Vaccine Against Sars-Cov-2 Directed Towards the Latin American Population: An Immunoinformatics Approach

Andrés F. Cuspoca
Universidad Pedagógica y Tecnológica de Colombia, Colombia https://orcid.org/0000-0002-2420-4898
Laura L. Díaz
Universidad Pedagógica y Tecnológica de Colombia, Colombia https://orcid.org/0000-0001-9030-6043
Alvaro F. Acosta
Universidad Pedagógica y Tecnológica de Colombia, Colombia https://orcid.org/0000-0003-0849-8248
Marcela K. Peñaloza
Universidad Pedagógica y Tecnológica de Colombia, Colombia https://orcid.org/0000-0002-2736-0538
Yardany R. Mendez
Universidad Pedagógica y Tecnológica de Colombia, Colombia https://orcid.org/0000-0003-1528-2672
Diana C. Clavijo
Pontificia Universidad Javeriana Cali, Colombia https://orcid.org/0000-0003-0941-4552
Juvenal Yosa Reyes (juvenal.yosa@unisimonbolivar.edu.co)
Universidad Simón Bolivar Barranquilla, Colombia https://orcid.org/0000-0002-6492-2640

Research Article

Keywords: SARS-CoV-2., HLA-I , HLA-II, Immunoinformatics, Molecular Simulation, Molecular Dynamics

Posted Date: September 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-70414/v1

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. 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 vaccine against SARS-CoV-2. Using bioinformatics, we constructed a library of potential vaccine peptides to predict immunological complexes among antigenic, non-toxic and non-allergenic peptides extracted from the conserved regions of 92 proteomes. We included the most common HLA-I and II molecules in the Latin American population. 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 optimised final construct that meets the necessary safety and physicochemical requirements for conducting experimental tests.

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 February 20, 2020, 19 countries had reported cases and mortalities caused by coronavirus disease 2019 (COVID-19). In March of 2020, the 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 had travelled to Italy[5]. Subsequently, cases were reported in other LATAM countries, including Chile[6] and Colombia[7]. 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, as well as mass migration[8]. Indeed, the effective reproductive number estimated for LATAM countries suggests that aggressive outbreak behaviour is likely[9].

To date, 24,088,692 COVID-19 cases and 824,368 deaths have been registered worldwide[10]. COVID-19 has a broad range of clinical manifestations including asymptomatic patients and those with acute respiratory distress syndrome leading to death. There is a greater risk of mortality in patients with comorbidities such as arterial hypertension, chronic obstructive pulmonary disorder, diabetes, and vascular diseases, especially cerebrovascular disease[4,11]. While there is no evidence to suggest that the disease generates immunological memory post-COVID-19[12], studies have shown that recovered individuals present antibodies against the virus. Although some do not present detectable levels of immunoglobulin G (IgG), they do present low levels of neutralising antibodies[13]

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 the expression of Toll-like receptor (TLR) 3 and 4 are associated with better outcomes in infected rats[14]. In addition, individuals with 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[15]. 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[16].

From these findings, we can infer that the preservation of IFN-1 production and some alleles of major human histocompatibility complex 1 and 2 (HLA-I and HLA-II) may be related to the asymptomatic state and the presentation of only mild symptoms in COVID-19[16]. In contrast, changes to IFN-1-producing signalling 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[17].

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[3,18]. Coronavirus genomes are composed of 6-11 open reading frames (ORFs)[19], 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), five 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[3,20,21]. The genome structure is shown in (Supplementary Fig. S1).

Medication for COVID-19

An effective treatment against COVID-19 is urgently required; thus, potential medications such as antiviral drugs that can inhibit viral protease[22], anti-malarial drugs that can inhibit the endosomal entrance of the virus[23], and corticoids that can interrupt the inflammation caused by this virus[24] have been explored. To date, however, no medication has demonstrated sufficient efficacy to be implemented as a definitive treatment[25]. Recently, Wang et al. discovered a new antibody (47D11) that binds the subunit S1 of the S gene and disables its union with the ACE2 receptor, thereby blocking coronavirus infection in in vitro models. Although it has not been sufficiently tested on humans to prove efficacy[26], this antibody has potential as a prophylactic and treatment.

Related vaccines

Previously, Channappanavar et al.[27] proposed the use of vaccines based on cytotoxic T lymphocytes (CTLs) epitopes in mice infected with SARS-CoV strains; these vaccines were associated with decreased mortality and an effective immune response via TL CD8+ in mice administered a lethal dose of the virus. In an alternative approach, Minghai et al.[28] studied the peptides contained in SARS-CoV NP and SP proteins, their binding motifs to HLA-A*02:01 molecules and their ability to stimulate CTLs. Thus, the SP-derived peptide (KLPDDFMGCV) induces an effective and specific reaction[28]. The MER-CoV SP binding domain has also been reported to generate long-term protective antibodies[29].

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[30]. According to a WHO report, vaccine studies for COVID-19 are using different vaccine strategies, such as non-replicating viral vectors, RNA, inactivated, and DNA approaches. To date, there there are more than 100 vaccine candidates in preclinical trials[31]. The main vaccine candidates are summarised in (Supplementary Table 1). In both SARS-CoV and SARS-CoV-2, several candidate structural proteins have been studied as vaccine targets. SP has been widely studied due to its capacity to induce neutralising antibodies that prevent the virus from binding and fusing with ACE2 receptor. Various vaccine models have been proposed using the complete protein, SP binding domain, virus-like particles, DNA or viral vectors[32,33]. In SARS-CoV, SP has been shown to play an important role in inducing immunity by stimulating the production of neutralising antibodies and the activation of TLs[32,33].

In SARS-CoV studies of the NP, it was found to be expressed in large amounts during infection. The NP is highly immunogenic and a potential vaccine target: it develops a memory response from TLs that remains for up to 11 years after infection[34]. Another important target is ORF3, which plays a crucial role in the viral assembly of SARS-CoV in conjunction with the MG and envelope proteins. These proteins participate in pathogenesis and provide important immunogenicity[35]. In convalescent patients with SARS, the response to structural proteins such as SP or NP has shown greater activity in peripheral blood monocytes. This causes a more dominant and lasting response by TLs than that of other structural proteins; this could be used to create a vaccine for SARS-CoV-2 if preserved[34].

We believe that the participation of research institutions in LATAM in the development of a universal vaccine is essential 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[5,36–38].

In LATAM, the outbreak has had negative sociocultural, economic and political effects. Everyone, but particularly the most vulnerable, would be affected by further outbreaks and a second wave. The construction of a vaccine is therefore imperative for reducing the health costs and short- and long-term morbimortality associated with COVID-19[5,36–38].

In the present study, by using an immunoinformatics approach, we propose a multi-epitope peptide vaccine model, which is based on peptide binding properties extracted from conserved regions of SARS-CoV-2 proteomes and the HLA-I and II alleles most frequently found in LATAM (Fig. 1).

Results

Recovery of SARS-CoV-2 proteomes

Proteomes from 92 isolates of SARS-CoV-2 were recovered in FASTA format using the GenBank database for reference. Metadata related to the collection date, city, host, and source of isolation were also downloaded. Most 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 March 11, 2020 (the SARS-CoV- 2/human/USA/PC00101P/2020 isolate). Most sequences came from the USA (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 (Supplementary Table 2).

Identification of structural and non-structural proteins, their mechanism of action in SARS-CoV-2, and their conservation in

proteomes

"The spike protein acts like a Trojan horse".

SP plays a fundamental role in SARS-CoV-2 infection because it mediates the entry of the virus into host cells through its S1 domain, which binds to 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 neutralisation by antibodies. The immune response through viral and RNA receptors is the first line of defence against SARS-CoV-2 infection, producing cytokines and IFN-1[39]. Some proteins encoded by the coronavirus genome are capable of interfering with the innate immune response from the early phase of infection, which affects various signalling pathways important in maintaining immunity[40] (Supplementary Fig. S2). Thus, identifying the mechanisms of action, characteristics of structural and 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. In particular, NSPs necessary for replication and produced at an early stage should be considered. Considering structural protein production is also important as it may be an essential means by which to cover the SARS-CoV-2 life cycle.

We considered the structural, non-structural and accessory proteins of 92 proteomes to identify conserved regions according to frequency and alignment. Our findings are shown in (Supplementary Table 3). The conserved sequence blocks were used in their entirety for the prediction of conserved epitopes.

Most frequently identified HLA-I and HLA-II alleles in LATAM

We identified 168 of the most frequent HLA-I alleles in 13 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 (Supplementary Table 4), the frequencies found between the HLA-I loci were A (n = 126), B (n = 264) and C (n = 64); those above the mean frequencies were A (n = 48), B (n = 91) and C (n = 27). For HLA-II, the allele and allele groups had the following order of frequency: DRB1*04, DRB1*13, DRB1*07, DRB1*11, DRB1*08: 02, DRB1*15, DRB1*08, DRB1*01, DRB1*03 and DRB1*16:02. However, all DRB1 alleles listed by NetMHC 3.2, spanning 16 countries, had above average frequencies (Supplementary Table 5). (Supplementary Fig. S3) shows the distribution of the main alleles and allelic groups by country according to HLA-I and HLA-II.

HLA systems and SARS-CoV-2 proteomes

HLA systems in humans are located in the most polymorphic region of the genome, which is related to continuous selection in its interaction with extracellular and intracellular pathogens. Furthermore, this diversity corresponds to rapid adaptation to environmental change, as well as records of social conditions. This has become a hallmark of various human populations after their first few migrations[41].

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 that develops memory. While this would only work in certain

populations as 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.

We propose a scenario in which all the proteins encoded by the SARS-CoV-2 genomes may be susceptible to recognition by HLA-I and II molecules. This is possible because post-transcriptional changes do not occur in the proteins encoded by the viral genome. However, post-translational modifications may occur, which is inconvenient for experimental interpretation of results obtained from immunogenic *in silico* analyses.

Proteins synthesised in the cytoplasm without prediction of the transmembrane domain may be less likely to be changed by glycosylation. However, in studies of coronaviruses closely related to SARS-CoV-2, the polyprotein pp1a/b synthesised in the cytoplasm can result in NSP3, NSP4 and NSP6, which despite being non-structural proteins are susceptible to glycosylation given their proximity to the endoplasmic reticulum[42]. The structural proteins SP, MG and envelope protein can harbour a greater number of post-translational modifications, with N-glycosylations being most likely. These modifications can vary among proteins and are not necessarily related to the evasion of the immune system. NP, although a structural protein, has modifications related to phosphorylation that optimise its recognition of RNA. It is therefore an important vaccine target because of its greater expression during the infection[42].

In the SP of SARS-CoV-2, glucan shields with densities that limit probable antibody recognition have not been identified[43]. However, its interaction with innate immunity remains unknown. With these modifications, HLA may be able to recognise and present the peptides derived from antigenic processes when they are captured by dendritic cells[44]. This has been most frequently studied in cancer, evidenced in T and B cell primers, which are in some cases even more immunogenic[45–47]. Because they are naturally recognised as epitopes, their in vitro recognition can be included in complementary studies. It quickly identifies the immunogenic regions susceptible to being glycosylated and therefore optimises the potential vaccine formulations.

The best option for producing a SARS-CoV-2 vaccine in a short period is not only the construction of the best targets with potential immunogenicity but also the development of libraries of potential vaccine peptides, including the identification of experimental post-translational changes, especially those affecting immune recognition. Affinity and antigenicity should be selected as the most important characteristics to initiate specific antigenic recognition. The proteins responsible for virus replication and those that perform cleavage to become functional are important vaccine and pharmacology targets since they take part in the early stages of infection. The antigen presentation process is shown in (Supplementary Fig. S4), with special emphasis placed on SARS-CoV-2. Identification of potential epitopes (proteins or antigenic peptides with a high affinity for HLA-I or II molecules) from a certain population with more frequent HLAs could lead to adequate antigen presentation of TLs. This may be used to activate a favourable adaptive response and generate immunological memory. HLA-I molecules present epitopes to TL CD8+, which are cytotoxic, lyse infected cells, initiate apoptosis and prevent the formation of new virions. When HLA-II molecules present epitopes to TL CD4+[1,48], they promote the formation and maturation of new lineages of TLs for the production of specific antibodies by B lymphocytes (BLs).

Recent experimental studies on SARS-CoV in patients with COVID-19 have identified the structural proteins NP, SP and MG of SARS-CoV-2 as targets for the antibodies in the serum of patients with a neutralising capacity[49]. These proteins are the main targets for stimulating the creation of antibodies by BLs. These characteristics are included in our vaccine proposal, which is summarised in Figs. 2 and 3.

Potential epitopes of TLs, CD4+, CD8+ and BLs are contained in the conserved sequences of 92 SARS- CoV-2 proteomes

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 = 63) and NP (n = 23). Discrimination on the origin of predictions from the conserved sequence is included in (Supplementary Table 6).

HLA-II alleles found to be strong binders of various peptides (promiscuous alleles) were identified in the following order of frequency: DRB1*04:02 (n = 50), DRB1*09:01 (n = 35), DRB1*08:01 (n = 33), DRB1*16:02 (n = 33), DRB1*04:03 (n = 30), DRB1*07:01 (n = 24), DRB1*01:01 (n = 23) and DRB1*04:05 (n = 20). (Supplementary Table 7) summarises the conserved sequences present in the SP binding domain of interest; the 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 relatively more frequently in LATAM; therefore, it may be related to clinical outcomes and deserve to be studied in more detail.

The 15 predicted mers, with antigenic and strong binding characteristics, were evaluated for allergenic and toxic character- istics using AllerTOP and Toxin Ped, respectively. To achieve 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 being the most accurate by IEBD (as of March 15, 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: the ORF1 transcript (n= 573), SP (n = 163), MG (n = 51), the ORF-3a transcript (n = 43) and NP (n = 34). The contribution made by each conserved sequence to these proteins is included in (Supplementary Table 8).

A prediction was not possible only for the HLA-B*51:10 allele because it was not found on the available list. In addition to the antigenicity and strong binding characteristics of these groups, we also considered those with vaccine potential, and potential immunogenic, non-allergenic and non-toxic characteristics to be on our list. Adaptive immunity protection is now known to change the natural course of the disease by neutralising SARS-CoV-2[50]. The most recent experimental structures of the SP and NP proteins, found in RCSB PDB (https://www.rcsb.org/) 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[51], and conformational pre-fusion in the open and closed state. Analyses were performed using the Discotope 2.0 server adjusted to 80% specificity.

We used the prediction algorithm on all proteins based on their available 3D structure, using the specific SP or NP chains. The predictions, including the characteristics of the complexes found, the intervals, and frequency with which probable epitopes they were identified in each structure, are shown in (Supplementary Table 9). For SP we have indicated three possible regions of continuous epitopes and one discontinuous 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 this predictions, is absence of glycosylation in this residues.

For NP, we identified five regions of possible continuous epitopes and two discontinuous ones for 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. NP and SP with linear and discontinues epitopes zones described above, are illustrated in (Supplementary Fig. S5).

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, which included predicting epitopes in the functional domains using RDB or the RNA binding site for NP, antigenicity, toxicity, allergenicity and the identification of conserved sequences in SARS-CoV with experimental immunological evidence available at IEBD, were used to identify two ideal options. These are presented in (Supplementary Table 10). After the filters were applied, the best options were selected to choose the BL candidates to be incorporated into the multi-epitope construction. These sequences are summarised in (Supplementary Table 11).

Peptides with vaccine potential are characterised by more of the most frequent alleles of LATAM in con- served sequences from SARS-CoV-2 proteomes

Peptide 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 summarised in Table 1. These 'potential promiscuous vaccine peptides' (PPVPs) share characteristics for the optimised construction of multi-epitopes, including being antigenic, non-toxic and non-allergenic. This group of peptides contain conserved and partial sequences that are experimentally proven to bind to various molecular targets that mediate the specific immune response against SARS-CoV. For HLA-II molecules, DDSEPVLKGVKLHYT, the peptide conserved in SARS-CoV and contained in the SP, is likely the best candidate for an antigenic peptide; it is also valid because of experimental evidence of its binding to targets in restricted HLA, TLs and BLs. Other partial sequences with experimental validity were identified in a 90% BLAST of structural proteins including MG and SP.

For HLA-I molecules, the majority of predictions corresponded to the ORF1 transcript, in which the KVKYLYFIK peptide was experimentally valid and conserved in SARS-CoV with a high antigenicity score. Only two other non-ORF1 peptides, ITLCFTLKR conserved in ORF7a and WTAGAAAYY of SP, were found in this group.

The ORF1 transcript and the structural proteins SP and MG have the highest recognition by the most frequent alleles in LATAM

The peptides found with the highest promiscuity in HLA-I molecules were MPYFFTLLL (n = 66), which is from the ORF1 transcript, followed by epitopes conserved in SARS-CoV SP, FAMQMAYRF (n = 58), and ORF1, FLLNKEMYL (n = 42). Those with the highest promiscuity in HLA-II molecules were a partially conserved SARS-CoV epitope found in MG, SFRLFARTRSMWSFN (n = 7), followed by two predictions from the SP protein, QSIIAYTMSLGAENS (n = 4) and VLSFELLHAPATVCG (n = 4). The latter was partially recognised by tests on BL and HLA molecules in SARS-CoV.

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. The alleles found in order of frequency were as follows: HLA-C*08:01-HLA-C*08:03 (n = 98), HLA-C*16:01 (n = 97), HLA-C*03:05 (n = 94), HLA-C*03:03-HLA-C*03:04 (n = 92), HLA-C*03:02-HLA-C*15:03 (n = 90) and HLA-C*16:02-HLA-C*12:03-HLA-C*01:06-HLA-C*15:02 (n = 89). Alleles with relatively low, but higher than average, frequencies belonged to the countries Colombia-Brazil, Colombia-NIC and Colombia-Costa Rica-NIC. Those around the 50th percentile belonged to the countries Brazil-Colombia-NIC-Peru, Colombia-Venezuela and Brazil-Colombia-NIC, respectively with the alleles previously mentioned. 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 recognise peptides derived from SARS-CoV-2 these alleles were ranked 81st, 119th and 154th, respectively.

Flexible peptide-protein coupling signals favourable energy and anchorage residues to HLA molecules

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

The best models were those contained in clusters with the highest density, in accordance with the average RSMD obtained from each simulation and their adequate position in the HLA cleft. The results are shown in Fig. 4. 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) included. They also display some interactions with β-2 globulin, which serve to stabilise the α1 and α2 chains where the binding groove is formed and facilitate peptide bonding[52].

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.

Construction of the proposed multi-epitope vaccine

In the N terminal, β-defensins have a range of immune responses related to the maturation of cells that mediate innate immunity, such as dendritic cells and TLs, as well as antiviral activities[53,54]. Another adjuvant used was the universal memory TL helper peptide (TpD), an auxiliary peptide that can aid memory generation as a target of TL CD4+[55]. 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 generation of high and specific IgG titers. It can also overcome barriers, indicated by the high diversity among HLA molecules; hence, it reaches a larger population and is safe[56,57]. Towards the C terminal, a peptide domain capable of interacting with M cells and mediating up to an 8-fold increase in intestinal absorption was added[58].

The linkers used as spacers between the HTL GPGPG and AYY epitopes are recognised as suitable spacers that facilitate antigen presentation by directly interacting with transport and assembly mediators to HLA molecules[59]. 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[60]. The following linkers were used in the vaccine, which contained 510 amino acids: 6 EAAK, 6 GPGPG, 11 AAY and 5 KK. The proposed order of the construct is presented in Fig. 5.

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 construct that was safe, stable and capable of evoking 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, respectively. The physicochemical characteristics of the multi-epitope construct, along with special consideration for the final CTGKSC peptide, are shown in Table 2. 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 *Escherichia coli* (*E. coli*), which is the bacteria most commonly used to produce recombinant proteins[61].

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.

Three-dimensional structure modelling 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 analysed using the SOPMA and PSIPRED servers, which revealed the presence of 43% α-helix, 21% β-foil, 30% coils and 6% β-turns in the vaccine's construction (Supplementary Figs. S6 and S7). 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 analyse 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 Robbeta, based on homology models, and we used analysis of the Ramachandran diagram and ERRAT to validate the quality of this approach. The results are summarised in (Supplementary Fig. S8).

Molecular docking, interactions with heterodimer TLR-4/MD-2

By using the ClusPro 2.0 server to simulate molecular docking between the vaccine construct and TLR-4, 30 models were generated (Supplementary Table 12); these were classified by cluster size according to their representative position. The lowest energy -1192.6 (Supplementary Fig. S9), 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. This cluster with the balanced adjustment was chosen because the pose of this was closer than those of adjustments 2 and 3, which were related to a majority of hydrophobic and electrostatic interfaces[62].

This model positions the N terminal of the multi-epitope construct and a 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 the terminal C of the multi-epitope construct.

In addition, interactions are present in the internal part and running towards Myeloid Differentiation Factor 2 (MD-2); these are mostly hydrogen and some saline bridges, the latter being maintained primarily by the ARG 157 and ARG 159 residues of the SFRLFARTRSMWSFN peptide. These are illustrated in (Supplementary Figs. S9-S11).

Molecular Dynamics

The root mean square deviation RMSD for the vaccine-receptor complex was evaluated during the complete simulation time (Fig. 6 A). The RMSD plot showed a considerably increase until it stabilizes 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 finding the best position with respect to its receptor. After 20 ns from beginning of the simulation it is remained stable, which is an indicator that the vaccine reaches the best conformation to form a stable complex. In addition, the root mean square fluctuation RMSF of all residue, for vaccine and its receptor was evaluated. Residues from the vaccine start at GLY-1483 and finalize at residue CYS-1992 (see Fig. 6B), The RMSF showed a stable conformation from residue GLY-1483 to LYS1683 from the vaccine segment, which make part of the β-defensin 3 to TpD of the vaccine construct, Fig. 5a. This stabilization, is made due to the salt bridges formed into residue ASP-1424:ARG1639, ASP-756:ARG-1641, GLU-1180:ARG-1524, GLU-1183:ARG1520 and GLU-1556:ARG930 which

showed a regular contacts in the whole simulation (Supplementary Fig.12). On the other hand, a segment of the vaccine, which corresponds to residue PHE-1684 to CYS-1992 (Fig. 6B), showed a higher RMSF, displaying a pronounced motion compared with the GLY-1483 to LYS1683 segment, while TLR-4 receptor keeps stable. Beside this motion, the complex remains stable during the 68 ns of simulation.

The radius of gyration Rg was calculated to determine the compactness of vaccine-receptor complex system during the simulation, Fig. 6C. It does not show a significant change. The complex exhibit a compact folded structure in the whole simulation, with an average Rg of about 4.8 nm. This folded complex structure is stabilized by around 9 H-bonds in average (Fig. 6B), and 5 salt-bridges which were described above. Thus, the Vaccine-TLR4 complex showed a stable conformation during the 68 ns of simulation.

Optimisation 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-optimised according to the JCAT server. Furthermore, during optimisation 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 (Supplementary Fig. S13).

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 organism for cloning purposes since multiple-epitopes 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 optimised cDNA was then inserted near the ribosome binding site using Snapgene (Supplementary Fig. S13).

Immune system simulation

To identify the immunogenic profile of the vaccine construct, we used the C-IMMSIM immune server. As shown in Fig. 7, the secondary and tertiary responses showed greater global responses and the presence of memory T and B cells. This may have been due to the cumulative effect of cells at the serum level that possess 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+ T cells shown. In addition, 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. These results are consistent with the construct having a natural immunogenic capacity, which was applied without 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 (Supplementary Fig. S14).

Discussion

In this study, we used a reverse engineering approach to design a multi-epitope vaccine for SARS-CoV-2, which was based on the identification of PPVPs from the conserved regions of proteins in 92 SARS-CoV-2 proteomes. Because of our methodology, our approach has advantages over other vaccine design strategies: it is a more powerful, specific, safe, effective and complete response[35]. Furthermore, it was optimised for the LATAM population by identifying the most frequent HLA-I and HLA-II alleles in this region. In HLA-I molecules, we found a greater number of predicted agglutinations associated with the C locus, with a median of 82 agglutinations. In the complete dataset, B*27:05 (n = 8) was the least frequent allele, whereas C*08:01 (n = 98) was the most frequent; this has been reported in other populations[56]. Thus, the C locus should be a focus of future research on increased recognition of SARS-CoV-2 and could be used in future rational vaccine formulations.

We searched all conserved proteins to identify PPVPs capable of activating CTLs by breaking the machinery necessary for viral replication. This process causes programmed death through the degradation of genetic material and proteins, including those recently synthesised. We hypothesise that this could help avoid an early expansive immunopathological response, similar to findings for other viruses such as HIV or HCV[55,57].

It has been reported that an immune response maintained by HTL-2 cells instead of HTL-1 may be due to the stimulation of BLs by unassembled virus proteins[63]. Proteins <70 kDa evoke a direct anti-inflammatory response; proteins >70 kDa, such as SP, indirectly stimulate BLs via follicular dendritic cells through complements or FCγ receptor. This explains the recognition of non-structural protein antibodies with less surface access in SARS-CoV and SARS-CoV-2 but with neutralising capacity[55,57]. Accordingly, we used the complete structural proteins of SP, NP and MG to identify PPVPs related to HLA-II and BLs. In SARS-CoV-2, SP and NP are known to be antigenic and play important roles in pathogenic host interaction, in addition to the BL and HLA-I and II epitopes that overlap[36,64].

Here, we corroborated NP as a relevant vaccination target. In NP, we found areas with predictions of linear and discontinuous epitopes longer than those of SP; therefore, there was a greater chance of antibody production. We also highlighted the immunogenicity of SP: we identified a greater number of strong agglutination predictions in association with more frequent HLA-II molecules in LATAM. We emphasise 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 that were associated with HLA-I and II molecules. This was also in the most frequent phenotypes in LATAM, e.g. B*16:02. Thus, this is an important immunogenic region that is also associated with the linear epitopes of BLs. However, it should be noted that mutations in this immunodominant region could imply changes in antigenic recognition with loss of protective capacity[65].

The routine identification of mutations during the course of the pandemic could therefore be updated in parallel libraries of potential vaccine peptides located in the immunodominant regions of SARS-CoV-2, including the example identified here (Supplementary tables 6 and 8). This would also enable validation of potential epitopes subject to post-

translational changes, allowing for the continuation of related research in the future. We have shared this resource as an online spreadsheet at the following link to encourage community-curated collaboration: https://tinyurl.com/y84n9ttq.

From 18 epitopes comprising the final construct, predicted as PPVPs with targets for HLA-I and II molecules, another five have been identified as possible vaccine targets: P9, P7, P5, P18 (in SP) and P12 (in ORF7)[66–71]. Interestingly, there is experimental evidence to show that P9 conserved in SARS-CoV is a potential vaccine target[66]. Most of these epitopes originate in the ORF1a/b transcript or from MG and SP. In comparison to other proteins, these could harbour a greater number of conserved immunogenic regions because they were found in predictions with greater agglutinations against LATAM alleles. 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[72]. Some HLAs and supertypes have been associated with a lack of response, e.g. the oral rotavirus vaccine in infants[73].

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[74,75]. Human β-defensin 3, which activates human monocytes dependent on TLR1/2[76], 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+ T cells[77]. This approach was previously used to construct a bovine herpes 1 DNA vaccine, in which it increased the production of IFN-γ-dependent LT CD8+[78].

We also used TpD as adjuvant, which has previously been studied extensively in animal models and in peripheral blood samples, in which it resulted in increased TLs and the robust generation of antibodies[55]. The final construct was catalogued as antigenic, non-allergenic and non-toxic, which establishes it as a safe and powerful candidate vaccine against SARS-CoV-2.

The peptides selected in the final construct confer a greater benefit in the early phase of infection when CD8+ target cells are prominent and NSPs are translated directly in the cytoplasm. In addition, they protect the host in more advanced phases. Structural proteins and the production of antibodies against SP and NP were considered, along with CD4+ targets and stimulation of B cells against proteins that are overly expressed during infection[79]. 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 construct. Without the extra adjuvant (without LPS), the vaccine construct promoted an immune profile that indicated adequate antigenic presentation and the subsequent generation of immune memory in groups of T and B cells. In addition, there was a polarised 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 C-IMMSIM algorithm have been conducted with experimentally validated peptides and correlations with in vitro studies were found[80]; this suggests that in vitro experiments with our construct may potentially evoke similar cellular behaviours.

These predictions are consistent with analysis of flexible protein-peptide docking, which provides evidence for 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, and it is associated with an effective immune interaction and therefore sufficient presentation of CD4+ and CD8+ T cells that, when activated, trigger an immune response[52].

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, and this 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[81]. This position has been found in other vaccine models constructed using different methods. Therefore, it seems to be relevant to the stabilisation of the multi-epitope construct[80].

In addition to the immunogenicity and stability identified in our molecular dynamics analysis, the physicochemical analysis of the construct revealed other desirable characteristics in the multi-epitope construct related to a safe profile. In particular, it was not predicted to be toxic or allergenic. It was also thermally stable, which suggests that it would be suitable for 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.

Conclusion

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, especially in regions with emerging economies. A vaccine is considered to be the best option to limiting the medium- and long-term effects of COVID-19 on a global scale. By using bioinformatics tools and considering the immunological profiles of vulnerable and diverse regions in LATAM, we created a multi-epitope peptide vaccine with immunogenic capacity that surpassed the safety standards required to begin complementary experimental studies (including phase 1 clinical trials). We hope to quickly validate the vaccine and take action against the ongoing pandemic.

Methods

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). The metadata was extracted manually, and provenance, 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.

Identification of amino acid sequences conserved in SARS-CoV-2 proteomes

In order to exclude the potential non-synonymous substitutions in the amino acid sequences from the recovered SARS-CoV-2 proteomes, we performed a multiple alignment using the MAFFT server v7.0 (https://mafft.cbrc.jp/alignment/server/)[82], with the genome NC_04551 as a reference, including structural and accessory proteins, in addition to the transcript ORF1a/b. The resulting preserved sequences formed non-redundant sequences, which were used for further analysis.

Identification of the most frequent HLA-I and HLA-II alleles in the 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 region were downloaded using a specialist search tool (http://www.allelefrequencies.net/default.asp). The phenotype percentage (PF) was calculated using the median of the Equation 1. The most frequent phenotypes were separated and PF values greater than the median were grouped by locus and country. Tableau v2020.2[83] was used to create a geographic representation of the alleles found most frequently.

$$PF = 1 - (1 - \text{Allele Frequency})^2 \quad (1)$$

HLA-I and HLA-II allele selection

Only the alleles corresponding to the DRB loci were identified for analysis in HLA-II because of the higher precision of these predictions due to 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: >3,000 isolates have been identified in humans[84]. This variability seems to be due to a rapid evolutionary response to the diversity of extracellular pathogens[85].

The HLA-DQ loci were not taken into account 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 responses[86].

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. This HLA-I group of genes are also isolated more frequently than those in DRB, with the B locus being the most diverse.

Prediction of the HTL epitope

In order to predict the TL CD4+ epitopes, conserved sequences from SARS-CoV-2 structural proteins with at least 15 mers were used. The algorithm NetMHCII v2.3 was used; this approach is based on training individual molecules from complex experiments, which gives greater accuracy[87]. 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 affinities <IC50. The predictions that remained following filtering were considered to be potential HLA-II epitopes capable of being recognised by HTL CD4+.

Prediction of the CTL epitope

In order to break through the SARS-CoV-2 viral assembly at an early stage of infection, we used structural 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, namely NetMHCpan and MHCflurry[88,89]. These approaches used cross-trained and individually trained neural networks,

respectively, which are based on experimental data from the IEBD database and integration of eluted peptides and other ligands identified by mass spectrometry.

A portable version of NetMHCpan 4.0 was used following a request to the "Health Tech" area of the Technical University of Denmark (https://services.healthtech.dtu.dk/cgi-bin/sw_request). MHCflurry v1.6.1 was downloaded from the author's github repository (https://github.com/openvax/mhcflurry/releases). 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 (http://tools.iedb.org/immunogenicity/)[90] 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 to be candidates for further evaluation of antigenicity, allergenicity and toxicity.

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[91]. The Vaxijen 2.0 server (http://www.ddg-harmfac.net/vaxijen/VaxiJen/VaxiJen.html) 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[92]. This process discriminates between probable and unlikely antigens. Peptides with strong agglutinations and an antigenicity threshold >0.4 were considered to be potential immunogens.

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/)[93] and ToxinPred (http://crdd.osdd.net/raghava/toxinpred/)[94]. AllerTOP employs a similar approach to Vaxijen 2.0: it uses an auto-covariance transformation to normalise 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; thus, the method has 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[89]. 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. To obtain accuracy close to 97%, the position and frequency were analysed along with other characteristics of the amino acids that are most abundant in toxic peptides. Non-toxic and non-allergenic peptides with immunogenic potential were considered to be potential SARS-CoV-2 vaccine targets.

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 characterised as follows: being peptides conserved in proteomes of 9 or 15 mer, having immunogenic characteristics, being non-toxic and non-allergenic, and having a strong affinity to the most common alleles found in LATAM. These matrices were grouped by allelic HLA class in order to identify possible non-redundant combinations capable of covering all the HLA molecules tested. Where peptides had equal predictions, the highest score from Vaxijen was used to decide the best candidate peptide. 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/).

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+ T cell receptors[52]. 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[95]. A standalone python package of CABS v 0.9.16 was downloaded from a repository (https://bitbucket.org/lcbio/cabsdock/downloads/).

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 sec- ondary structure of the peptides by PSIPRED through the RPBS web portal (https://bioserv.rpbs.univ-paris-diderot.fr/index.html). 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/)[96].

Prediction of the BL epitope

To identify potential continuous epitopes capable of stimulating a BL response, the artificial neural network-based ABCpred tool[97] (http: //crdd.osdd.net/raghava/abcpred/) was used. The cutoff threshold was set to 0.90, which allows a greater specificity to be selected. To predict potential discontinuous BL epitopes, Discotope v2.0[98] (server version) (http://www.cbs.dtu.dk/services/DiscoTope/) 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 as it had a par- tial or total presence in the experimental sequences identified as SARS-CoV immunological targets in IEBD with a 90% BLAST alignment. In addition, it was located in potentially immunogenic regions, either because they are shown in the Discotope-provided areas from various experimental structures or because they are 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 (ARP)[99] (http://crdd.osdd.net/raghava/algpred/submission.html) were chosen as potential BL epitopes and were added to the multi- epitope construct.

Constructive design of the multi-epitope vaccine

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, β-defensin 3, TpD and the PADRE sequence, using the EAAAK linker. 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[58].

Analysis of antigenicity and allergenicity of the vaccine construction

To define whether the final design of the multi-epitope proposal was safe and viable, its antigenic capacity was considered by predicting antigenicity using Vaxijen values with a threshold >0.4. Its toxicity was predicted using ToxinPred and its allergenicity using Allergen FP[100] (http://ddg-pharmfac.net/AllergenFP/).

Analysis of the physicochemical properties of the multi-epitope vaccine construct

To analyse the construct's physicochemical properties, the Expasy server (https://web.expasy.org/protparam/) and the ProtParam tool[101] were used to determine optimal recognition by the immune system. Negative values of average hydropathy (GRAVY)[102] and stability are especially important for adequate antigen presentation as well as determining solubility, which was performed using SolPro (http://scratch.proteomics.ics.uci.edu/). Other parameters related to production and expression were also considered including the aliphatic index (related to the thermostability of the construct).

Vaccine structure prediction and validation

The secondary structure of the multi-epitope design was predicted using SOPMA[103] (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/ and PSI-PRED[104] (http://bioinf.cs.ucl.ac.uk/psipred/). The tertiary structure was predicted using the Robetta server (http://robetta.bakerlab.org/) based on homology[105]. An evaluation of quality was carried out on the structure using the Ramachandran diagram in PDBsum[106] (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and by the ERRAT server[107] (https://servicesn.mbi.ucla.edu/ERRAT/).

Refinement was performed using 3D Refine[108] (http://sysbio.rnet.missouri.edu/3Drefine/) and then GalaxyRefine[109] (http://galaxy.seoklab.org/c bin/submit.cgi?type=REFINE).

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/MD-2 complex and the multi-epitope construct using molecular coupling, which was performed using the Cluspro 2.0[110] server (https://cluspro.bu.edu/login.php). 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 analysed by sum PDBs and plotted by UCSF Chimera v1.14[111].

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. Minimisation and molecular dynamic protocols were performed with AMBER 16[112]. Force field parameters were used for the amino acid residues ff14SB43[113]. The complex was subjected to unrestricted molecular dynamic simulations for all atoms in an explicit dissolvent using the PMEMD GPU version algorithm in Amber16[112].

The Leap module integrated within Amber16 was used to add missing hydrogen atoms and add Cl^- ions for neutralisation. The systems were immersed in an orthorhombic box using the TIP3P[114] water model. The long-range electrostatic interactions were calculated using the particle mesh Ewald method[115], with a direct space and a vdW cutoff of 12 Å//. An initial minimisation was applied, using a potential of 500 kcal mol^{-1} $Å//^2$ to the solute, for 10,000 steps using the steepest descent algorithm followed by 10,000 steps with the conjugate gradient method. Subsequently, 10,000 unrestricted minimisation 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}-1 $Å//^2$, which was 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 and using the Berendsen barostat to maintain the pressure at 1 bar. The SHAKE[116] algorithm was used to restrict the bonds of all hydrogen atoms, and a 2 fs time-step was used with the precision model SPFP[117] 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 paths were analysed for 2 ps of the simulation using CPPTRAJ and PTRAJ[118].

Codon optimisation 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 *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. In addition, it is an optimised lineage for overexpression of recombinant proteins[119]. To identify the best cloning strategy, the complete protein sequence of the multi-epitope construct was first converted into cDNA using Backtranseq reverse translation (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/). The resulting cDNA sequence was further optimised by adapting the most frequently used codons of *E. coli* K 12 to enhance protein expression using the JCAT server[120] (http://www.jcat.de/Start.jsp). 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 optimised genetic sequence of the construct.

Immune simulation

The immunogenic behaviour of the multi-epitope construct was simulated using the C-IMMSIM server (https://kraken.iac.rm.cnr.it/C- IMMSIM/). 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 behaviours 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 matrices of some HLA molecules. The algorithm can also identify probable linear B cell 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 the other values predetermined.

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.

Declarations

Author contributions statement

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.

Competing Interests

The authors declare no competing interests.

Data availability

For any data related please contact juvenal.yosa@unnisimonbolivar.edu.co

References

1. Prompetchara, , Ketloy, C. & Palaga, T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. *Asian Pac J Allergy Immunol* **38**, 1–9 (2020).
2. Li, G. *et al.* Coronavirus infections and immune responses. *medical virology* **92**, 424–432 (2020).
3. Jin, *et al.* Virology, epidemiology, pathogenesis, and control of COVID-19. *Viruses* **12**, 372 (2020).

4. Adhikari, S. *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. diseases poverty* **9**, 1–12 (2020).
5. Rodriguez-Morales, J. *et al.* COVID-19 in Latin America: The implications of the first confirmed case in Brazil. *Travel. medicine infectious disease* (2020).
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. *Trop. Medicine Reports* 1–4 (2020).
7. Millán-Oñate, J. *et al.* Successful recovery of COVID-19 pneumonia in a patient from Colombia after receiving chloroquine and clarithromycin. *Annals Clin. Microbiol. Antimicrob.* **19**, 1–9 (2020).
8. The, L. The unfolding migrant crisis in Latin America. *Lancet (London, England)* **394**, 1966 (2019).
9. Ochoa, 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).
10. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* **20**, 533–534 (2020).
11. Wang, , Li, R., Lu, Z. & Huang, Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. *Aging (Albany NY)* **12**, 6049 (2020).
12. Organization, H. & Others. Immunity Passports in the Context of COVID-19. *Sci. Brief* **24** (2020).
13. Melgaço, G., Azamor, T. & Ano Bom, A. P. D. Protective immunity after covid-19 has been questioned: What can we do without sars-cov-2-igg detection? *Cell. Immunol.* **353**, 104114, DOI: https://doi.org/10.1016/j.cellimm.2020.104114 (2020).
14. Totura, L. & Baric, R. S. SARS coronavirus pathogenesis: host innate immune responses and viral antagonism of interferon. *Curr. opinion virology* **2**, 264–275 (2012).
15. Chan, S. F. *et al.* Homozygous L-SIGN (CLEC4M) plays a protective role in SARS coronavirus infection. *Nat. genetics* **38**, 38–46 (2006).
16. Wang, -F. *et al.* 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 immunology* **24**, 421–426 (2011).
17. Tay, 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.* 1–12 (2020).
18. Guo, -R. *et al.* The origin, transmission and clinical therapies on coronavirus disease 2019 (covid-19) outbreak–an update on the status. *Mil. Med. Res.* **7**, 1–10 (2020).
19. Song, Z. *et al.* From sars to mers, thrusting coronaviruses into the spotlight. *Viruses* **11**, 59 (2019).
20. Wu, *et al.* Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell host & microbe* (2020).
21. Narayanan, K., Huang, C. & Makino, S. SARS coronavirus accessory proteins. *Virus research* **133**, 113–121 (2008).
22. Boopathi, S., Poma, A. B. & Kolandaivel, Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J. Biomol. Struct. Dyn.* 1–10 (2020).
23. Tu, -F. *et al.* A review of SARS-CoV-2 and the ongoing clinical trials. *Int. journal molecular sciences* **21**, 2657 (2020).
24. Russell, C. D., Millar, J. E. & Baillie, J. K. Clinical evidence does not support corticosteroid treatment for 2019-nCoV lung injury. *The Lancet* **395**, 473–475 (2020).
25. S¸ ims¸ek Yavuz, & Ünal, S. Antiviral treatment of covid-19. *Turkish J. Med. Sci.* **50**, 611–619, DOI: 10.3906/ sag-2004-145 (2020).

26. Wang, *et al.* A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* **11**, 1–6 (2020).
27. Channappanavar, R., Fett, C., Zhao, J., Meyerholz, D. K. & Perlman, S. Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus *J. virology* **88**, 11034–11044 (2014).
28. Zhou, M. *et al.* Screening and identification of severe acute respiratory syndrome-associated coronavirus-specific CTL *The J. Immunol.* **177**, 2138–2145 (2006).
29. Wang, , Shang, J., Jiang, S. & Du, L. Subunit vaccines against emerging pathogenic human coronaviruses. *Front. microbiology* **11**, 298 (2020).
30. Ahn, D.-G. *et al.* Current status of epidemiology, diagnosis, therapeutics, and vaccines for novel coronavirus disease 2019 (COVID-19). *microbiology biotechnology* **30**, 313–324 (2020).
31. World Health Organization. DRAFT landscape of COVID-19 candidate vaccines. *World* (2020).
32. Dhama, K. *et al.* COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. *Vaccines & Immunother.* 1–7 (2020).
33. Le, T. *et al.* The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* **19**, 305–306 (2020).
34. Ahmed, F., Quadeer, A. A. & McKay, M. R. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* **12**, 254 (2020).
35. Enayatkhani, *et al.* Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *J. Biomol. Struct. Dyn.* 1–16 (2020).
36. Ong, E., Wong, U., Huffman, A. & He, Y. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *BioRxiv* (2020).
37. Prasanna, L., Abilash, V. G. & Others. Coronaviruses pathogenesis, comorbidities and multi-organ damage–A review. *Life Sci.* 117839 (2020).
38. 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. *Atencion Primaria/Sociedad Espanola de Medicina de y Comunitaria* (2020).
39. Olejnik, J., Hume, A. J. & Mühlberger, E. Toll-like receptor 4 in acute viral infection: Too much of a good thing. *PLoS pathogens* **14** (2018).
40. DeDiego, L. *et al.* Coronavirus virulence genes with main focus on SARS-CoV envelope gene. *Virus Res.* **194**, DOI: 10.1016/j.virusres.2014.07.024 (2014).
41. Vina, A. F. *et al.* Tracking human migrations by the analysis of the distribution of hla alleles, lineages and haplotypes in closed and open populations. *Philos. Transactions Royal Soc. B: Biol. Sci.* **367**, 820–829 (2012).
42. Fung, S. & Liu, D. X. Post-translational modifications of coronavirus proteins: roles and function. *Futur. Virol.* **13**, 405–430 (2018).
43. Watanabe, , Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the sars-cov-2 spike. *Science* (2020).
44. Wolfert, A. & Boons, G.-J. Adaptive immune activation: glycosylation does matter. *Nat. chemical biology* **9**, 776–784, DOI: 10.1038/nchembio.1403 (2013).
45. Lipinski, *et al.* Enhanced immunogenicity of a tricomponent mannan tetanus toxoid conjugate vaccine targeted to dendritic cells via Dectin-1 by incorporating β-glucan. *J. immunology (Baltimore, Md. : 1950)* **190**, 4116–4128, DOI: 10.4049/jimmunol.1202937 (2013).

46. Hanisch, -G. & Ninkovic, T. Immunology of O-glycosylated proteins: approaches to the design of a MUC1 glycopeptide- based tumor vaccine. *Curr. protein & peptide science* **7**, 307–315, DOI: 10.2174/138920306778018034 (2006).
47. Roulois, D., Grégoire, M. & Fonteneau, -F. MUC1-specific cytotoxic T lymphocytes in cancer therapy: induction and challenge. *BioMed research international* **2013**, 871936, DOI: 10.1155/2013/871936 (2013).
48. Blum, J. S., Wearsch, A. & Cresswell, P. Pathways of antigen processing. *Annu. review immunology* **31**, 443–473 (2013).
49. Liang, M.-f. *et al.* SARS patients-derived human recombinant antibodies to S and M proteins efficiently neutralize SARS-coronavirus infectivity. *Environ. Sci.* **18**, 363 (2005).
50. Addetia, A. *et al.* Neutralizing antibodies correlate with protection from sars-cov-2 in humans during a fishery vessel outbreak with high attack rate. *medRxiv* DOI: 1101/2020.08.13.20173161 (2020). https://www.medrxiv.org/content/ early/2020/08/14/2020.08.13.20173161.full.pdf.
51. Ter Meulen, J. *et al.* Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS medicine* **3** (2006).
52. Liu, J. & Gao, G. Major histocompatibility complex: Interaction with peptides. *eLS* (2011).
53. Howell, M. D., Streib, J. E. & Leung, D. M. Antiviral activity of human beta-defensin 3 against vaccinia virus. *J. allergy clinical immunology* **119**, 1022–1025 (2007).
54. Chen, , Zaro, J. L. & Shen, W.-C. Fusion protein linkers: property, design and functionality. *Adv. drug delivery reviews* **65**, 1357–1369 (2013).
55. Fraser, C. *et al.* Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. *Vaccine* **32**, 2896–2903 (2014).
56. Fast, E. & Chen, B. Potential T-cell and B-cell Epitopes of 2019-nCoV. *bioRxiv* (2020).
57. Ip, P. *et al.* Alphavirus-based vaccines encoding nonstructural proteins of hepatitis C virus induce robust and protective T-cell responses. *Mol. Ther.* **22**, 881–890 (2014).
58. Fievez, *et al.* In vitro identification of targeting ligands of human M cells by phage display. *Int. journal pharmaceutics* **394**, 35–42 (2010).
59. Dorosti, H. *et al.* Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. *Biomol. Struct. Dyn.* **37**, 3524–3535 (2019).
60. Arai, , Ueda, H., Kitayama, A., Kamiya, N. & Nagamune, T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein engineering* **14**, 529–532 (2001).
61. Idicula-Thomas, S. & Balaji, V. Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli. *Protein science : a publication Protein Soc.* **14**, 582–592, DOI: 10.1110/ps.041009005 (2005).
62. Vajda, *et al.* New additions to the ClusPro server motivated by CAPRI. *Proteins* **85**, 435–444, DOI: 10.1002/prot.25219 (2017).
63. Martin, J. B Cells Over-Activation by Viral Proteins <70 kDa Causes Th2 Immune Suppression in COVID-19 *Preprints* DOI: (doi:10.20944/preprints202005.0244.v1 (2020).
64. Neefjes, , Jongsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).

65. Wang, -D. *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. virology* **78**, 5612–5618 (2004).
66. Almofti, 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.* **6**, 94–114 (2018).
67. Rakib, *et al.* Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. *Comput. Biol. Medicine* 103967, DOI: 10.1016/j.compbiomed.2020.103967 (2020).
68. AP, , VS, A. & Others. Design of multi-epitope vaccine candidate against SARS-CoV-2: a In-Silico study. *J. Biomol. Struct. & Dyn.* 1–10 (2020).
69. Su, Q.-d. *et al.* The biological characteristics of SARS-CoV-2 Spike protein Pro330-Leu650. *Vaccine* (2020).
70. Mitra, , Shekhar, N., Pandey, J., Jain, A. & Swaroop, S. Multi-epitope based peptide vaccine design against SARS-CoV-2 using its spike protein. *bioRxiv* (2020).
71. Joshi, , Joshi, B. C., Mannan, M. A.-u. & Kaushik, V. Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. *Informatics Medicine Unlocked* 100338 (2020).
72. Zhao, L., Zhang, M. & Cong, H. Advances in the study of HLA-restricted epitope vaccines. *vaccines & immunotherapeutics* **9**, 2566–2577 (2013).
73. Liu, *et al.* Association of human leukocyte antigen alleles and supertypes with immunogenicity of oral rotavirus vaccine given to infants in China. *Medicine* **97** (2018).
74. Wu, -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. biomedical science* **17**, 88 (2010).
75. Wang, *et al.* Recombinant heat shock protein 65 carrying PADRE and HBV epitopes activates dendritic cells and elicits HBV-specific CTL responses. *Vaccine* **29**, 2328–2335 (2011).
76. Funderburg, *et al.* Human β-defensin-3 activates professional antigen-presenting cells via Toll-like receptors 1 and 2. *Proc. Natl. Acad. Sci.* **104**, 18631–18635, DOI: 10.1073/pnas.0702130104 (2007).
77. Ferris, L. K. *et al.* Human beta-defensin 3 induces maturation of human Langerhans cell–like dendritic cells: an antimicrobial peptide that functions as an endogenous adjuvant. *Investig. Dermatol.* **133**, 460–468 (2013).
78. Mackenzie-Dyck, S., Kovacs-Nolan, J., Snider, M., Babiuk, L. A. & Others. 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 *Clin. Vaccine Immunol.* **21**, 463–477 (2014).
79. Davidson, D. *et al.* Characterisation of the transcriptome and proteome of sars-cov-2 using direct rna sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. *BioRxiv* (2020).
80. Khan, M. A. A. *et al.* An immunoinformatic approach driven by experimental proteomics: in silico design of a subunit candidate vaccine targeting secretory proteins of leishmania donovani *Parasites & Vectors* **13**, 1–21 (2020).
81. Maeshima, & Fernandez, R. Recognition of lipid a variants by the tlr4-md-2 receptor complex. *Front. cellular infection microbiology* **3**, 3 (2013).
82. Katoh, K., Rozewicki, J. & Yamada, D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings bioinformatics* **20**, 1160–1166 (2019).
83. Academic Programs | Tableau Software (2020).

84. Robinson, *et al.* The IPD and IMGT/HLA database: allele variant databases. *Nucleic acids research* **43**, D423—-D431 (2015).
85. Manczinger, M. *et al.* Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLoS biology* **17**, e3000131 (2019).
86. Niehrs, A. & Altfeld, M. Regulation of NK-cell function by HLA class II. *Cell. Infect. Microbiol.* **10** (2020).
87. Jensen, K. K. *et al.* Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**, 394–406 (2018).
88. Jurtz, *et al.* NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *The J. Immunol.* **199**, 3360–3368 (2017).
89. O'Donnell, J. *et al.* MHCflurry: open-source class I MHC binding affinity prediction. *Cell systems* **7**, 129–132 (2018).
90. Calis, J. J. A. *et al.* Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS computational biology* **9** (2013).
91. Berzofsky, J. A. Intrinsic and Extrinsic Factors in Protein Antigenic Structure. *Science* **229**, 932–940 (1985).
92. Doytchinova, I. A. & Flower, D. R. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit *BMC bioinformatics* **8**, 4 (2007).
93. Dimitrov, , Bangov, I., Flower, D. R. & Doytchinova, I. AllerTOP v. 2—a server for in silico prediction of allergens. *J. molecular modeling* **20**, 2278 (2014).
94. Gupta, S. *et al.* In silico approach for predicting toxicity of peptides and proteins. *PloS one* **8** (2013).
95. Kurcinski, *et al.* Cabs-dock standalone: a toolbox for flexible protein–peptide docking. *Bioinformatics* **35**, 4170–4172 (2019).
96. Xue, L. C., Rodrigues, J. , Kastritis, P. L., Bonvin, A. M. & Vangone, A. Prodigy: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).
97. Saha, & Raghava, G. P. S. Prediction of continuous b-cell epitopes in an antigen using recurrent neural network. *Proteins: Struct. Funct. Bioinforma.* **65**, 40–48 (2006).
98. Kringelum, J. , Lundegaard, C., Lund, O. & Nielsen, M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS computational biology* **8** (2012).
99. Saha, S. & Raghava, G. S. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic acids research* **34**, W202—-W209 (2006).
100. Dimitrov, I., Naneva, L., Doytchinova, I. & Bangov, AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics* **30**, 846–851 (2014).
101. Walker, M. *The proteomics protocols handbook* (Springer, 2005).
102. Solanki, , Tiwari, M. & Tiwari, V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. *Sci. reports* **9**, 1–19 (2019).
103. Geourjon, C. & Deleage, G. SOPM: a self-optimized method for protein secondary structure prediction. *Protein Des. Sel.* **7**, 157–164 (1994).
104. Buchan, D. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic acids research* **47**, W402—-W407 (2019).
105. Song, *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–1742 (2013).
106. Laskowski, A., Jabłon´ska, J., Pravda, L., Va vreková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein science* **27**, 129–134 (2018).

107. Colovos, & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science* **2**, 1511–1519 (1993).

108. Bhattacharya, D., Nowotny, , Cao, R. & Cheng, J. 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic acids research* **44**, W406—-W409 (2016).

109. Heo, L., Park, H. & Seok, C. GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic acids research* **41**, W384—-W388 (2013).

110. Kozakov, *et al.* The ClusPro web server for protein–protein docking. *Nat. protocols* **12**, 255 (2017).

111. Pettersen, E. *et al.* Ucsf chimera—a visualization system for exploratory research and analysis. *J. computational chemistry* **25**, 1605–1612 (2004).

112. Case, D. A. *et al.* AmberTools 16, University of California, San Francisco, 2016. *Google Scholar There is no corresponding record for this reference* (2020).

113. Maier, J. A. *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *chemical theory computation* **11**, 3696–3713 (2015).

114. Jorgensen, L. I. Am. Chem. Sot. 106 (1984) 6638. WL Jorgensen, J. Chandrasekhar, JD Madura, RW Impey and ML Klein. *J. Chem. Phys* **79**, 0 (1983).

115. Harvey, J., Giupponi, G. & Fabritiis, G. D. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. chemical theory computation* **5**, 1632–1639 (2009).

116. Ryckaert, J. , Ciccotti, G. & Berendsen, H. J. C. Numerical integration of a System with Constraints: of the Cartesian Equations of Motion Molecular Dynamics of n-Alkanes. *J Comput. Phys* **23**, 327–341 (1977).

117. Le Grand, S., Götz, A. & Walker, R. C. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380 (2013).

118. Roe, D. R. & Cheatham III, E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. chemical theory computation* **9**, 3084–3095 (2013).

119. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in escherichia coli: advances and challenges. *microbiology* **5**, 172 (2014).

120. Grote, A. *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic acids research* **33**, W526—-W531 (2005).

121. Blander, J. M. Coupling Toll-like receptor signaling with phagocytosis: potentiation of antigen presentation. *Trends immunology* **28**, 19–25 (2007).

122. Hegde, R., Chevalier, M. S. & Johnson, D. C. Viral inhibition of MHC class II antigen presentation. *Trends immunology* **24**, 278–285 (2003).

123. Bartlett, B. L., Pellicane, A. J. & Tyring, K. Vaccine immunology. *Dermatol. therapy* **22**, 104–109 (2009)

Tables

Peptide		Protein of Origin	Sequence	Position	IEBD-90% (ID)	Experimental	Vaxijen	AllerTOP	Alelos-HLA
1	AAAYYVGYLQPRTFL	SP	1	262-276			0.48	NO allergen	2
2	DDSEPVLKGVKLHYT	SP	12	1259-1273	7868	MHC/B cell/ T cell	1.18	NO allergen	1
3	LVIGAVILRGHLRIA	MG	1	138-152		p MHC/B cell	0.88	NO allergen	2
4	QSIIAYTMSLGAENS	SP	10	690-704			0.57	NO allergen	4
5	QSLLIVNNATNVVIK	SP	3	115-129			0.43	NO allergen	3
6	SFRLFARTRSMWSFN	MG	1	99-113		p MHC/B cell	0.80	NO allergen	7
7	VLSFELLHAPATVCG	SP	9	512-526		p MHC/B cell	0.48	NO allergen	4
8	CISTKHFYW	ORF1-NSP4	36	3147-3155			1.90	NO allergen	3
9	FAMQMAYRF	SP	11	898-906		p MHC/B cell/ T cell	1.03	NO allergen	58
10	FLLNKEMYL	ORF1-NSP4	36	3183-3191	16737	MHC	0.44	NO allergen	42
11	GYKSVNITF	ORF1-NSP3	11	835-843			2.37	NO alergen	8
12	ITLCFTLKR	ORF7a	1	110-118			2.02	NO allergen	12
13	KRAKVTSAM	ORF1-NSP8	43	4022-4030			0.76	NO allergen	18
14	KVKYLYFIK	ORF1-NSP9	44	4225-4233	34083	MHC	1.06	NO allergen	7
15	LEMELTPVV	ORF1-NSP3	15	1012-1020			1.97	NO allergen	31
16	MPYFFTLLL	ORF1-NSP3	26	2169-2177			0.49	NO allergen	66
17	VMYASAVVL	ORF1-NSP6	42	3684-3692		p MHC	0.48	NO allergen	39
18	WTAGAAAYY	SP	8	258-266			0.63	NO allergen	34

Table 1. Potential promiscuous vaccine peptides for the most frequent HLA-I and HLA-II alleles in LATAM. Protein of origin:
Refers to the proteins identified from the proteomes obtained from NCBI. Sequence: Refers to the preserved sequence number of the source protein in the dataset used. Position: refers to the first and last amino acid residues in the range occupied by the
peptide in the protein of origin. IEBD: Refers to peptide identification in IEBD base at a blast of 90%. Experimental: Refers to
the experimental studies found in which the peptide is recognized as a target of HLA, BL, or TL molecules.
A (p) is also added when the sequence was partially part of one of the experimental sequences.
Vaxijen: Antigenicity described by a threshold greater than 0.4. Allergenic: Prediction if the peptide is a probable allergen. HLA-alleles: Refers to the number of alleles covered by the peptide.

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 hours	30 hours
Aliphatic index	81.81	82.62
Size (amino acids)	510	499

Table 2. Physicochemical characteristics of the multi-epitope construct. The default thresholds for ToxinPred, Vaxijen> 0.4,
and allergen FP V1.0 were chosen to be the best balance between specificity, sensitivity, and precision while taking account of linear and non-linear reasons in the results. In addition, the physicochemical characteristics, construct, antigenic, non-allergenic, non-toxic, soluble, and stable characteristics are demonstrated suggesting that an interaction with the immune system can allow and maintain recognition by the innate immune system, as well as by BL.

Figures

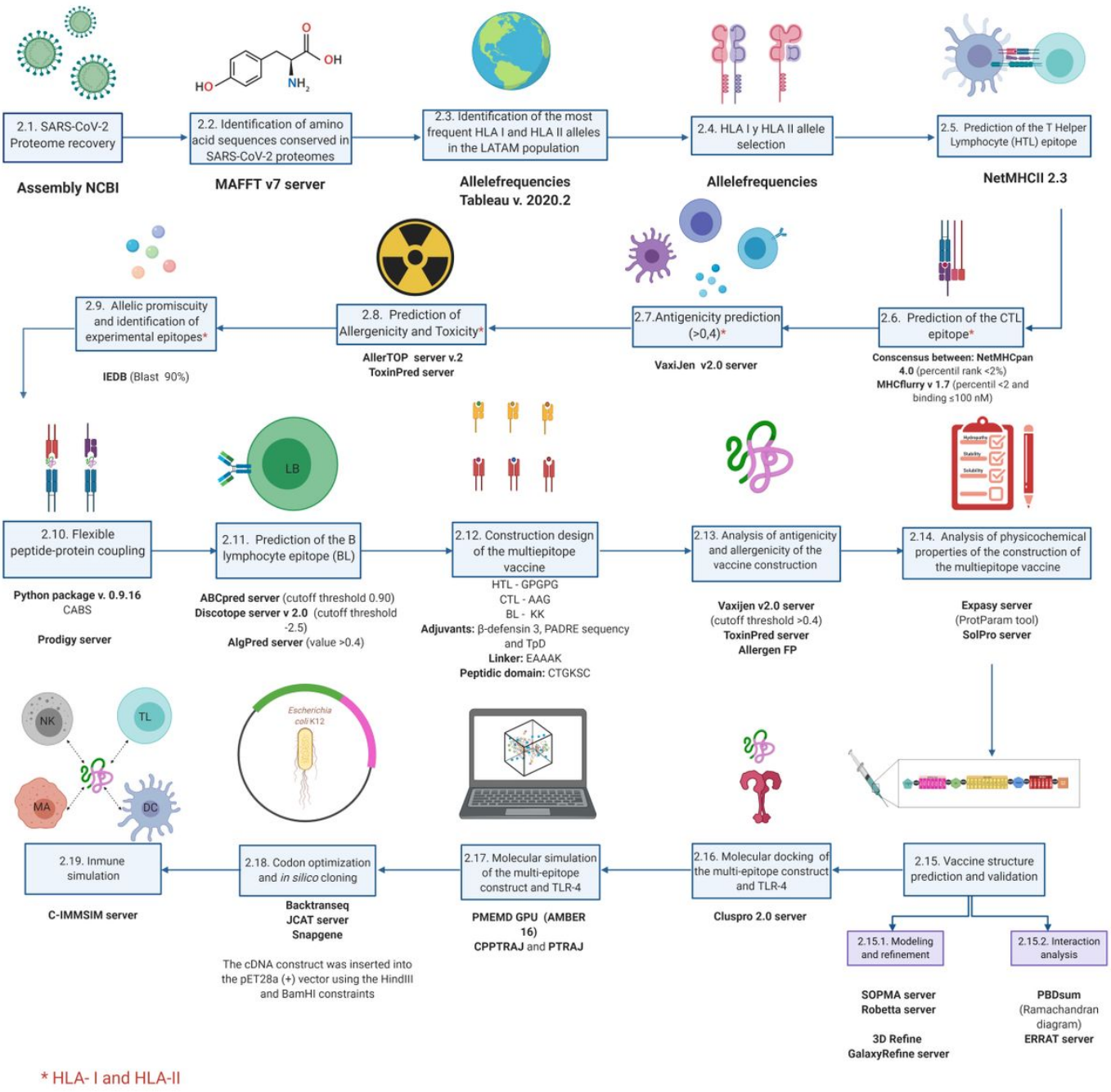

Figure 1

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.

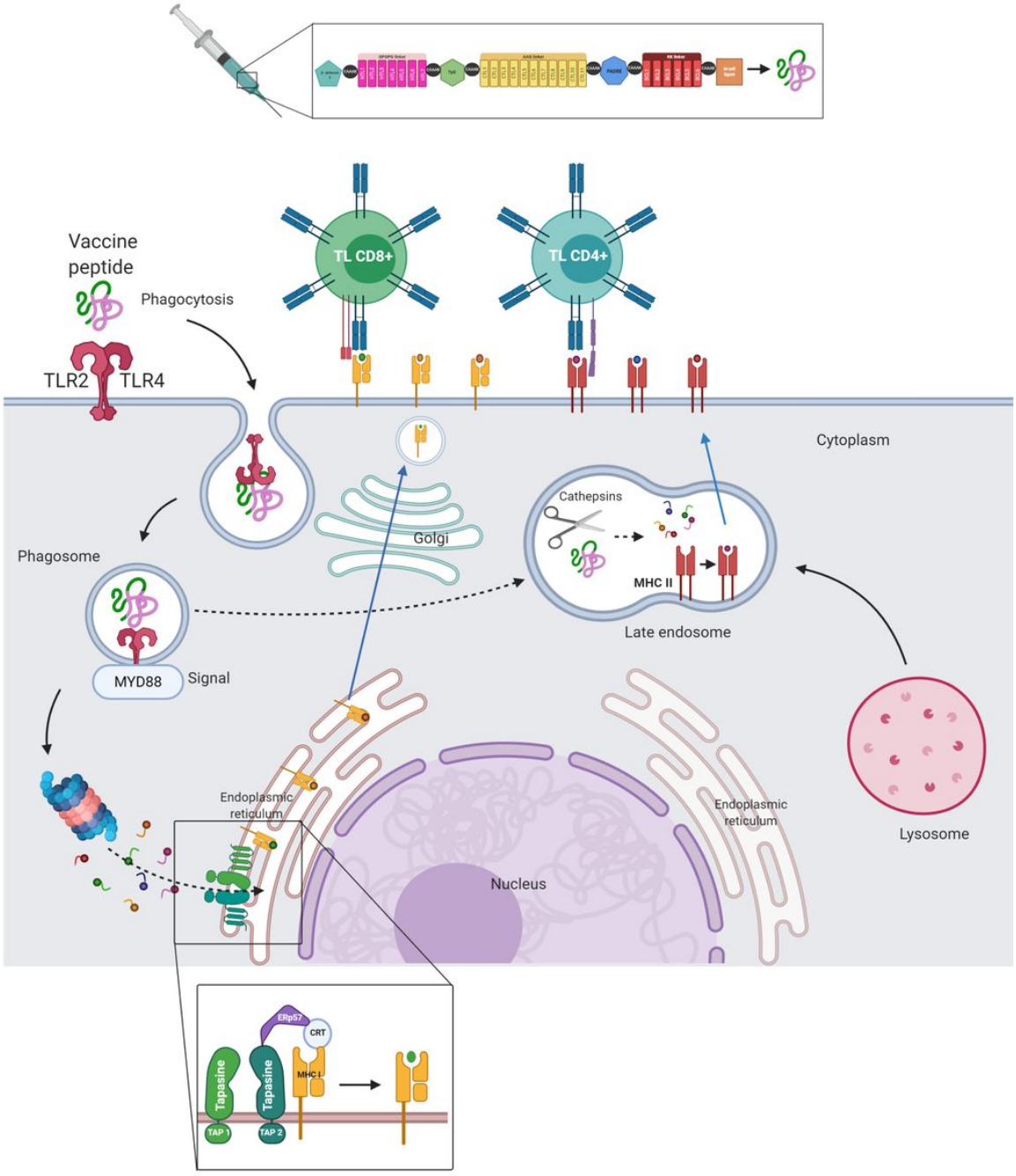

Figure 2

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 phagocytised by the cell together with TRL, 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 internalisation 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

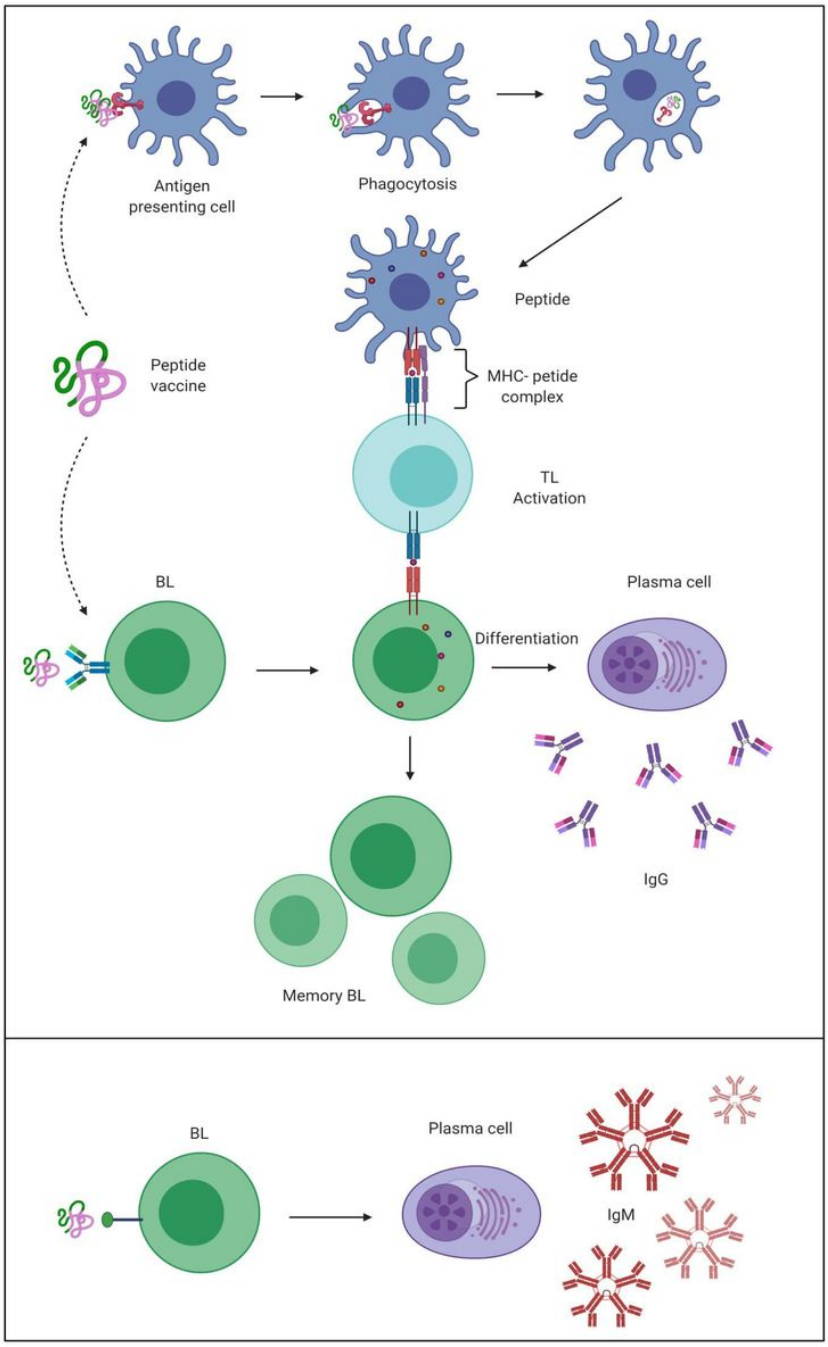

Figure 3

Effect of the multi-epitope construct, including predictions of B lymphocyte epitopes, on antigen presentation and development of cellular and humoral immunity: two types of immunological responses to the vaccine can coexist: T-dependent and non-T-dependent responses. In the T-dependent response (upper panel), the antigenic peptide based on viral proteins is recognised and phagocytised in the antigen-presenting cell (the dendritic cell). Once processed at the intracellular level, the peptide bound to the HLA molecules is expressed at the membrane level. CPA presents the antigen to HTL type 2. Once recognised, T lymphocytes (TLs) are activated and generate cytokines (IL-4, IL-5 and IL-6) that promote B lymphocyte (BL) differentiation in plasma cells, thereby promoting the generation of specific immunoglobulin G (IgG) and BL memory. In the non-T-dependent response (lower panel), the peptide is directly recognised by BL for its immunoglobulin M (IgM) receptor at the level of lymphoid tissues; once processed, the BL produces and secretes IgM

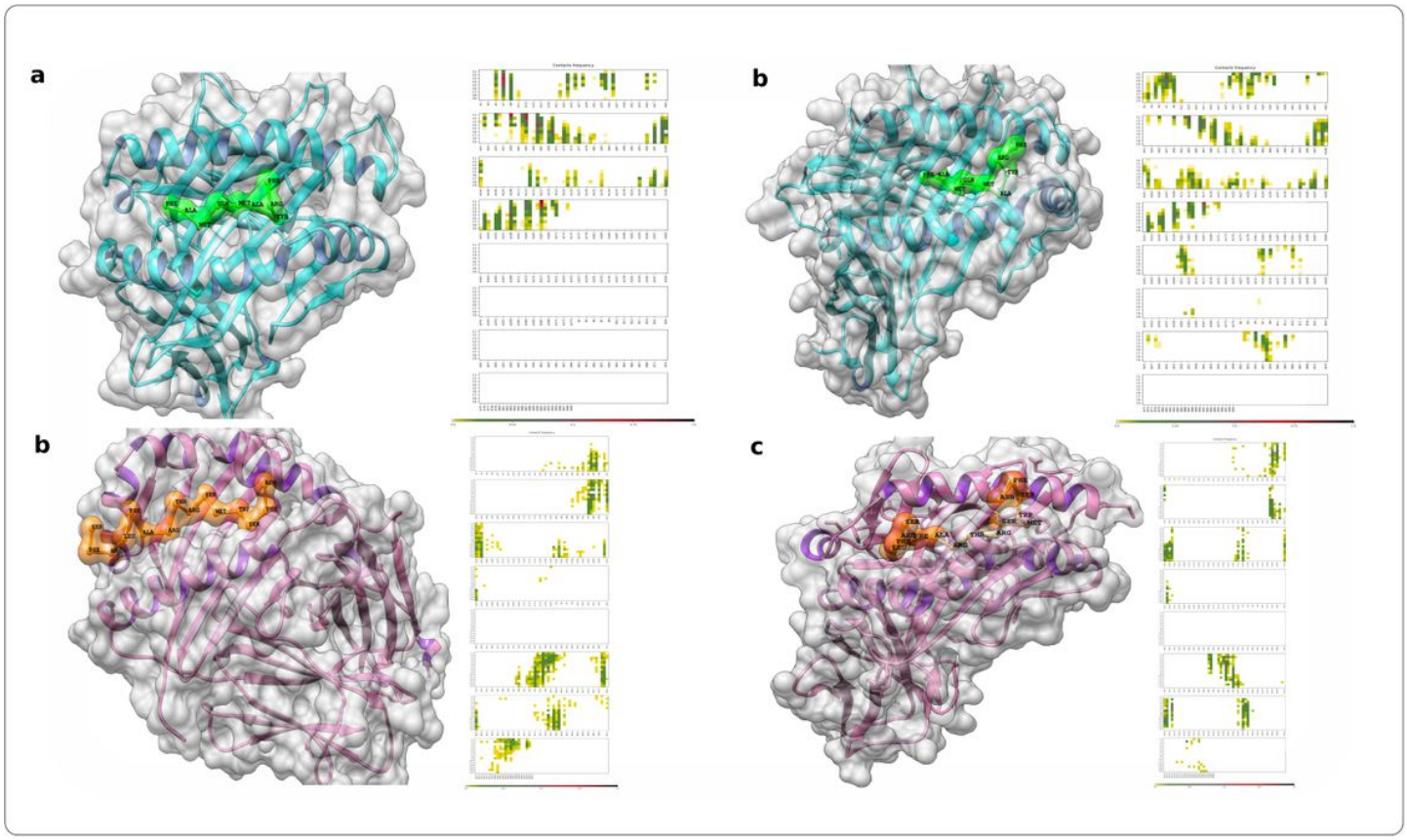

Figure 4

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. \textbf{a,} Cluster 1, density of 239, ΔG of -5.2. \textbf{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. \textbf{c,} Cluster 1, density of 138, ΔG of -4.6. \textbf{d,} Cluster 2, density of 168, ΔG of -4.4.

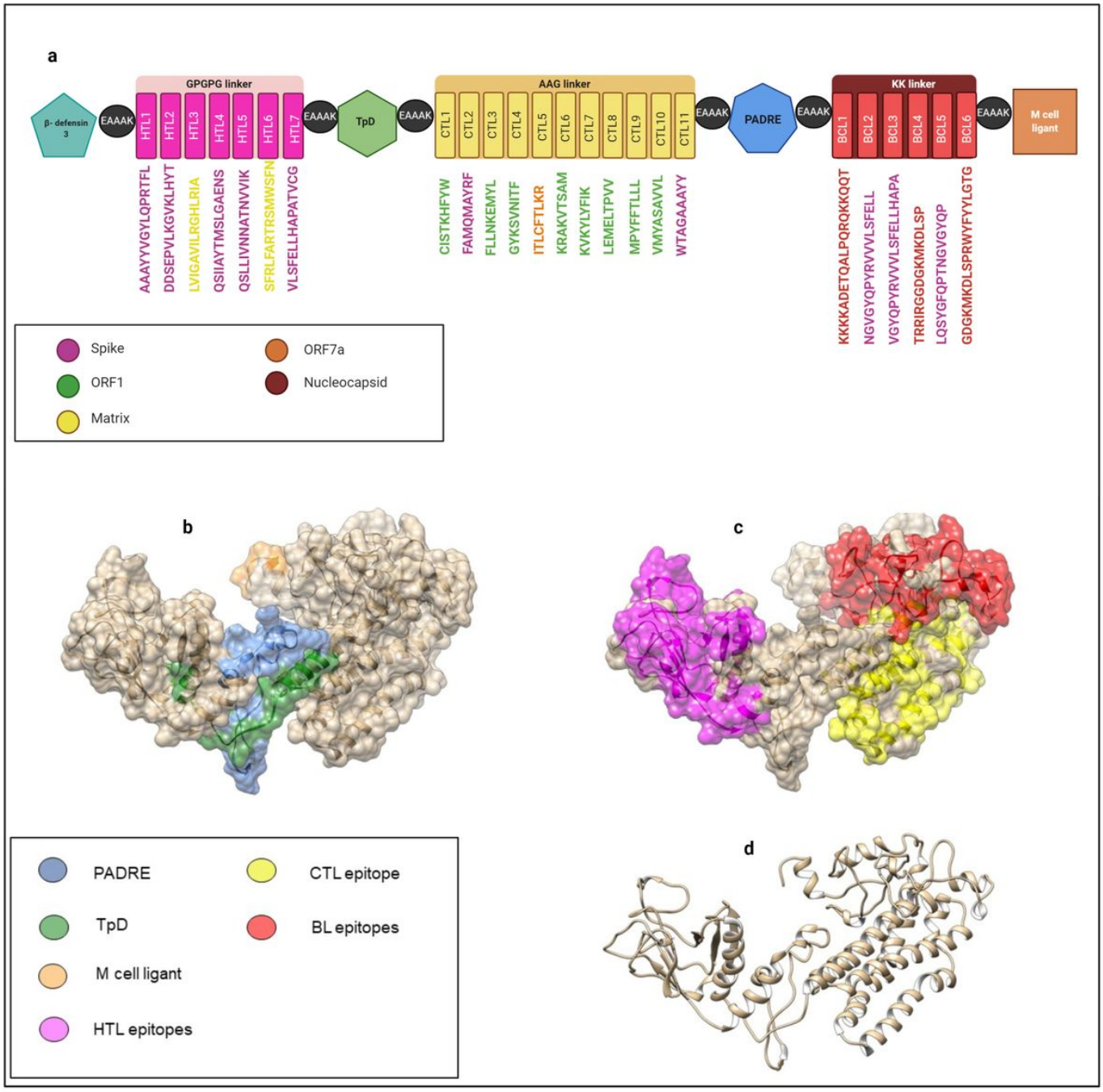

Figure 5

Structure of the multi-epitope construct: three-dimensional diagrammatic structure of the multi-epitope construct showing \textbf{a,} the separators, origin of the peptides and adjuvants used; \textbf{b,} the location of the adjuvants PADRE, TpD and M cell ligand; and \textbf{c,} the location of the epitope's cytotoxic T lymphocytes (CTLs), T helper lymphocytes (HTLs) and B lymphocytes (BLs). \textbf{d,} Ribbon diagram of multiepitope construct.

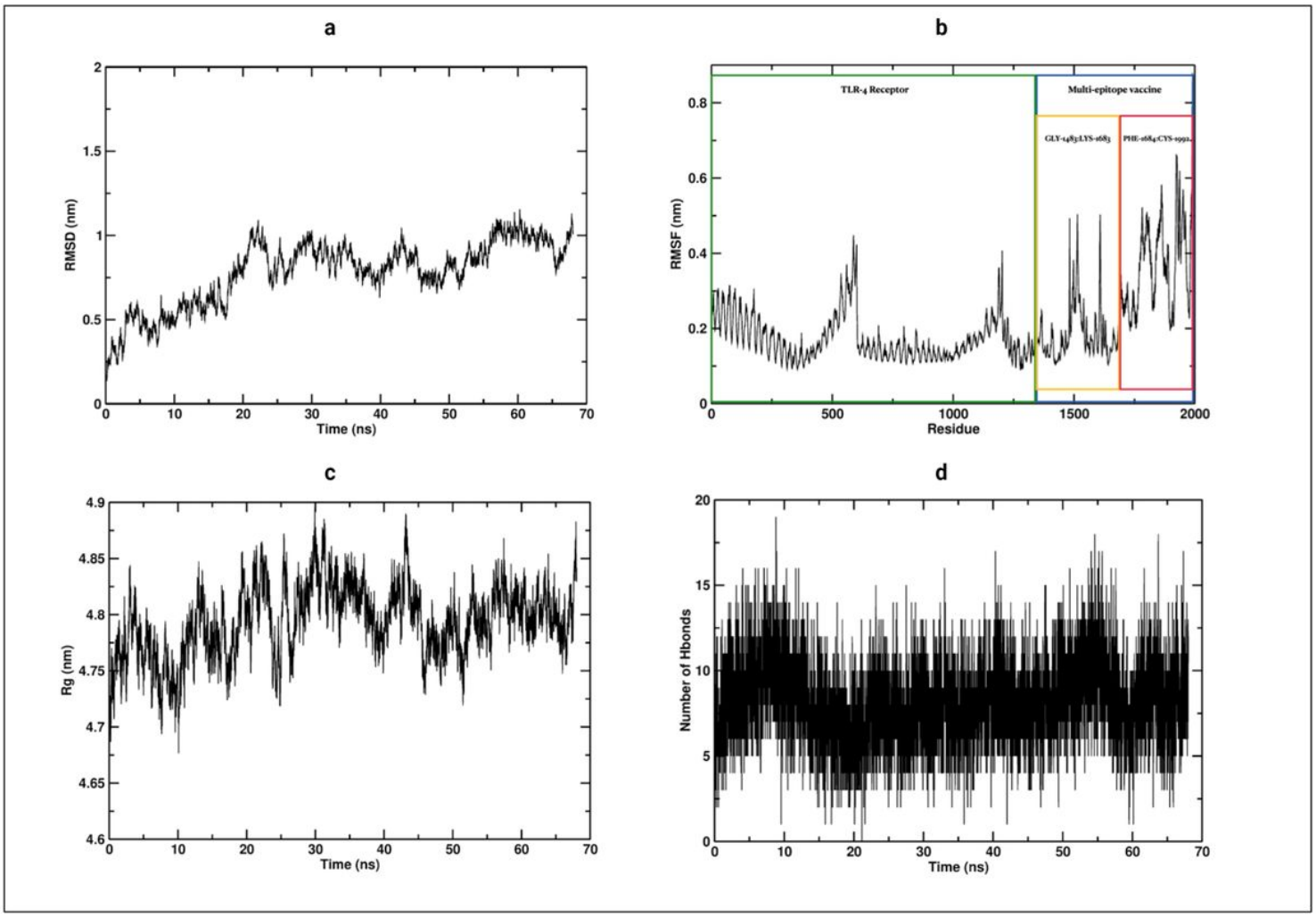

Figure 6

Analysis of trajectory.\textbf{a,} Root mean square deviation (RMSD) Cα atoms. \textbf{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. \textbf{c,} Rg plot; vaccine construct is stable in its compact form during the simulation time. \textbf{d,} Changes in the number of hydrogen bonds between the TLR-4 receptor and multi-epitope vaccine molecule during MD simulation.

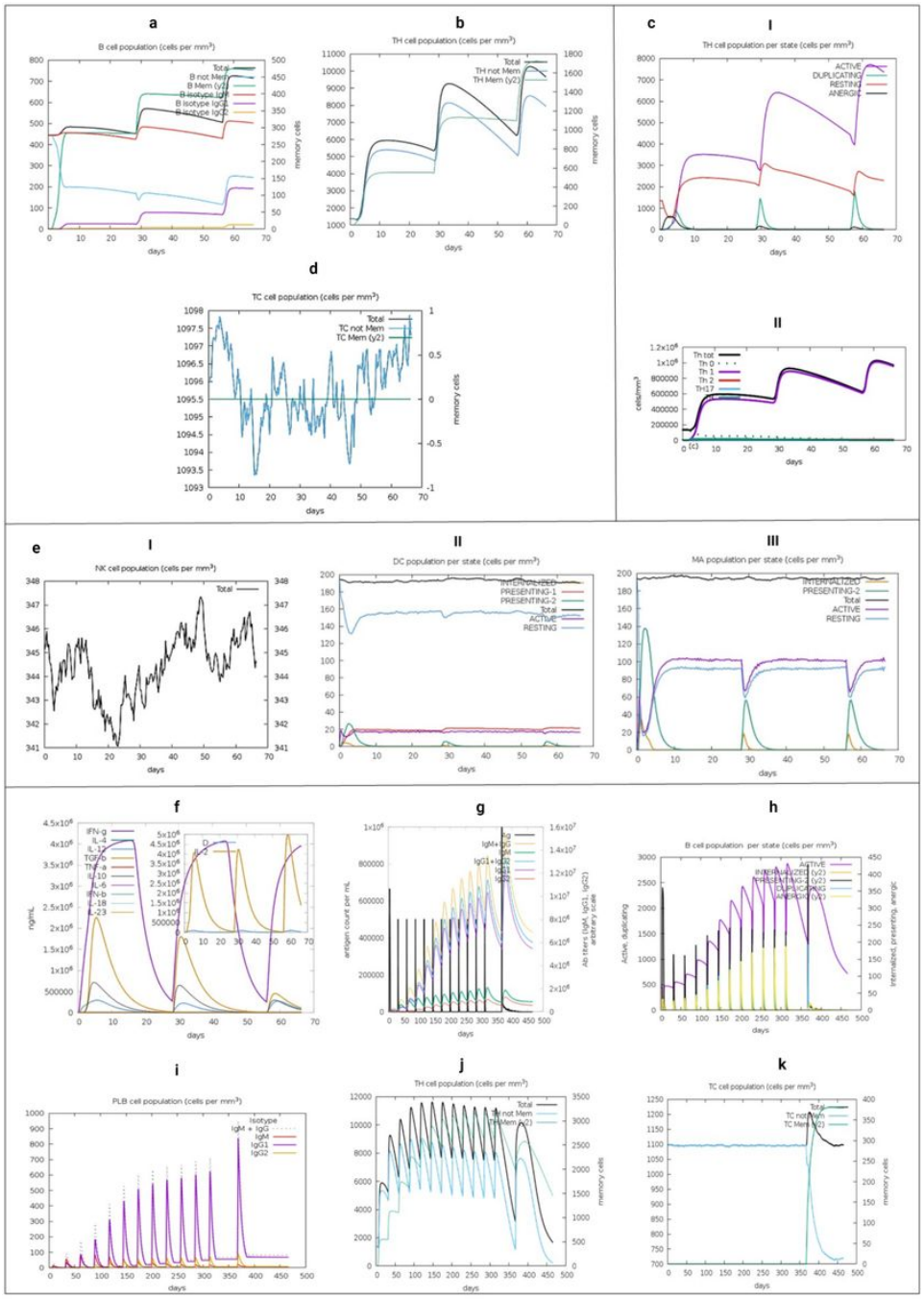

Figure 7

Immune simulation using the multi-epitope construct. \textbf{a,} B cells, \textbf{b,} CD4+ T cells and \textbf{d,} CD8+ T cells were simulated, presenting a cumulative effect towards the third injection on the 56th day of simulation. This suggests the early presence of T cell memory and a change towards the immunoglobulin isotype, immunoglobulin G (IgG), being more predominant. While the antigenic stimulus lasts, there is a polarisation towards a certain type of response \textbf{c,} HTL-1, which is consistent with \textbf{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 \textbf{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. \textbf{g,} An indifference in immunoglobulin M (IgM) production and increased IgG response suggests favourable conditions for viral antibody clearance in the B cell. \textbf{h,} The duplication and antigenic presentation of B cells corresponds to the

stimulated response by the simulated virus, which lasts beyond the viral challenge. \textbf{i,} An IgG isotype shows a substantial response to the viral challenge, with subtype IgG1 most notably responding. \textbf{j,} The CD4+ T cell 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. \textbf{k,} The population of memory CD8+ cells are stimulated by the viral challenge, acting directly on viral clearance.

Supplementary Files

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

- MULTIEPITOPEPEPTIDEVACCINEAGAINSTSARSCOV2SUPPLEMENTARY2.pdf