



## Immunoinformatics study to search epitopes of spike glycoprotein from SARS-CoV-2 as potential vaccine

Ramírez-Salinas Gema Lizbeth<sup>a</sup>, García-Machorro Jazmín<sup>b</sup>, Correa-Basurto José<sup>a</sup> and Martínez-Archundia Marlet<sup>a</sup>

<sup>a</sup>Laboratorio de Diseño y Desarrollo de Nuevos Fármacos e Innovación Biotecnológica (Laboratory for the Design and Development of New Drugs and Biotechnological Innovation), Escuela Superior de Medicina, Instituto Politécnico Nacional, México City, México; <sup>b</sup>Laboratorio de medicina de Conservación, Escuela Superior de Medicina, Instituto Politécnico Nacional, México City, México

Communicated by Ramaswamy H. Sarma

### ABSTRACT

The Coronavirus disease named COVID-19 is caused by the virus reported in 2019 first identified in China. The cases of this disease have increased and as of June 1<sup>st</sup>, 2020 there are more than 216 countries affected. Pharmacological treatments have been proposed based on the resemblance of the HIV virus. With regard to prevention there is no vaccine, thus, we proposed to explore the spike protein due to its presence on the viral surface, and it also contains the putative viral entry receptor as well as the fusion peptide (important in the genome release). In this work we have employed *In Silico* techniques such as immunoinformatics tools which permit the identification of potential immunogenic regions on the viral surface (spike glycoprotein). From these analyses, we identified four epitopes E332-370, E627-651, E440-464 and E694-715 that accomplish essential features such as promiscuity, conservation grade, exposure and universality, and they also form stable complexes with MHCII molecule. We suggest that these epitopes could generate a specific immune response, and thus, they could be used for future applications such as the design of new epitope vaccines against the SARS-CoV-2.

**Abbreviations:** 3D: three-dimensional; ACE2: angiotensin-converting enzyme 2; APN: aminopeptidase N; CoVs: Coronaviruses; E: envelope; FP: fusion peptide; ICTV: International Committee on Taxonomy of Viruses; M: membrane; MERS: Middle East Respiratory Syndrome; MHC I: major histocompatibility complex class I; MHC II: major histocompatibility complex class II; MHC: major histocompatibility complex; N: nucleocapsid; P1: Pocket 1; P4: Pocket 4; P6: Pocket 6; P9: Pocket 9; PB: Pocket B; PDB: Protein Data Bank; PF: Pocket F; RBD: receptor binding domains; SARS: Severe Acute Respiratory Syndrome; TCRs: T cell receptors; WHO: World Health Organization

### ARTICLE HISTORY

Received 3 April 2020

Accepted 7 June 2020

### KEYWORDS

Epitope; vaccine; SARS-CoV-2; coronavirus; spike glycoprotein



## Introduction


Coronaviruses (CoVs) are enveloped, single positive stranded RNA viruses that belong to the subfamily Coronavirinae. The genomes of CoVs are within the range of 26 to 32 kilobases in length and is probably the largest viral RNA known (Cui et al., 2019; Fehr & Perlman, 2015). One of the main features of the Coronavirus is that it has spike projections that stand out from the virion surface and resemble to a crown shape, which is the reason of its name.

COVs are considered causal agents of several diseases in birds and mammals including, bats, bovines, mice, dogs, cats, civets, and camels (Fehr & Perlman, 2015). Coronaviruses were not considered a highly pathogenic virus until "Severe Acute Respiratory Syndrome" (SARS) was found in 2002 and 2003 in the Chinese Province of Guangdong (Cui et al., 2019). One decade after, another highly pathogenic coronavirus emerged named as the Middle East Respiratory Syndrome (MERS-CoV), affecting different

countries of the Middle East. Recently in December 2019 (Ji et al., 2020), a novel CoV identified as the pathogen causing outbreaks of illness in the Chinese city of Wuhan was firstly named as 2019-nCoV by the World Health Organization (WHO), however, later, it was renamed SARS-CoV-2 by the International Committee on Taxonomy of Viruses (ICTV). SARS-CoV-2 belongs to lineage B  $\beta$ -COVs, subgenus Sarbecovirus (Ji et al., 2020; Wang et al., 2020).

CoVs have four principal proteins in their structure, spike glycoprotein (S), membrane (M), envelope (E) and nucleocapsid (N) (Fehr & Perlman, 2015). The homotrimers of S protein constitute the spike structure on the virus surface. The spike glycoprotein mediates the binding of the receptor to the Host through RBD (Fehr & Perlman, 2015). The 3.5 Å resolution cryo-EM structure of the SARS-CoV-2 shows that its predominant state of the trimer has one of the three RBDs (Wrapp et al., 2020) Of worthy interest is that the described receptors for coronaviruses depend on the lineage, therefore,

**CONTACT** Martínez-Archundia Marlet  [marletm8@gmail.com](mailto:marletm8@gmail.com)  Laboratorio de Diseño y Desarrollo de Nuevos Fármacos e Innovación Biotecnológica (Laboratory for the Design and Development of New Drugs and Biotechnological Innovation), Escuela Superior de Medicina, Instituto Politécnico Nacional, México City, 11340, México.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2020.1780944>.

© 2020 Informa UK Limited, trading as Taylor & Francis Group

for the alpha lineage, aminopeptidase N (APN) has been described as their receptor, while SARS-CoV and HCoVNL63 use angiotensin-converting enzyme 2 (ACE2) as their receptor. In the case of MERS-CoV, it uses the protein CD26 as a Host, infecting epithelial bronchial non ciliated cells and type II pneumocytes (Cao et al., 2020; Li et al., 2020). In the case of SARS-CoV-2 it has been suggested that it could use the enzyme ACE2 as Host despite the presence of mutations in the binding domain (Li et al., 2020; Lu et al., 2020).

Additionally, spike glycoprotein has a hydrophobic segment (a “fusion loop” or “fusion peptide”) and is considered a Class I fusion protein (Fehr & Perlman, 2015). S protein activation is subjected to a diverse conformational changes and insertion of a putative fusion peptide (FP) into the target membrane (Ou et al., 2016). Thus, the study of the spike glycoprotein of the virus and the determination of structural similarity in the binding domains of the receptor of COVs (Lu et al., 2020) as well as its importance in the fusion process of the replication cycle are relevant.

SARS-CoV-2 has infected a lot of people in about 216 countries, until June 1<sup>st</sup> 2020 OMS had reported 6,194,533 people infected by the SARS-CoV-2 and 376,320 deaths (<https://www.who.int/emergencies/disease/novel-coronavirus-2019>), indicating that we are facing a new pandemic (Nishiura et al., 2020). Unfortunately, there is no treatment or vaccines. Currently, the best preventive strategy to treat diseases from diverse pathogens is vaccine application.

Vaccine application can prevent some diseases from pathogens. Currently, one of the most common strategies is the development of peptide based vaccines (immunogenic regions) by using immunoinformatics tools (Calis et al., 2019). These peptides, better known as epitopes can form MHC-epitope complexes, driving the activation of an immune response to prevent further infection.

Major histocompatibility complex (MHC) class I and MHC class II-dependent immune responses are considered important. Class I molecules (MHCI) are expressed in all nucleated cells and interact with T cell receptors (TCRs) to induce cytotoxic immune system by lymphocytes T. MHCII molecules bind antigenic peptides generated in the endocytic pathway (endosomes and phagosomes) of professional APCs (macrophages, B lymphocytes and dendritic cells) and then present to the CD4+ T cell receptor (Agudelo & Patarroyo, 2010). Both types of molecules participate in the immune response to eliminate antigens by different mechanisms and the two MHC molecules show grooves in which peptides are bound. Generally peptides within the range of 8 to 11 amino acids size commonly bind to MHCI, whereas peptides of 10–20 amino acid size bind MHCII (Agudelo & Patarroyo, 2010; Apostolopoulos et al., 2008). The MHC allele set present in each chromosome is called MHC haplotype. In this work, the haplotype class I (HLA-A\*2402 and HLA-A\*0301) and haplotype class II (HLA-DRB1\*0401 and HLA-DRB1\*0701); were included due to their response against viruses, including neutralizing antibody (CD4+ T-cell mediated) and CD8+ T-cell mediated antibodies (humoral and cellular immunity responses, respectively).

Furthermore, in this work we have used bioinformatics tools to find spike glycoprotein epitopes from SARS-CoV-2 as

possible immunogenic peptides that could be used as possible vaccines in the future. Adequate selection of these epitopes through bioinformatics tools will help to induce a more direct and potent immune response. By using this methodology, we have considered some epitopes features which have been suggested in other works, for example, promiscuity, (Saraav et al., 2016), conservation (Doytchinova & Flower, 2005; Sidney et al., 2008) and exposure to the surface of the trimeric spike glycoprotein (Liang et al., 2009).

## Methodology

### Consensus sequences of spike glycoprotein SARS-CoV-2

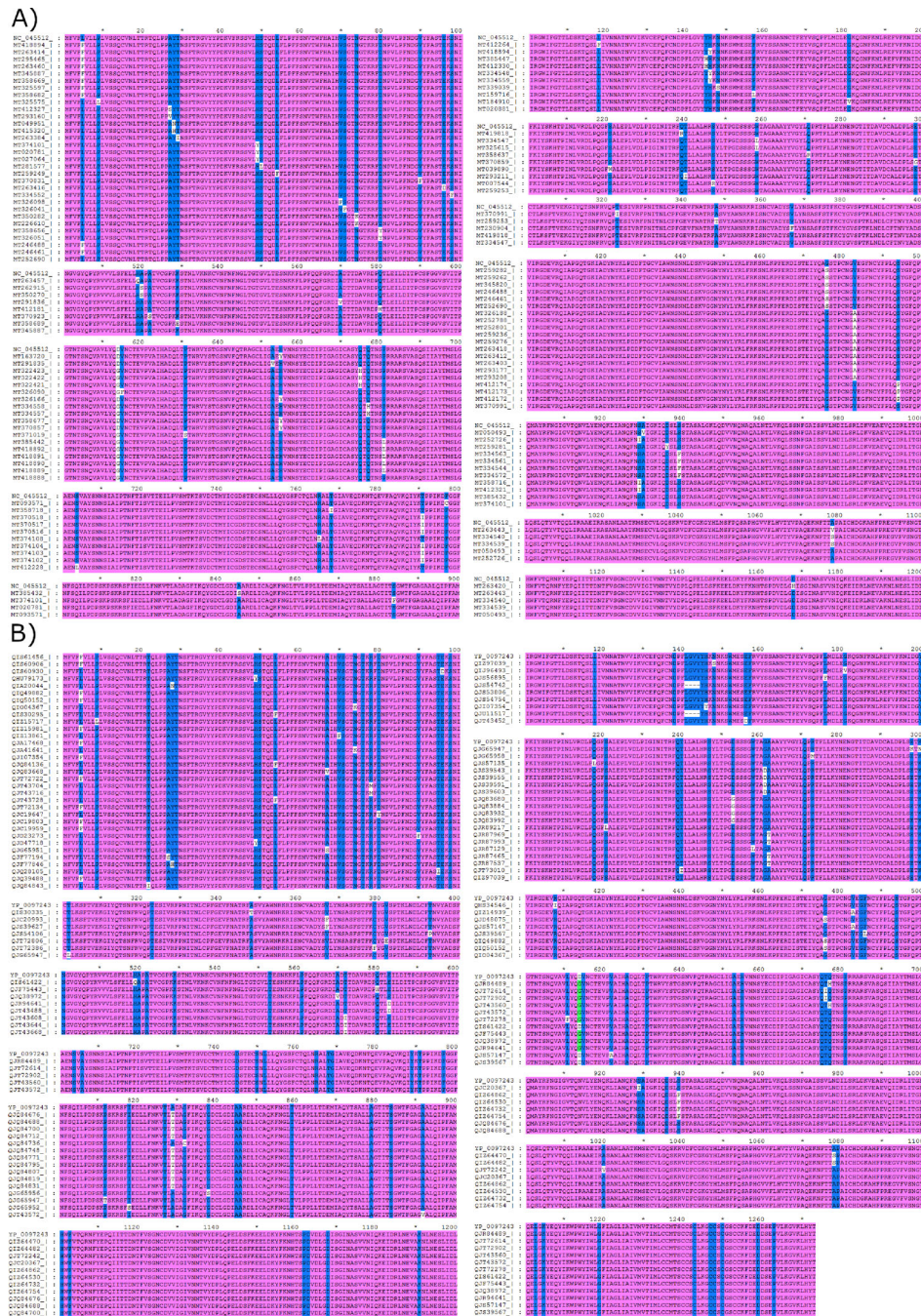
2745 protein and 1702 complete genome sequences reported sequences available on May 18<sup>th</sup> 2020 were downloaded from the COVID 19 data hub. Only complete genomes of high-coverage and the ones that do not have regions with clearly identified amino acids or nucleotides (NN or XX) were sequences included in this study. Further multiple alignments were carried out by Clustal X Program by employing default parameters provided by the Program (Larkin et al., 2007). Afterwards, complete genome sequence alignment was translated to protein sequences. Sequences with the following IDs NC\_045512 and YP\_009724390.1 were considered as reference sequences (refseq) since they are both available at the COVID 19 data hub.

### Immunoinformatics studies of the spike glycoprotein SARS-CoV-2

From the consensus sequence of the SARS-CoV-2 spike glycoprotein, HLA supertypes molecules of MHCI and MHCII proteins were identified. The affinity of each of the peptides to HLA-Supertype molecules was determined employing the following servers: NetMHC 4.0 and NetMHCII 2.3. NetMHC 4.0 server was used to predict 10 mer peptides which could have affinity to HLA supertypes of MHCI proteins (Andreatta & Nielsen, 2016) whereas NetMHCII 2.3 (Jensen et al., 2018) was employed to predict 25mer peptides which could show affinity for HLA Supertypes for MHCII proteins.

### Molecular modeling of the spike glycoprotein SARS-CoV-2

At the time we prepared the manuscript, there was no crystal structure in Protein Data Bank (Rose et al., 2017), thus, a three-dimensional (3D) model of the protein was built employing molecular modeling techniques and the Program Swissmodel (<https://swissmodel.expasy.org/>) (Waterhouse et al., 2018). From the NCBI database, the primary sequence of SARS-CoV-2 spike glycoprotein (ID Sequence: YP\_009724390.1) was found. This sequence was used to build a 3D model of the SARS-CoV-2 spike glycoprotein by using the SwissModel Software. In this process, the crystal structure (PDB: 6ACD) of the SARS-CoV spike glycoprotein structure was employed as template to build the SARS-CoV-2 spike glycoprotein. The SARS-CoV spike glycoprotein has an identity of 76.47% with SARS-CoV-2. The predictive 3D model from the consensus sequence was named SCons and



**Figure 1.** Alignment of spike glycoprotein sequences of SARS-CoV-2. A) Multiple sequence alignments of the complete genome sequences exhibit 81 amino acid changes (a fragment of the alignment is shown). B) Multiple protein sequences of the spike glycoprotein of SARS-CoV-2 (protein sequences) exhibit 164 amino acid changes (a fragment of the alignment is shown). Amino acids that are identical are shown in pink color. Variations of the amino acids are shown in blue color.

represents the structure of the SARS-CoV-2 spike glycoprotein. Furthermore this model was validated using the Software PDBSum (Laskowski et al., 2018).

Additionally, at the time we were preparing the manuscript, two crystal structures that represent distinct conformational states of the spike glycoprotein were released. The PDB ID: 6VYB represents the up conformation, whereas PDB ID: 6VXX constitutes the down conformation. Both of the crystals showed missing fragments, thus, these were completed by employing molecular modeling techniques, using the Program Modeller 9.23 (script segment.py) (Webb & Sali, 2016). After this process, the resultant models were denoted as 6VXX-fill and 6VYB-fill.

### Epitope selection

In the process of epitope selection, only peptides that showed the following features: promiscuity, protein surface exposure and conservation grade were selected.

### Molecular modeling of the MHC I and MHC II molecules

In the PDB database we searched for the structures of the MHC-II proteins; for the HLA-DRB1\*0401 molecule, the crystal with the PDB ID: 6BIJ, for HLA-A\*2402 we used the crystal structure PDB ID: 3WOW, and for HLA-A\*0301 we used the structure PDB ID: 2XPG whereas for the HLA-DRB1\*0701

**Table 1.** Promiscuous epitopes. This table describes which molecules of HLA supertype show affinity to the epitopes by using NetMHC 4.0 and NetMHCII 2.3 servers. The affinity is indicated in parenthesis as either weak (WB) or strong (SB).

Promiscuous epitopes			
ID	Sequence	MHC I	MHCII
E136-167	CNDPFLGVVYHKNNKSWMESE FRVYSSANNCT	HLA-A*0301 (WB), HLA-A*2402 (SB), HLA-A*0101 (WB).	HLA-DRB1*0401 (WB), HLA-DRB1*0701 (WB), HLA-DRB1*0801 (WB), HLA-DRB1*0901 (SB), HLA-DRB1*1101 (WB), HLA-DRB1*1301 (WB), HLA-DRB1*1501 (WB).
E199-225	GYFKIYSKHTPINLVRDLPQG FSALEP	HLA-A*0301 (WB), HLA-A*2402 (WB), HLA-B*0801 (WB).	HLA-DRB1*0101 (SB), HLA-DRB1*0301 (SB), HLA-DRB1*0401 (WB), HLA-DRB1*0701 (SB), HLA-DRB1*0801 (SB), HLA-DRB1*0901 (WB), HLA-DRB1*1101 (SB), HLA-DRB1*1301 (WB), HLA-DRB1*1501 (WB).
E238-262	FQTLALHRSYLTPGDSSSG WTAGA	HLA-B*5801 (SB).	HLA-DRB1*0101 (WB), HLA-DRB1*0401 (WB), HLA-DRB1*0701 (WB), HLA-DRB1*0801 (WB), HLA-DRB1*1301 (SB), HLA-DRB1*1501 (WB).
E332-370	ITNLCPFGEVFNATRFASVYA WNRKRISN CVADYSVLYN	HLA-A*2601 (SB), HLA-A*0301 (WB), HLA-A*2402 (WB), HLA-B*0801 (WB), HLA-A*0101 (SB).	HLA-DRB1*0101 (WB), HLA-DRB1*0401 (WB), HLA-DRB1*0701 (WB), HLA-DRB1*0801 (SB), HLA-DRB1*0901 (WB), HLA-DRB1*1101 (SB), HLA-DRB1*1301 (SB), HLA-DRB1*1501 (WB).
E440-464	NLDSKVGGNYYLRLFRKS NLKPF	HLA-A*0301 (SB), HLA-A*2402 (WB), HLA-B*0801 (WB), HLA-B*1501 (WB).	HLA-DRB1*0101 (WB), HLA-DRB1*0701 (SB), HLA-DRB1*0801 (SB), HLA-DRB1*1101 (SB), DRB1-1201 (WB), HLA-DRB1*1301 (SB), HLA-DRB1*1501 (WB).
E481-506	NGVEGFNCYFPLQSYGFQP TNGVGY	HLA-A*0101 (WB), HLA-B*4001 (WB), HLA-A*2402 (WB), HLA-B*1501 (WB)	HLA-DRB1*0101 (WB), HLA-DRB1*0401 (SB), HLA-DRB1*0701 (WB), HLA-DRB1*0901 (WB), HLA-DRB1*1501 (WB).
E627-651	DQLTPTWRVYSTGSNVFQTR AGCLIG	HLA-A*0301 (WB), HLA-A*2402 (SB).	HLA-DRB1*0101 (WB), HLA-DRB1*0401 (SB), HLA-DRB1*0701 (SB), HLA-DRB1*0801 (WB), HLA-DRB1*0901 (SB), HLA-DRB1*1301 (WB).
E153-180	MESEFRVYSSANNCTFEYVS QPFLMDLE	HLA-A*0301 (WB), HLA-A*2402 (SB), HLA-A*0101 (SB), HLA-A*2601 (WB), HLA-B*4001 (WB).	HLA-DRB1*0101 (WB), HLA-DRB1*0401 (WB), HLA-DRB1*0701 (WB), HLA-DRB1*0801 (WB), HLA-DRB1*0901 (SB), HLA-DRB1*1501 (SB).
E694-715	AYTMSLGAENSVAYSNNNSIA IPTNF	HLA-A*0101 (SB), HLA-A*2601 (SB), HLA-B*1501 (WB), HLA-B*4001 (WB), HLA-A*2402 (WB).	HLA-DRB1*0101 (WB), HLA-DRB1*0401 (WB), HLA-DRB1*0901 (SB).
E911-939	VTQNVLVENQKLIANQFNSAI GKIQDSLS	HLA-A*2402 (WB), HLA-A*0301 (SB).	HLA-DRB1*0101 (WB), HLA-DRB1*0301 (WB), HLA-DRB1*0401 (WB), HLA-DRB1*0901 (WB), HLA-DRB1*1101 (WB), DRB1-1201 (WB), HLA-DRB1*1301 (WB), HLA-DRB1*1501 (WB).
E811-848	PSKRSFIEDLLFNKVTLAD AGFIKQYGDCLGDIAARDL	HLA-A*0301 (WB), HLA-B*4001 (WB), HLA-A*0101 (WB).	HLA-DRB1*0101 (SB), HLA-DRB1*0301 (WB), HLA-DRB1*0401 (WB), HLA-DRB1*1501 (WB).

molecule, we did not find a structure in the PDB, so we modeled the structure by homology modeling using Program Modeller 9.23 (Webb & Sali, 2016). We employed the crystal structure PDB ID: 3PDO as a template, with an identity of 100% for the alpha chain and an identity of 89.90% for the beta chain.

### Molecular modeling of the selected epitopes

Three dimensional (3D) structures of the chosen epitopes were built by employing PEP-FOLD 3 server which employs the *novo* peptide structure prediction strategy based on a Hidden Markov Model derived Structural Alphabet (SA) (Lamiable et al., 2016).

### Protein-Protein docking studies

MHCI-II proteins were docked with the chosen epitopes using the ClusPro 2.0 Server (Vajda et al., 2017). With regards to this Program, epitopes can be considered as ligands, whereas molecules such as HLA-DRB1\*0401, HLA-A\*2402, HLA-A\*0301 and HLA-DRB1\*0701 can be determined as receptors. Cluspro server employs distinct algorithms which permit to evaluate millions of predictive complexes by taking account their electrostatic energy as well as desolvation parameters (Larkin

et al., 2007). Subsequently, some MHC-epitope complexes which show the highest affinity were selected.

### Molecular dynamics simulations of the MCH II-epitope complexes

Eleven Molecular dynamics simulations of the MCH II epitopes-complexes were carried out by means of the NAMD Program and using the known inputs for NAMD standard MD simulations (Phillips et al., 2005). Complexes were solvated and the total charge was neutralized up to a final concentration of 0.15 M NaCl.

MD simulations were performed under periodic boundary conditions (PBC). The method of Particle-Mesh Ewald (PME) was used for the calculation of the electrostatic potential energy. A no bonded cutoff of 12 Å, switchdist of 10 Å and pairlistdist of 16 Å were implemented for these long range interactions.

In the equilibration step, a NTV (constant volume and temperature) protocol was applied. For the production step a NTP (constant temperature and pressure) ensemble was maintained with Langevin thermostat (310 K) and anisotropic Langevin barostat (1 atm). For these last two steps, a integration time step of 2 femtosecond (fs) was used, and all the bond lengths involved. Finally, MD simulations were run for about 10 ns.

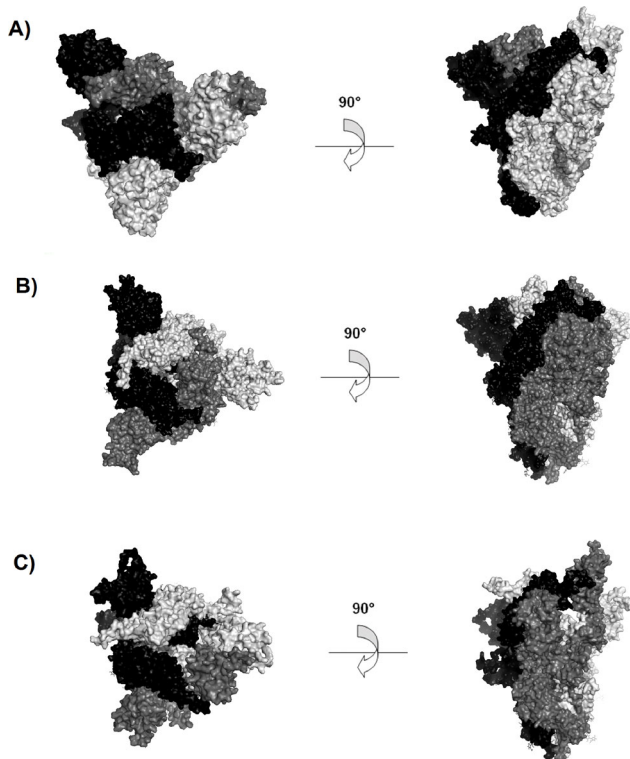
**Table 2.** Conserved epitopes: This table shows the amino acid changes and the frequency percentage of changes of the epitopes obtained from both alignments, between protein and nucleotide sequences respectively. (NC): indicates No changes, (AP): alignment of the protein sequences, (AN): alignment of the nucleotide sequences.

Conserved epitopes					
ID	Consensus sequence	Changes (AP)	Frequency (%)	Changes (AN)	Frequency (%)
E136-167	CNDPFLGVYYHK NNKSWMESEFRVY SSANNCT	D138H	0.07285	Y145H	0.11750
		L141-	0.07285	H146Y	0.17626
		G142-	0.07285	N148S	0.05875
		V143-	0.07285	F157L	0.05875
		Y144-	0.14571		
		Y145H	0.10928		
		H146Y	0.80145		
		N148S	0.03642		
		W152R	0.03642		
		M153I	0.03642		
		E156D	0.03642		
		F157L	0.03642		
		E153-180	MESEFRVYSSANN CTFEVVSQPFLMD LE	M153I	0.03642
		E156D	0.03642		
		F157L	0.03642		
		L176I	0.03642		
		L176F	0.03642		
		G181V	0.03642		
E199-225	GYFKIYSKHTPINLV RDLPPQGFSALEP	Y200S	0.03642	S221W	0.05875
		Q218L	0.03642		
		S221W	0.03642		
		S221L	0.03642		
E238-262	FQTLALHRSYLTP GDSSSGWTAGA	Q239R	0.03642	Q239R	0.05875
		T240I	0.07285	T240I	0.05875
		S247R	0.03642	S247R	0.05875
		D253G	0.18214	W258L	0.11750
		W258L	0.14571		
		G261D	0.14571		
		A262T	0.14571		
E332-370	ITNLCPFGEVFNAT RFASVYAWNRKRI SNCVADYSVLYN	A348T	0.03642	A348T	0.05875
		V367F	0.07285	V367F	0.05875
E440-464	NLDSKVGGNYYL YRLFRRKSNLKP	Y453F	0.18214	NC	0.0
E481-506	NGVEGFNCYFPLQ SYGFQPTNGVGY	V483A	0.40072	V483A	0.5865
		G485R	0.03642		
		F486L	0.03642		
		S494P	0.10928		
		N501T	0.03642		
E627-651	DQLTPTWRVYSTG SNVFQTRAGCLIG	P631S	0.03642	P631S	0.05875
E694-715	AYTMSLGAENSVA YSNNSIAIPTNF	S704L	0.10928	S704L	0.11750
		A706S	0.03642		
E811-848	PSKRFSFIEDLLFNK VTLADAGFIKQYGD CLGDIAARDL	P812T	0.03642	I818V	0.05875
		I818S	0.03642	A845S	0.05875
		I818V	0.03642		
		T827I	0.03642		
		A829T	0.36429		
		G832C	0.03642		
		G838D	0.03642		
		G838S	0.03642		
		A845S	0.32786		
		A846V	0.03642		
E911-939	VTQNVLYENQKLI NQFNSAIGKIQDSLS	S929I	0.10928	S929I	0.1762
		A930V	0.03642	A930V	0.05875
		D936Y	0.03642	D936Y	0.05875
		S939F	0.21857	S939F	0.23501

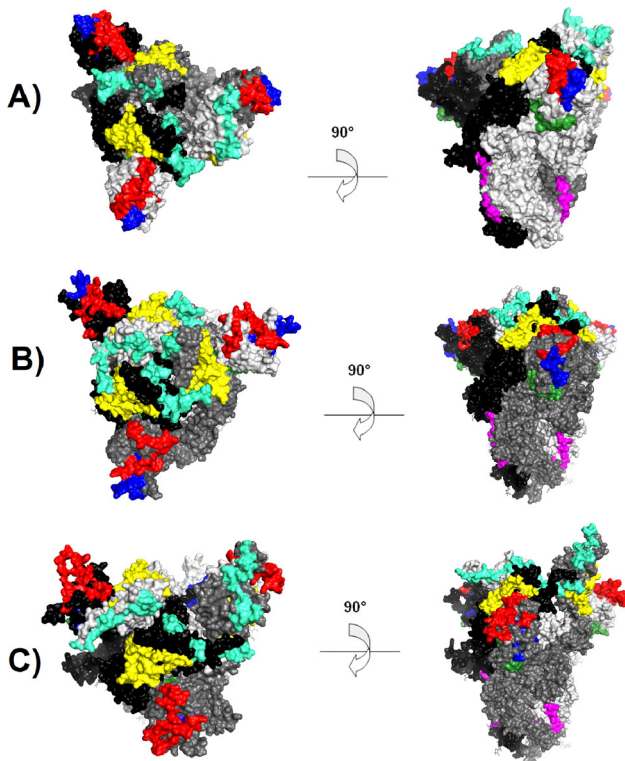
### Binding free energy calculations of the MCH II-epitope complexes

MCH II-epitope complexes binding free energy ( $\Delta G_{\text{bind}}$ ) were measure of the 10 ns MD simulation trajectories. A Stride of 10 was considered for the calculations, resulting in about

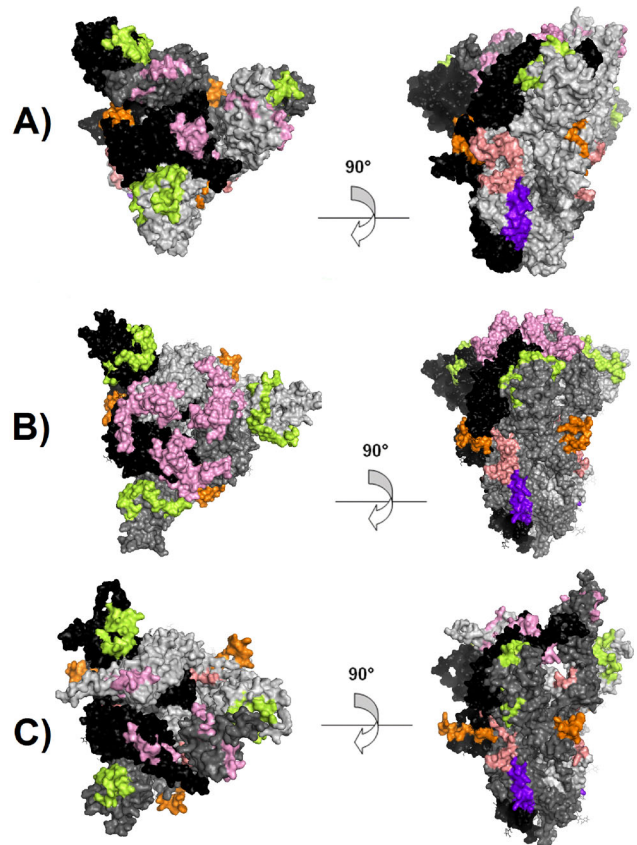
100 frames for the  $\Delta G_{\text{bind}}$  analysis. The estimations were carried out employing molecular mechanics combined with Poisson-Boltzmann surface area (MM-PBSA) method. MM-PBSA was applied by Calculation of Free Energy (CaFE) plugin (Liu & Hou, 2016) implemented to VMD program.



**Figure 2.** spike glycoprotein trimer. Figure is shown in black and white colors for a better view of the chains. Chain A is colored in black, chain B is colored in dark grey and chain C is colored in light grey. A) The shown trimer of the consensus sequence (SCons). B) The trimer is observed in the down conformation (6VXX-fill). C) The trimer is shown in the up conformation (6VYB-fill). The region of the ectodomain was modeled.



**Figure 3.** Epitopes exposed on the SARS-CoV-2 spike glycoprotein: The epitopes are shown in different colors as follows E136-167 (red), E199-225 (green), E238-262 (blue), E332-370 (yellow) E481-506 (turquoise) and E694-715 (magenta) are shown. A) The trimer of the consensus sequence (SCons). B) The trimer is observed in the down conformation (6VXX-fill). C) The trimer is shown in the up conformation (6VYB-fill).



**Figure 4.** Epitopes exposed in the SARS-CoV-2 spike glycoprotein. The epitopes are shown in different colors as follows; E440-464 (pink), E627-651 (orange), E153-180 (lime), E911-939 (purple) and E811-848 (salmon) are exposed in the SARS-CoV-2 spike glycoprotein. A) The consensus sequence of the trimer is shown (SCons). B) The trimer is observed in the down conformation (6VXX-fill). C) The trimer is shown in the up conformation (6VYB-fill).

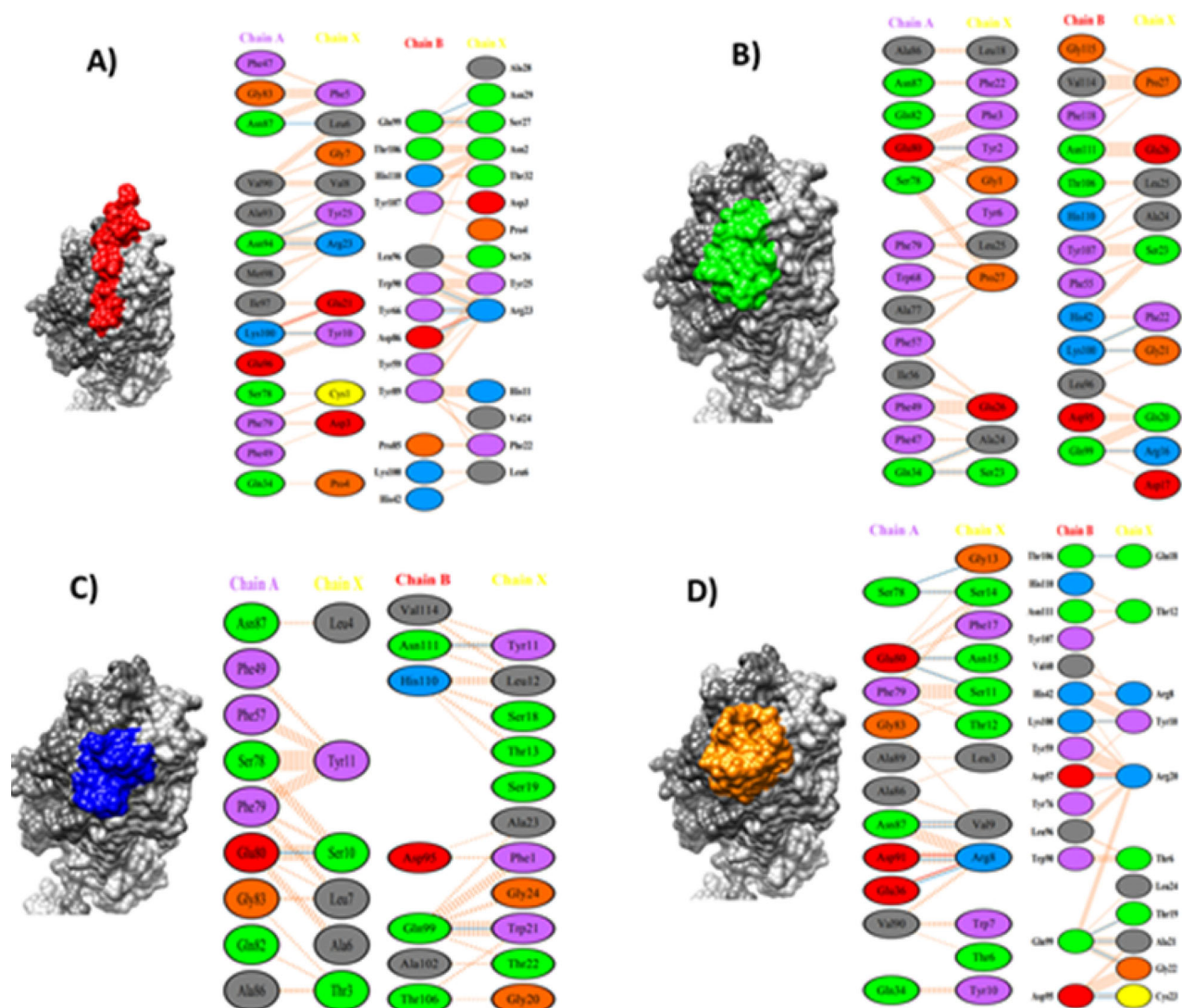
## Results

### Consensus sequences of the spike glycoprotein SARS-CoV-2

From the predicted alignment of the complete genome sequences, 81 amino acid changes were identified (Table S02 and Figure 1(A)) and from the spike glycoprotein SARS-CoV-2 alignment (protein sequences) 164 amino acid changes were found (Table S01 and Figure 1(B)). From these results, the mutation SNP D614G was identified, interestingly, this mutation is presumably found in the possible region of the S1-S2 binding site near the furin recognition site (R667) for the cleavage of the spike glycoprotein (Yin, 2020).

### Immunoinformatics of the SARS-CoV-2 spike glycoprotein

The sequence ID YP\_009724390.1 was analyzed through the Servers NetMHC 4.0 Server and NetMHCII/ 2.3 to predict immunogenic regions which could induce antibodies against the virus SARS-CoV-2. Both of the servers mentioned predicted more than 200 regions which could have affinity either for MHC I or MHC II proteins. After this process, we selected 11 epitopes because these regions showed promising characteristics which include promiscuity (Ebrahimi et al.,



**Figure 5.** Docking of the E136-167, E199-225, E238-262 and E627-651 peptides with the HLA-DRB1\*0401 molecules: A) the structural coupling of the E136-167 peptide with a lowest energy of  $-1903.2$  and a frequency of 24.3% was observed. B) E199-225 (green) is schematized with a lowest energy of  $-1191$  and a frequency of 15.4% C) We observe in blue color the peptide E238-264 interacting with the DRB1\*0401 complex, the peptide E238-262 is posed in the binding cavity to the ligand with a frequency of 36.5% and a lowest energy of  $-1173.3$ . D) E627-651 peptide (orange) binds in the binding cavity with a frequency of 37.7% and lowest energy of  $-1560.3$ . Furthermore, amino acids of the peptides interact with the amino acids in the cavity of HLA-DRB1\*0401.

2019; Saraav et al., 2016) conservation, affinity to HLA-Supertypes (Doytchinova & Flower, 2005; Sidney et al., 2008) and exposure to the surface of the trimeric spike glycoprotein (Liang et al., 2009).

a. Promiscuity and affinity revelation to HLA-Supertypes

After submitting peptides in the above mentioned Servers, 11 epitopes (Table 1) which showed affinity to the MHCI and MHCII molecules were identified as promiscuous since they were capable of binding to different molecules, and universal because displayed affinity towards HLA-Supertypes molecules, thus, epitopes would be efficient in different ethnic groups.

a. Conservation grade

The conservation grade of the epitopes was also calculated from the two multiple sequence alignments described above. Table 2 shows the most conserved epitopes, highlighting: E440-

464 (change of the residue in Y453F), E627-651 (change of the residue in P631S), E332-370 (changes of the these residues: A348T and V367F), and E694-715 (change of the residues S704L and A706S). Finally, the epitopes that showed more variations or higher number of residue changes are E811-848 and E136-167.

a. The exposure to the surface of the trimer spike glycoprotein

Figure 2 shows different trimers on which the epitopes are bound and they exhibited exposed, SCons (consensus sequence), 6VXX-fill (down conformation) and 6VYB-fill (up conformation).

In the up conformation (6VYB-fill), the E199-225 and E238-262 epitopes are not exposed; however, is the down conformation (6VXX-fill) and in the model of consensus sequence (SCons) they are exposed in the surface of the SARS-CoV-2 spike glycoprotein (Figure 3). It can be also depicted from the Figures 3 and 4 that the other epitopes are exposed in the all the

**Table 3.** Molecular interactions of peptides with MHCII: Reside interactions between epitopes and MHCII molecules. (NP) refers to non-interactions with the pocket.

Molecular interactions of peptides with MHCII				
ID	Pocket 1	Pocket 4	Pocket 6	Pocket 9
E136-167_HLA-DRB1*0401	Phe79A, Phe49A, His110B	Gln34A, Asn87A, His42B, Gln99B	Gln34A, Asn87A, Asn94A, Tyr59B, Lys100B, His42B	Ala93A, Asn94A, Glu96A, Ile97A, Met98A, Try95B, Tyr66B, Asp86B, Trp90B
E199-225_HLA-DRB1*0401	Phe49A, Phe79A, Phe57A, Trp68A, Ala77A, Glu80A, His110B, Val114B, Gly115B, Phe118B	Gln34A, Phe49A, Asn87A, Phe55B, Gln99B, Lys100B, Tyr107B	Gln34A, Ala86A, Asn87A, His42B, Lys100B	NP
E238-262_HLA-DRB1*0401	Phe57A, Phe79A, Glu80A, His110B, Val114B	Phe49A, Asn87A, Gln99B, Ala102B	Gln34A, Ala86A, Asn87A	NP
E627-651_HLA-DRB1*0401	Phe79A, Glu80A, His110B	Gln34A, Glu36A, Asn87A, Lys100B, His42B, Asp57B, Gln99B, Tyr107B	Gln34A, Ala86A, Asn87A, Ala89A, Asp91A, His42B, Asp57B, Tyr59B, Lys100B	Try95B, Trp90B
E153-180_HLA-DRB1*0401	Phe79A, Glu80, His110B, Val114B	Gln34A, Glu36A, Asn87A, Lys100B, His42B, Gln99B, Tyr107B	Gln34A, Ala86A, Asn87A, Asp91A, His42B, Tyr59B, Lys100B	Try95B, Trp90B
E332-370_HLA-DRB1*0401	Ala77A, Phe79A, Glu80, His110B, Val114B, Phe118B	Gln34A, Phe49A, Tyr107B	Gln34A	NP
E440-464_HLA-DRB1*0701	NP	Asn87A, Arg100B	Asn87A, Val90A, Asp91A, Asn94A, Arg100B, Glu57B, Trp38B, Glu57B, Leu59B	Ala93A, Asn94A, Ile97A, Arg101A, Trp38B, Leu59B, Val86B, Trp90B
E481-506_HLA-DRB1*0401	Phe79A, Phe57A, Trp68A, His110B, Val114B	Gln34A, Asn87A, His42B, Asp57B, Gln99B, Lys100B, Tyr107B	Gln34A, Ala86A, Asn87A, Val90A, Asn94A, His42B, Asp57B, Lys100B	Asn94A, Tyr59B, Trp90B
E694-715_HLA-DRB1*0401	Phe49A, Phe57A, Ala77A, Phe79A, Glu80A, His110B, Val114A	Gln34A, Phe49A, Asn87A, Lys100B, His42B, Tyr107B, Asp57B	Gln34A, Asn87A, Val90A, Asp91A, Lys100B, His42B, Asp57B, Tyr59B	NP
E811-848_HLA-DRB1*0401	Phe49A, Phe57A, Ala77A, Phe79A, Glu80A, His110B, Val114A	Gln34A, Phe49A, Asn87A, Lys100B, His42B, Tyr107B, Asp57B	Gln34A, Asn87A, Val90A, Lys100B, His42B, Asp57B	NP
E911-939_HLA-DRB1*0401	Phe49A, Ala77A, Phe79A, Glu80A, His110B, Val114B	Gln34A, Phe49A, Asn87A, His42B, Lys100B, Tyr107B	Gln34A, Ala86A, Asn87A, Val90A, Asp91A, Asn94A, Val40B, His42B, Tyr59B, Lys100B	Asn94A, Ile97A, Arg101A, Tyr59B, Asp86B, Tyr89B, Trp90B

**Table 4.** Molecular interactions of peptides with MHCI: Molecular interactions between peptides and MHCI molecules.

Molecular interactions of peptides with MHCI		
ID	Pocket B	Pocket F
E136-167_HLA-A*2402	Tyr31A, Lys90A, Phe123A, Tyr183A, Thr187A, Gly191A	His94A, Thr97A, Asp98A, Asn101A, Met121A, Lys170A, Trp171A
E332-370_HLA-A*2402	Tyr31A, Ser33A, Ala48A, Glu86A, Glu87A, Lys90A, Val91A, His94A, Phe123A, Tyr183A, Thr187A	His94A, Thr97A, Asn101, Met121A, Lys170A, Trp171A
E627-651_HLA-A*2402	Tyr31A, Lys90A, Phe123A, Tyr183A, His94A	His94A, Thr97A, Asn101, Met121A, Trp171A
E440-464_HLA-A*2402	Phe123A, Tyr183A, His94A	His94A, Thr97A, Met121A, Trp171A
E481-506_HLA-A*2402	Try31A, Lys90A, Glu86A, Thr187A, Tyr183A	Thr97A, Asn101A, Met121A, Lys170A, Trp171A
E153-180_HLA-A*2402	Glu86A, Lys90A, Tyr183A, His94A	His94A, Thr97A, Asn101, Met121A, Trp171A
E694-715_HLA-A*2402	Lys90A, Tyr183A, His94A, Phe123A, Tyr183A	His94A, Thr97A, Met121A
E811-848_HLA-A*0301	Tyr31A, Glu87A, Gly86A, Lys90A, Tyr183A, His94A, Tyr123A, Tyr183A	His94A, Thr97A, Arg121A
E911-939_HLA-A*2402	Tyr31A, Glu86A, Tyr183A, Thr187A	His94A, Thr97A, Asn101A, Ile104A, His138A, Thr167A, Trp171A

models. Specifically, the E481-506 epitope is more exposed in chain B (Figure 3(C) colored in gray) in comparison to the other conformations (Figure 3(A, B)).

Figure 4(B) shows down conformation (6VXX-fill) in which E440-464 epitope is more exposed on this structure than in the up conformation or in the SCons. Based on these results, we suggest that conformational state of the protein could influence the exposure grade of the peptides on the surface.

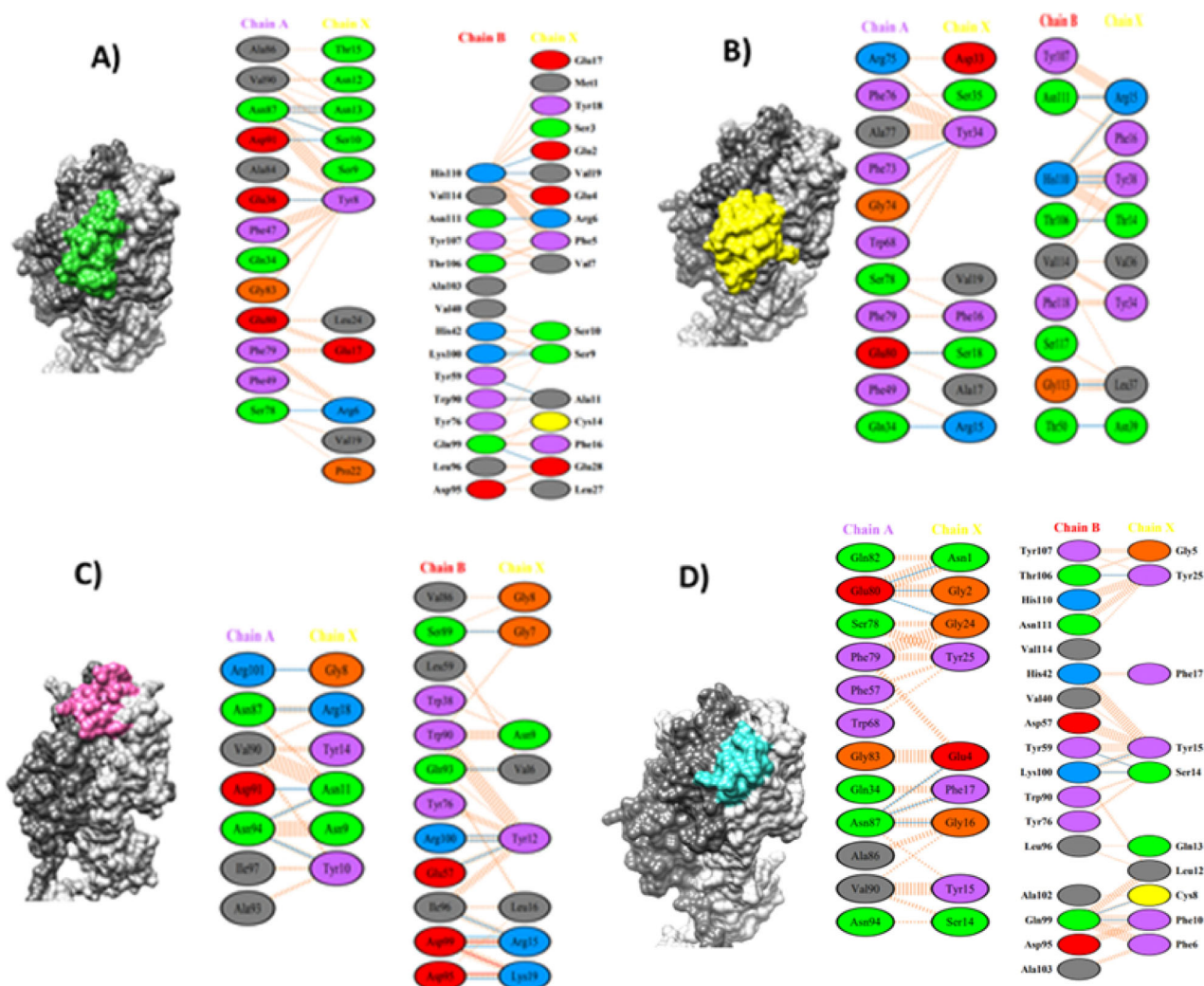
### Molecular docking of the chosen epitopes on the MHCI and MHCII proteins

The structure of the molecules MHCI and MHCII molecules are quite different. The cavity of MHCII shows different

pockets: Pocket1 (P1), Pocket 4 (P4), Pocket 6 (P6) and the Pocket 9 (P9) which are known to be crucial in the stability of the MHCII-epitope complex and they may play a role in the recognition of amino acids (Agudelo & Patarroyo, 2010). For the case of the MHCI complex, six specificity pockets A-F have been found: Pocket B (PB) and Pocket F (PF) which are known to be relevant in the anchorage process of some molecules on MHCI molecules (Hinrichs et al., 2010). In the present work we were focused on searching for other pockets that could be crucial for the stability of MHC-epitope complex.

The following peptides E136-167, E199-225, E238-262, E332-370, E153-180, E627-651, E481-506, E694-715, E811-848, and E911-939 were docked on the HLA-DRB1\*0401 molecule,





**Figure 6.** Docking of E153-180, E332-370 and E440-464 peptides with MHCII: A) The E153-180 peptide is coupled with the HLA-DRB1\*0401 haplotype to the peptide recognition cavity with a frequency of 16.3% and lowest energy of  $-1213.6$ . B) In this section, we observe the E332-370/HLA-DRB1\*0401 complex, This peptide is coupled with lowest energy  $-1361$  and a frequency of 20% in the peptide binding cavity. C) E440-464 peptide binds to HLA-DRB1\*0701 haplotype, with a 28.7% frequency and a lowest energy of  $-1196.7$ . D) The E481-506 binds to the cavity with a frequency of 28.3% and lowest energy  $-1545.9$ . In addition, the amino acids of the peptides that interact with the amino acids of the MHCII molecule cavity are shown.

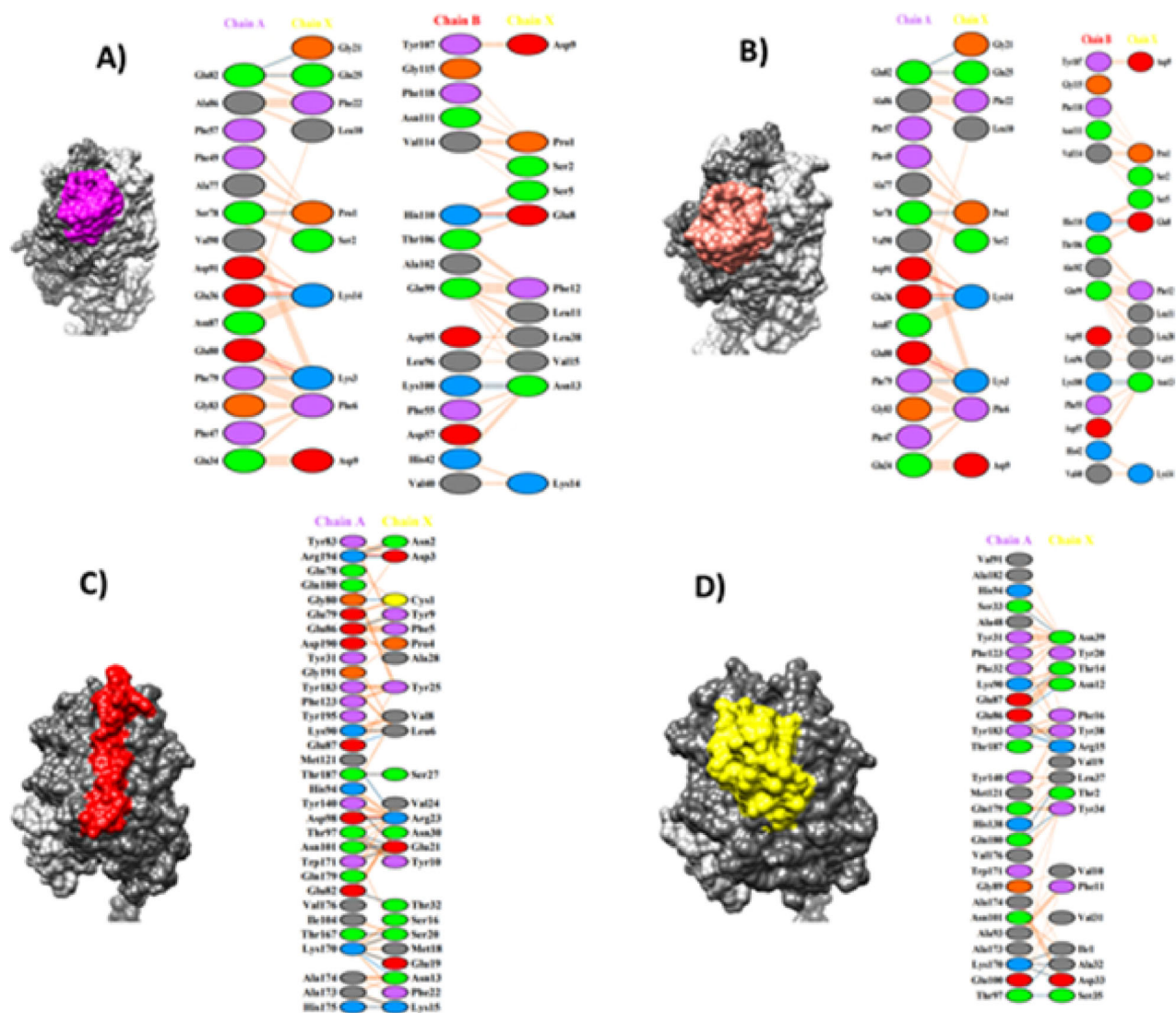
whereas the epitope E440-464 was coupled with HLA-DRB1\*0701. These peptides E136-167, E238-262, E332-370, E153-180, E481-506, E627-651 and E694-715 were coupled with the HLA-A\*2402 allele (Liu et al., 2010), whereas the peptide E811-848 was coupled with the HLA-A\*0301 allele. It is worth mentioning that the chosen alleles were identified as the most frequent among the ethnic groups. On the other side, the capability of the epitopes to form covalent and non-covalent interactions in the cavities of the alleles was also evaluated while highlighting the interactions present in the pockets and crucial for the stability for the MHCII-epitope complex.

Figure 5(A) and Table 3 show the results of the binding of the E136-167 peptide on HLA-DRB1\*0401 allele from which it could be observed that the peptide interacts with residues involved in the P1, P4, P6 and P9 pockets. Particularly, the residues Q99B (P4), N87A (P6, P9) and N94A (P6, P9) form hydrogen bonds, and specifically D86B (P9) is able to form hydrogen bonds and generate a saline bridge. This residue shows covalent interactions among the P4, P6 and P9 pockets.

The E153-180, E627-651 and E911-939 peptides form non-bonded interactions in the crucial pockets for both HLA-DRB1\*0401 and HLA-A\*2402 as depicted in Tables 3 and 4. The first peptide E153-180 is able to bind to the P1, P4, P6 and P9 pockets, which are located in the HLA-DRB1\*0401 allele. The PB and PF pockets display electrostatic interactions and hydrogen bonds with the HLA-A\*2402 allele (Figures 6(A) and 8(C)).

On the other side, E136-167 peptide interacts with the allele HLA-A\*2402 (Figure 7(C) and Table 4). In such complex, hydrogen bonds and saline bridges are present in the crucial pockets PB and PF. Thus, suggesting that the E136-167 peptide could form stable complexes with HLA-A\*2402 and HLA-DRB1\*0401 alleles.

The affinity of the E627-651\_HLA-DRB1\*0401 complex is attributed to three hydrogen bonds in the P1, P4 and P6 pockets and saline bridges sited in the P4 and P6 pockets (Figure 5(D)). For the E627-651\_HLA-A\*2402 complex, the presence of electrostatic interactions such as two hydrogen bonds located in the Pocket B of the HLA-A\*2402 allele are depicted (Figure 8(A)).



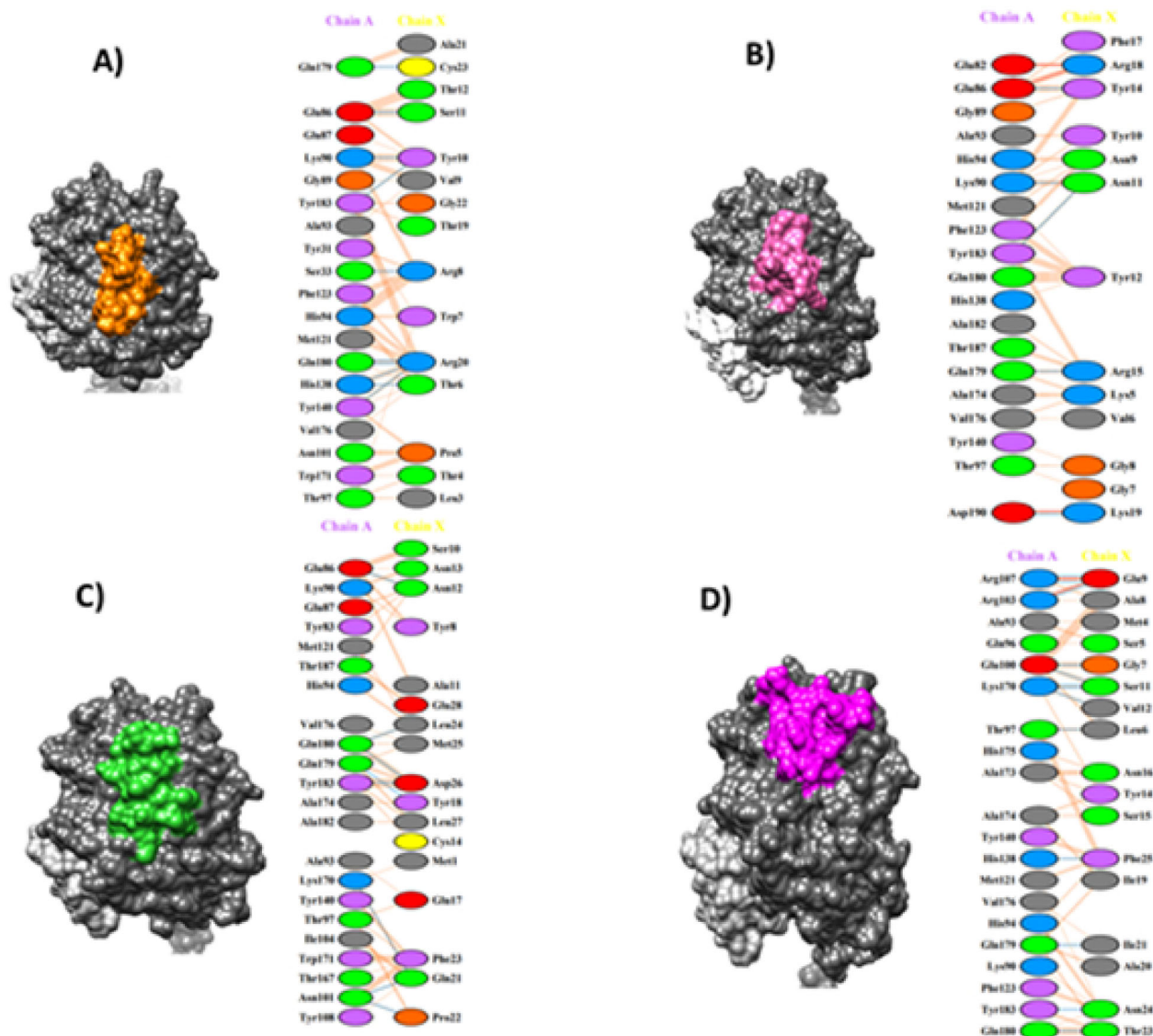
**Figure 7.** Docking of the peptides E694-715, E811-848, E136-167 and E332-370 with the MHC molecules: A) The peptide E694-715 (magenta) binds to the cavity of DRB1\*0401 with a frequency of 9.8% and a lowest energy of  $-1319.9$  B) The E811-848\_DRB1\*0401 complex is illustrated, where it was determined that the binding frequency of the ligand in this formation is 14.4% and a lowest energy of  $-1410.8$ . C) In this section it is shown that the E136-167 peptide binds to the cavity of the HLA-A\*2402 molecule with a frequency of 24.1% and a lowest energy of  $-1641.4$ . D) Finally in this figure the E332-370 peptide binds to HLA-A\*2402 with a frequency of 19.5% and lowest energy of  $-1441.4$ . In addition, the amino acids of the peptides that interact with the amino acids of the MHC molecule cavity are shown.

The E911-939 peptide is able to form electrostatic interactions that include four hydrogen bonds, three of which are located in the PF and PB pockets of the E911-939\_HLA-A\*2402 complex (Figure 9(B)). Figure 9(C) shows the E911-939\_HLA-DRB1\*0401 complex, and its affinity is mediated through hydrogen bonds located in the P4, P6 and P9 pockets.

The E332-370 and E440-464 peptides are situated in the RBD region, which corresponds to the binding domain of ACE2 receptor (Wrapp et al., 2020) making them even more interesting. The E332-370 peptide interacts within the P1, P4 and P6 pockets with the HLA-DRB1\*0401 allele. Figure 6(B) depicts the presence of 3 hydrogen bonds within the P1, P4 and P6 pockets (Table 3). The E332-370 peptide interacts with the PB y PF pockets in the HLA-A\*2402 molecule (Table 4) and the stability of the E332-370\_HLA-A\*2402 complex is mainly due to the presence of 7 hydrogen bonds (Figure 7(D)), of which 6 are located in the PB and PF pockets.

Interestingly, according to immunoinformatic predictions, the E440-464 peptide does not show affinity to HLA-DRB1\*0401, while E440-464 shows affinity for HLA-DRB1\*0701. From these theoretical predictions, it could be identified that these peptides bind to the P4, P6 and P9 pockets (Figure 6(C) and Table 3) in which electrostatic interactions, for example, 11 hydrogen bonds and 2 saline bridges, are found. Five of the hydrogen bonds are located in the P4, P6 and P9 pockets. The E440-464\_HLA-A\*2402 complex is stable because the E440-464 peptide is able to bind to the PB and PF pockets in the HLA-A\*2402 allele (Table 4 and Figure 8(B)). This complex shows non-covalent interactions, for example, 5 hydrogen bonds and 3 saline bridges. Additionally, there is the presence of one hydrogen bond in the PB pocket.

Figure 6(D) shows that the E481-506 epitope binds to the HLA-DRB1\*0401 molecule by non-covalent interactions, and three hydrogen bonds are located at the P4 and P6 pockets. In addition, Table 4 shows that the



**Figure 8.** Docking of the E627-651, E440-464, E153-180 and E694-715 peptides with the molecule HLA-A\*2402: A) the peptide E627-651 binds in the binding cavity to the ligand with a frequency of 23.2% and a lowest energy of  $-1361.4$ . B) It is shown in the section that the frequency of the E440-464\_HLA-A\*2402 complex is 35% and a lowest energy of  $-1171.3$ . C) The E153-180 is bound in the recognition cavity with a frequency of 19.7% and a lowest energy of  $-1252.9$ . D) E694-715 binds in the HLA-A\*2402 cavity with a lowest energy  $-1163.3$  and a frequency of 11.4%.

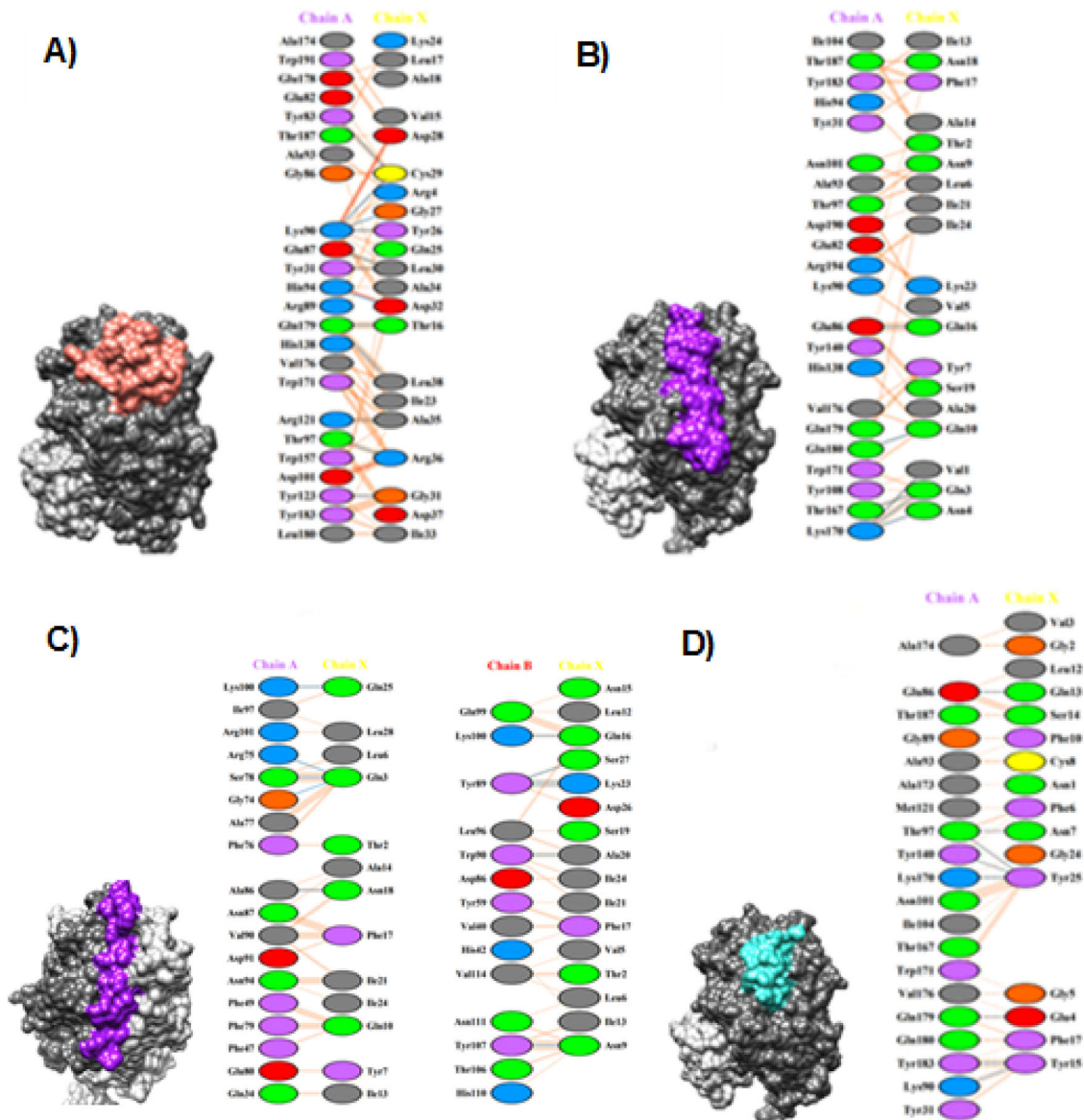
peptide binds to the crucial pockets. The E481\_506-HLA-A\*2402 complex (Figure 9(D) and Table 3) also binds to the crucial Pockets (PB and PF) and this binding is mediated seven hydrogen bridges, of which 5 are distributed in the crucial Pockets.

The E199-225, E238-262, E694-715 and E811-848 epitopes interact with the P1, P4 and P6 pockets of the HLA-DRB1\*0401 allele. The affinity of E199-225 and E238-262 peptides is due to non-covalent interactions, specifically through hydrogen bonds. The four hydrogen bonds which are formed in E199-225 are distributed in the P1, P4, P6 and P9 pockets (Figure 5(B)). The E238-262 peptide forms 3 hydrogen bonds, two of which are found in the of P1 and P4 pockets (Figure 5(C)).

The binding force of the E694-715 and E811-84 peptides is mediated by non-covalent interactions, including hydrogen bonds and saline bridges. The E694-715 peptide forms 6 hydrogen bonds, 3 of which are positioned in the P1, P4 and

P6 pockets, and 4 saline bridges, 3 of which are placed in the P1 and P6 pockets (Figure 7(A)). In the E811-848\_HLA-DRB1\*0401 complex, some hydrogen bonds can be depicted in the P1 and P4 pockets as well as saline bridges in the P1 pocket (Figure 7(B)). For the two complexes, E694-715\_HLA-A\*2402 and E811-848\_HLA-A\*0301, the bound peptides interact with the PB and PF pockets. Both of them show the formation of electrostatic interactions, which include hydrogen bonds and saline bridges between the peptides and MHC molecules. In the Figures 8(D) and 9 hydrogen bonds can be detected, of which 2 are positioned in the PB and PF pockets, as well as 2 saline bridges, which form the E694-715\_HLA-A\*2402 complex.

For the E811-848 peptide (Figure 9(A)), it could be seen that it was bound to the HLA-A\*0301 allele while displaying 9 hydrogen bonds, where 6 of them are sited at the PB and PF pockets. Additionally, 3 saline bridges were found and two of them are placed in the PB pocket.



**Figure 9.** Docking of the E811-848\_HLA-A\*0301, E911-939\_HLA-DRB1\*0401, E911-939\_HLA-A\*2402 and E481-506\_HLA-A\*2402 complexes: A) The E811-848 peptide binds to the cavity where the ligand is bound, with a frequency of 11.4% and lowest energy of  $-1163.3$ . B) It is observed that the frequency is 20.7% and lowest energy of  $-1453.9$  of the E911-939\_HLA-A\*2402 complex. C) The E911-939 peptide binds to the recognition site with a frequency of 31.1% and lowest energy of  $-1714.4$ . D) The E481-506\_HLA-A\*2402 complex binds to the cavity with a Lowest Energy of  $-1613.9$  and a frequency of 41.6%.

### Structural analysis of MD simulations

For these studies, we have focused mainly on the MHC II molecule since it has been widely described the potential roles of MHC-II molecules in inducing immune responses to vaccination although the mechanisms are still unknown (Eunju et al., 2014; Takenaka et al., 2012). Two MD simulations correspond to alleles HLA-DRB10401 and HLA-DRB10701 and the others to the MHCII-epitopes complexes.

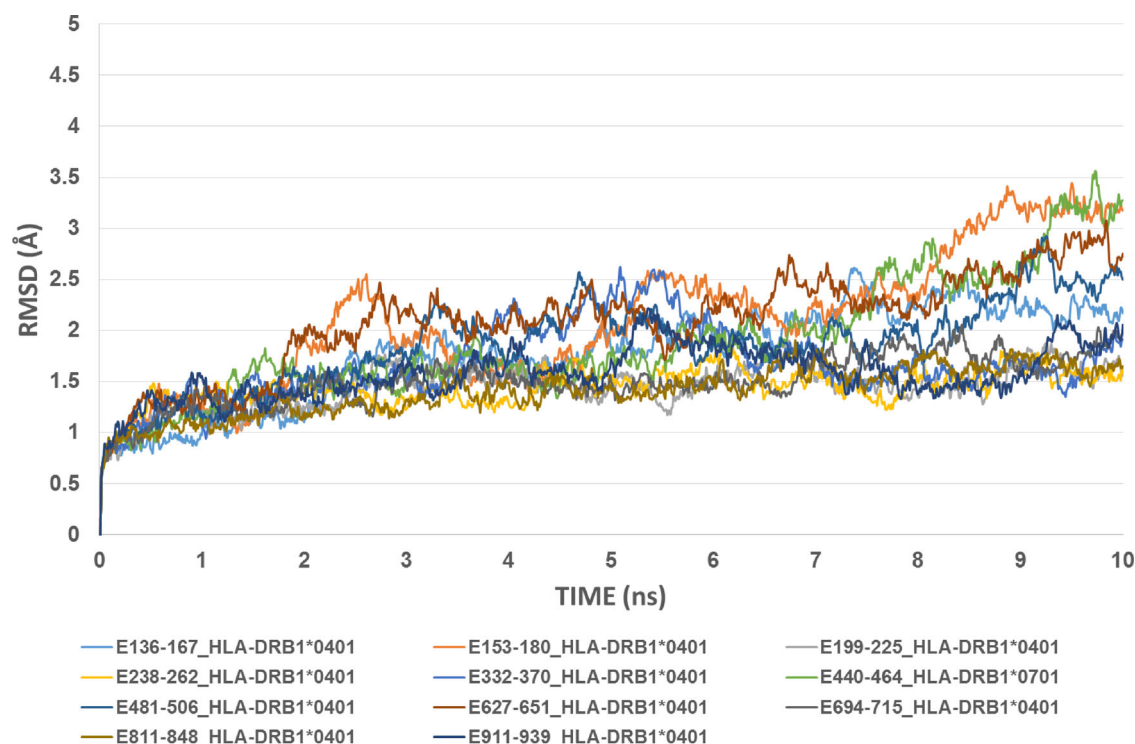
We carried out structural analysis of the 11 MHCII-epitopes MD simulations by calculating Root mean Square deviation (RMSD) while considering only the alpha carbon atoms of the MHCII molecules along the 10ns trajectories. Calculations were performed by using Carma Program (Glykos, 2006).

Figure 10 shows that RMSD values of all the protein structures are stable after about 2 ns. RMSD values of the MHCII molecules E199-225\_HLA-DRB1\*0401, E238-262\_HLA-DRB1\*0401, E694-

715\_HLA-DRB1\*0401 and E811-848\_HLA-DRB1\*0401 fluctuate between 1 to 2 Å. For the MHCII molecules E136-167\_HLA-DRB1\*0401, E332-370\_HLA-DRB1\*0401 and E911-939\_HLA-DRB1\*0401 RMSD values fluctuate between 1.5 and 2.5 Å. RMSD value for the MHCII molecule E440-464\_HLA-DRB1\*0701 oscillates between 1 to 3.5 Å. Finally, MHCII molecules E153-180\_HLA-DRB1\*0401, E481-506\_HLA-DRB1\*0401 and E627-651\_HLA-DRB1\*0401 show RMSD values between 1.5 to 3 Å.

### Binding free energy calculations

Table 5 shows the binding free energy calculations of the MHCII-epitopes complexes by using MM-PBSA method. From these results, it could be seen that the epitope with the lowest energy ( $-1810.9855$  kcal/mol) corresponds to E811-848\_HLA-DRB1\*0401. Other epitopes that show low binding free energy values correspond to E694-715\_HLA-DRB1\*0401,



**Figure 10.** Root mean Square Deviation (RMSD) calculations. Graph plots the curves from the RMSD calculations of MHCII molecules along the 10 ns trajectories. Calculations only considered the alpha carbon atoms of each of the proteins.

**Table 5.** Binding free energy calculations of the MHCII-epitopes. Table shows binding free energy values of the MHCII-epitopes complexes calculated through MM-PBSA methods. Epitopes that.

Free energy binding calculations			
ID	Energy complex (kcal/mol)	Energy receptor (kcal/mol)	Energy epitope (kcal/mol)
E136-167_HLA-DRB1*0401	-19226.2934	-17892.1930	-1329.5953
E153-180_HLA-DRB1*0401	-18544.7318	-17895.4225	-1193.5754
E199-225_HLA-DRB1*0401	-18916.5742	-17923.2356	-990.0683
E238-262_HLA-DRB1*0401	-18658.2674	-17908.5144	-726.5607
<b>E332-370_HLA-DRB1*0401</b>	<b>-19053.3214</b>	<b>-17845.2289</b>	<b>-1736.2376</b>
<b>E440-464_HLA-DRB1*0701</b>	<b>-18872.9431</b>	<b>-17407.1707</b>	<b>-1448.8533</b>
E481-506_HLA-DRB1*0401	-18523.1587	-17888.0617	-635.6636
E627-651_HLA-DRB1*0401	-18990.4175	-17864.4075	-1104.7079
<b>E694-715_HLA-DRB1*0401</b>	<b>-19735.7899</b>	<b>-17925.8473</b>	<b>-1798.6098</b>
<b>E811-848_HLA-DRB1*0401</b>	<b>-19722.2810</b>	<b>-17916.9543</b>	<b>-1810.9855</b>
E911-939_HLA-DRB1*0401	-19033.4650	-17893.9879	-1106.6094

E332-370\_HLA-DRB1\*0401 and E136-167\_HLA-DRB1\*0401. These epitopes are marked in bold in the table.

## Discussion

Currently, immunoinformatics represents a useful interface between computer science and experimental immunology. In this work, we have used computational tools, which permitted to find spike glycoprotein epitopes from SARS-CoV-2 as possible immunogenic peptides. In the process of selection of these epitopes, we considered different epitope features such as promiscuity, conservation and exposure to the surface of the trimeric spike glycoprotein.

Table 1 lists E136-167, E199-225, E238-262, E332-370, E481-506, E627-651, E153-180, E694-715 and E911-939 epitopes which show affinity to the HLA-DRB1\*0401 allele,

according to docking studies. Such alleles are expressed in different ethnic groups: Amerindian, Arab, Asian, Austronesina, Black, Caucousoid, Jew, kurd, Melanesian, Polynesian, Mixed, Mixed ancestry, North african-spain, Mulatto, Mextizo, Micronesian, Oriental, and Persian (Liu et al., 2010). Additionally, the E136-167, E199-225, E332-370, E440-464, E481-506, E627-651, E153-180, E694-715 and E911-939 peptides display affinity towards the HLA-A\*2402 allele, which is expressed in different ethnic groups Amerindian, Arab, Asian, Aboriginal, Austronesina, Caucousoid, Hispanic, Jew, Melanesian, Mextizo, Mixed, Mixed ancestry, Mulatto, Oriental Persian, Polynesian y Siberian (Gonzalez-Galarza et al., 2011). Overall, all these peptides mentioned in Table 1 were identified as promiscuous and showed affinity to HLA-Supertypes. These two properties of peptides are important because they can be recognized by different proteins of both the MCHI and MHCII families, providing greater capacity to generate antibodies. In addition, as mentioned in the databases where the frequency of the alleles has been investigated, we found that the HLA-DRB1\*0401 and HLA-A\*2401 alleles are distributed in different ethnic groups, therefore these peptides could be used as a base to design new vaccines for different human races or ethnic groups, which are important because SARS-CoV-2 is not only in China but also distributed in different countries.

From this immunoinformatics analysis, it could be depicted that the E332-370, E627-651, E440-464 and E694-715 epitopes are potential candidates as immunogenic agents because they show four important features: promiscuity, universality (HLA-Supertypes), exposure to the trimer and are conservation grade. Interestingly, the mentioned epitopes also showed the lowest binding free energies from the MM-

PBSA calculations, indicating also structural stability of the MHCII-epitope complexes.

The E440-464 peptide showed not only the mentioned features before, but also a region that can be crucial for the receptor interaction of the target cells, which could also initiate the infection process. Additionally, Structural alignments have demonstrated that the amino acid Leu455 is crucial in the affinity of spike glycoprotein for ACE2 protein (Wan et al., 2020).

On the other hand, Epitope 694-715 is located near the peptide fusion (Ou et al., 2016) in the trimeric spike glycoprotein, therefore, it could be recognized by antibodies, and this could prevent structural changes that are necessary for the virus to fuse to the endosome membrane and release genome for cycle replication. We suggest that the proposed epitopes by using immunoformatics tools could interfere in functional regions of the spike glycoprotein, which are crucial in the steps of the virus infection.

## Conclusion

From *In Silico* studies we identified that epitopes E332-370, E627-651, E440-464 and E694-715 accomplish essential features such as promiscuity, conservation grade, exposure and universality, and also they also form stable complexes with MHCII molecule as observed from the binding free energy calculations (MM-PBSA). On the whole, these characteristics could provide higher probability to generate a specific immune response. The epitope E440-464 is located in the functional region of the spike glycoprotein while the epitope E694-715 is located near this region. On the other hand, the antibodies that could be potentially generated by the epitope E440-464 could recognize the RBM region while avoiding the formation of the spike glycoprotein-ACE2 complex, and this could result in the interruption of the virus entry in a Host cell. Finally, these immunoformatics tools permitted us to identify epitopes that can be used as potential vaccines in pandemic events. This work proposed a set of epitopes that could be synthesized and assayed in animal model to determine their potential to induce immunogenic responses and even used for future applications such as the design of new epitope vaccines against the SARS-CoV-2, which could also represent a preventive strategy to stop the pandemic worldwide.

## Acknowledgements

We thank to CONACYT Project 254600. GLRS, JGM, JCB, MMA, are members of the National System of Researchers, SNI.

## Disclosure statement

The authors declare that they have no conflict of interest.

## References

Agudelo, W. A., & Patarroyo, M. E. (2010). Quantum chemical analysis of MHC-peptide interactions for vaccine design. *Mini Reviews in*

- Medicinal Chemistry*, 10(8), 746–758. <https://doi.org/10.2174/138955710791572488>
- Andreata, M., & Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics (Oxford, England)*, 32(4), 511–517. <https://doi.org/10.1093/bioinformatics/btv639>
- Apostolopoulos, V., Lazoura, E., & Yu, M. (2008). MHC and MHC-like molecules: Structural perspectives on the design of molecular vaccines. *Advances in Experimental Medicine and Biology*, 640, 252–267. [https://doi.org/10.1007/978-0-387-09789-3\\_19](https://doi.org/10.1007/978-0-387-09789-3_19)
- Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Peters, B. (2019). Properties of MHC class I presented peptides that enhance immunogenicity. *International Review of Immunology*, 38(6), 307–322. <https://doi.org/10.1080/08830185.2019.1657426>
- Cao, Y., Li, L., Feng, Z., Wan, S., Huang, P., Sun, X., Wen, F., Huang, X., Ning, G., & Wang, W. (2020). Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discovery*, 6, 11. <https://doi.org/10.1038/s41421-020-0147-1> eCollection 2020.
- Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, 17(3), 181–192. <https://doi.org/10.1038/s41579-018-0118-9>
- Doytchinova, I. A., & Flower, D. R. (2005). In silico identification of super-types for class II MHCs. *Journal of Immunology (Baltimore, MD)*, 174(11), 7085–7095. <https://doi.org/10.4049/jimmunol.174.11.7085>
- Ebrahimi, S., Mohabatkar, H., & Behbahani, M. (2019). Predicting promiscuous T cell epitopes for designing a vaccine against *Streptococcus pyogenes*. *Applied Biochemistry and Biotechnology*, 187(1), 90–100. <https://doi.org/10.1007/s12010-018-2804-5>
- Eunju, O., Lee, Y.-T., Ko, E.-J., Kim, K.-H., Lee, Y.-N., Song, J.-M., Kwon, Y.-M., Kim, M.-C., Perez, D. R., & Kang, S.-M. (2014). Roles of major histocompatibility complex class II in inducing protective immune responses to influenza vaccination. *Journal of Virology*, 88(14), 7764–7775. <https://doi.org/10.1128/JVI.00748-14>
- Fehr, A. R., & Perlman, S. (2015). Coronaviruses: An overview of their replication and pathogenesis. *Methods in Molecular Biology (Clifton, N.J.)*, 1282, 1–23. [https://doi.org/10.1007/978-1-4939-2438-7\\_1](https://doi.org/10.1007/978-1-4939-2438-7_1)
- Glykos, N. M. (2006). Software news and updates carma: A molecular dynamics analysis program. *Journal of Computational Chemistry*, 27(14), 1765–1768. <https://doi.org/10.1002/jcc.20482>
- Gonzalez-Galarza, F. F., Christmas, S., Middleton, D., & Jones, A. R. (2011). Allele frequency net: A database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Research*, 39(Database issue), D913–D919. <https://doi.org/10.1093/nar/gkq1128>
- Hinrichs, J., Föll, D., Bade-Döding, C., Huyton, T., Blasczyk, R., & Eiz-Vesper, B. (2010). The nature of peptides presented by an HLA class I low expression allele. *Haematologica*, 95(8), 1373–1380. <https://doi.org/10.3324/haematol.2009.016089>
- Jensen, K. K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J. A., Yan, Z., Sette, A., Peters, B., & Nielsen, M. (2018). Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology*, 154(3), 394–406. <https://doi.org/10.1111/imm.12889>
- Ji, X., Wang, W., Zhao, X., Zai, J., & Li, X. (2020). Cross-species transmission of the newly identified coronavirus 2019-nCoV. *Journal of Medical Virology*, 92(4), 433–440. <https://doi.org/10.1002/jmv.25682>
- Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., & Tufféry, P. (2016). PEP-FOLD3: Faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research*, 44(W1), W449–W454. <https://doi.org/10.1093/nar/gkw329>
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Laskowski, R. A., Jabłońska, J., Právda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science: A Publication of the Protein Society*, 27(1), 129–134. <https://doi.org/10.1002/pro.3289>
- Li, R., Qiao, S., & Zhang, G. (2020). Analysis of angiotensin-converting enzyme 2 (ACE2) from different species sheds some light on cross-

- species receptor usage of a novel coronavirus 2019-nCoV. *The Journal of Infection*, 80(4), 466–469. <https://doi.org/10.1016/j.jinf.2020.02.013>
- Liang, S., Zheng, D., Zhang, C., & Zacharias, M. (2009). Prediction of antigenic epitopes on protein surfaces by consensus scoring. *BMC Bioinformatics*, 10, 302. <https://doi.org/10.1186/1471-2105-10-302>
- Liu, H., & Hou, T. (2016). CaFE: A tool for binding affinity prediction using end-point free energy methods. *Bioinformatics*, 32(14), 2216–2218. <https://doi.org/10.1093/bioinformatics/btw215>
- Liu, J., Wu, P., Gao, F., Qi, J., Kawana-Tachikawa, A., Xie, J., Vavricka, C. J., Iwamoto, A., Li, T., & Gao, G. F. (2010). Novel immunodominant peptide presentation strategy: A featured HLA-A\*2402-restricted cytotoxic T-Lymphocyte epitope stabilized by intrachain hydrogen bonds from severe acute respiratory syndrome coronavirus nucleocapsid protein. *Journal of Virology*, 84(22), 11849–11857. <https://doi.org/10.1128/JVI.01464-10>
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., ... Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet (London, England)*, 395(10224), 565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Nishiura, H., Jung, S.-m., Linton, N. M., Kinoshita, R., Yang, Y., Hayashi, K., Kobayashi, T., Yuan, B., & Akhmetzhanov, A. R. (2020). The extent of transmission of novel Coronavirus in Wuhan. *Journal of Clinical Medicine*, 9(2), 330. <https://doi.org/10.3390/jcm9020330>
- Ou, X., Zheng, W., Shan, Y., Mu, Z., Dominguez, S. R., Holmes, K. V., & Qian, Z. (2016). Identification of the fusion peptide-containing region in betacoronavirus spike glycoproteins. *Journal of Virology*, 90(12), 5586–5600. <https://doi.org/10.1128/JVI.00015-16>
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781–1802. <https://doi.org/10.1002/jcc.20289>
- Rose, P. W., Plić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., Costanzo, L. D., Duarte, J. M., Dutta, S., Feng, Z., Green, R. K., Goodsell, D. S., Hudson, B., Kalro, T., Lowe, R., Peisach, E., Randle, C., Rose, A. S., Shao, C., ... Burley, S. K. (2017). The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Research*, 45(D1), D271–D281. <https://doi.org/10.1093/nar/gkw1000>
- Saraav, I., Pandey, K., Sharma, M., Singh, S., Dutta, P., Bhardwaj, A., & Sharma, S. (2016). Predicting promiscuous antigenic T cell epitopes of Mycobacterium tuberculosis mymA operon proteins binding to MHC Class I and Class II molecules. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 44, 182–189. <https://doi.org/10.1016/j.meegid.2016.07.004>
- Sidney, J., Peters, B., Frahm, N., Brander, C., & Sette, A. (2008). HLA class I supertypes: a revised and updated classification. *BMC Immunology*, 9(1), 1. <https://doi.org/10.1186/1471-2172-9-1>
- Takenaka, M., Tiriveedhi, V., Subramanian, V., Hoshinaga, K., Patterson, A. G., & Mohanakumar, T. (2012). Antibodies to MHC class II molecules induce autoimmunity: Critical role for macrophages in the immunopathogenesis of obliterative airway disease. *PLoS One*, 7(8), e42370. <https://doi.org/10.1371/journal.pone.0042370>
- Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins*, 85(3), 435–444. <https://doi.org/10.1002/prot.25219>
- Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *Journal of Virology*, 94(7), 1–9. <https://doi.org/10.1128/JVI.00127-20>
- Wang, R., Zhang, X., Irwin, D. M., & Shen, Y. (2020). Emergence of SARS-like Coronavirus poses new challenge in China. *The Journal of Infection*, 80(3), 350–371. <https://doi.org/10.1016/j.jinf.2020.01.017>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>
- Webb, B., & Sali, A. (2016). Comparative protein structure modeling using modeller. *Current Protocols in Protein Science*, 86, 1–37. <https://doi.org/10.1002/cpps.20>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the SARS-CoV-2 spike in the prefusion conformation. *Science*, 367(6483), 1260–1263. <https://doi.org/10.1126/science.abb2507>
- Yin, C. (2020). Genotyping coronavirus SARS-CoV-2: Methods and implications. *Genomics*, 1–9. <https://doi.org/10.1016/j.ygeno.2020.04.016>