RESEARCH ARTICLE

In silico prediction of SARS-CoV-2 epitopes for vaccine development

Kitz Paul D. Marco*[1], Julia Patricia B. Llagas[1], Maria Teresa A. Barzaga[2], Francisco M. Heralde III*[1,2]

*Corresponding author's email address: kdmarco@up.edu.ph, fmheralde1@up.edu.ph.

[1]*Department of Biochemistry and Molecular Biology, College of Medicine, University of the Philippines Manila, Manila City*
[2]*Molecular Diagnostics and Cellular Therapeutics Laboratory, Lung Center of the Philippines, Quezon City*

ABSTRACT

The ongoing coronavirus disease (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is causing major damages in health and economies worldwide. The development of safe and effective vaccines for COVID-19 is of utmost importance yet none have been licensed to date. One of the strategies for vaccine development utilizes dendritic cells which express class I and class II human leukocyte antigen (HLA) molecules. These HLA molecules present the antigenic peptides to T cells which mediate the immune response. Thus, the study aimed to identify SARS-CoV-2 peptides with potential binding to HLA class I and class II molecules using different bioinformatics tools. SYFPEITHI and IEDB were used to predict epitopes for the most common HLA class I and II alleles among Filipinos. The top predicted epitopes were subjected to de novo and template-based molecular docking. Then, binding energies of the generated peptide-HLA complexes to putative T cell receptors were predicted using a homology modeling approach. Several predicted epitopes showed promising MHC and TCR binding, although results varied considerably between the prediction methods used. In particular, the results of de novo and template-based docking methods did not coincide, the latter of which generated complexes that more closely resemble typical peptide-HLA complexes. The results of this study will be validated by the next stage of the vaccine development project which is the in vitro assessment of the T cell responses elicited by dendritic cells pulsed with the candidate peptides.

Keywords: *COVID-19, SARS-CoV-2, vaccine, molecular docking*

Introduction

The Coronavirus disease (COVID-19) is the infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that is responsible for the ongoing pandemic affecting over 200 countries and territories globally [1]. SARS-CoV-2 is a novel strain of coronavirus, initially believed to be spread mainly through contact and droplet transmission but is now considered to be potentially airborne [2-4]. It enters human cells via the angiotensin-converting enzyme 2 (ACE2) receptor in the presence of TMPRSS2, which are co-expressed in the lungs, esophagus, intestines, and other tissues [5-7]. COVID-19 is primarily a respiratory illness, characterized by cough, pneumonia, or acute respiratory distress syndrome (ARDS) depending on the severity [8]. However, it may also present with extrapulmonary manifestations such as lymphopenia, myocardial infarction, and stroke [9]. In severe cases, COVID-19 may lead to respiratory failure and death [10]. As of 1 November 2020, over 45 million people have been confirmed to be infected with SARS-CoV-2 and around 1.2 million mortalities have been recorded due to COVID-19, with observed case-fatality rates ranging from 0.1 to 29% in different countries [1,11]. The Philippines is one of the countries that are greatly affected by this pandemic. It currently ranks 22nd and 2nd among countries with the highest total number of cases globally and in Southeast Asia, respectively, with over 378 thousand reported cases. It also has the highest mortality rate in Southeast Asia with 67.82 deaths per million of population [1,12].

Currently, clinical management of COVID-19 remains largely supportive, and therapies specifically indicated for the disease are still only investigational [13-15]. There are several antiviral and anti-inflammatory pharmacotherapies as well as passive immunotherapies currently undergoing

clinical trials for the treatment of infected patients [16]. However, in the absence of a vaccine for COVID-19, slowing down disease transmission currently relies on hygiene, disinfection, and personal protective measures, as well as drastic quarantine policies, resulting in massive disruptions to normal life [17]. Hence, while a safe and effective vaccine has yet to be licensed, vaccine research and development efforts remain a priority pandemic response. To date, 45 and 156 candidate vaccines for COVID-19 are in clinical and preclinical stages, respectively [18]. Current platforms under investigation include numerous RNA, DNA, recombinant protein, virus-like particle, viral vector-based, inactivated, and even live attenuated vaccines [18,19].

One under-investigated vaccine platform are dendritic cell vaccines. Dendritic cells are the most potent antigen-presenting cells (APCs); they can activate both CD8+ T cells to kill virus-infected cells and CD4+ T cells to induce antibody-production in plasma cells via major histocompatibility complex (MHC) class I and II molecules, respectively [20]. A fraction of these short-lived effector T cells survive to become memory cells that quickly respond to subsequent antigenic challenge, providing long-term immunity [21]. While most researches on dendritic cell vaccines focus on cancer therapy, animal and human studies on hepatitis viruses, human immunodeficiency virus 1 (HIV-1), and herpes simplex virus 1 (HSV-1) among others suggest a potential for dendritic cell vaccination in viral infections [22-25]. Using synthetic peptides, which require no handling of the infective SARS-CoV-2 virion, dendritic cells can be loaded with viral epitopes and induced to mature for potential use in a COVID-19 vaccine.

The priming of T cells by dendritic cells requires the loading of the antigenic peptide onto the MHC molecule, also called human leukocyte antigen (HLA) in humans, and recognition of the peptide-MHC complex (pMHC) by a T cell receptor (TCR) [26]. Hence, in the design of peptide-loaded dendritic cell vaccines, the binding affinities of peptides to MHC and pMHC to TCR are important considerations. In this paper, we predicted SARS-CoV-2 epitopes for the most common HLA class I and II alleles in Filipinos, modelled their structures using de novo and templated-based methods, and performed docking to determine pMHC binding affinity. We also attempted to determine the affinity of the generated pMHCs for HLA class I to available TCRs. By focusing on highly immunogenic SARS-CoV-2 proteins, i.e. spike, envelope, and membrane proteins, as well as the ORF1ab polyprotein encoded by two-thirds of the viral genome, this bioinformatics approach may facilitate the design of a protective SARS-CoV-2 vaccine [27,28].

Methodology

SARS-CoV-2 epitope prediction

The allele frequency of HLA class I and class II alleles in the Philippines were obtained from the Allele Frequency Net Database [29]. The most common alleles were found to be HLA-A*02:01 and HLA-A*24:02 for class I and HLA-DRB1*15:02 and HLA-DRB1*12:02 for class II. SYFPEITHI and Immune Epitope Database (IEDB) algorithms were used to predict 9-mer HLA class I and 15-mer class II epitopes, respectively, from the sequences of SARS-CoV-2 proteins [30,31]. The HLA Class I peptides were based on the selection list of FMH while the HLA Class II peptides were based on the selection list of Dennis Macapagal.

HLA structures

The experimental structures of HLA-A*02:01 (3D25, 1.30 Å) and HLA-A*24:02 (3WL9, 1.66 Å) were obtained from the Protein Data Bank. The three-dimensional structures of HLA-DRB1*12:02 and HLA-DRB1*15:02 were modeled with SWISS-MODEL using the sequences obtained in the IPD-IMGT/HLA Database [32,33].

Molecular docking on HLA

The predicted epitopes were docked on the HLA molecules using de novo and template-based molecular docking. 'Stuffer' peptides, or self-peptides that are able to bind to the HLA molecule and are displaced by antigenic peptides, were also included in the analysis. Phogrin$_{331-339}$, probable C-mannosyltransferase DPY19L4$_{139-147}$, and CLIP$_{87-101}$ were used as stuffer peptides in this study [20,34,35]. Peptides that do not bind the HLA alleles of interest – Hantaan virus nucleoprotein$_{131-139}$ (HNTV-NP), human immunodeficiency virus I Gag-Pol polyprotein$_{273-282}$, and herpes simplex virus I tegument protein$_{519-533}$ (VP11/12) – were used as negative controls [36-38]. The stuffer peptides and negative control used for each HLA allele in the study are outlined in Table 1. The model with the highest binding affinity (greatest binding energy) was selected to represent the predicted binding of the selected SARS-CoV-2 peptides. Peptide-HLA interactions were visualized using UCSF Chimera 1.14 [39].

De novo molecular docking

The three-dimensional structures of the predicted epitopes were generated using PEP-FOLD3, which performs de novo

Table 1. *Stuffer and negative control peptides used in the molecular docking of predicted SARS-CoV-2 epitopes on HLA molecules.*

HLA Allele	Stuffer Peptide		Negative Control	
	Protein	Sequence	Protein	Sequence
A*02:01	Phogrin	GMAELMAGL	HNTV-NP	VPILLKALY
A*24:02	DPY19L4	LYPELIASI	Gag-Pol	VPLDKDFRKY
DRB1*12:02	CLIP	PVSKMRMATPLLMQA	VP11/12	YTHMGEVPPRLPARN
DRB1*15:02	CLIP	PVSKMRMATPLLMQA	VP11/12	YTHMGEVPPRLPARN

Table 2. *TCRs used for the modeling of TCR-peptide-MHC complexes.*

HLA Allele	Target	TCRA			TCRB		
		CDR3	V	J	CDR3	V	J
A*02:01	S	CATEGDSGYSTLTF	TRAV17	TRAJ11	CASSLQGGNYGYTF	TRBV6-5	TRBJ1-2
	ORF1ab	CADLQTNARLMF	TRAV5	TRAJ31	CSGGQGMVDGYTF	TRBV29-1	TRBJ1-2
	E	CAFTSGTYKYIF	TRAV38-1	TRAJ40	CASSIVQGSNQPQHF	TRBV19	TRBJ1-5
	M	CALSGPNTGNQFYF	TRAV19	TRAJ49	CASSLGILGGSLEPQHF	TRBV13	TRBJ1-5
A*24:02	S	CAVNRGTALIF	TRAV12-2*01	TRAJ15*01	CASTPENQETQYF	TRBV28*01	TRBJ2-5*01
	ORF1ab	CAAKEGYSTLTF	TRAV13-1*01	TRAJ11*01	CASSSTGGGEKDQPQHF	TRBV5-4*01	TRBJ1-5*01
	E	CAAKEGYSTLTF	TRAV13-1*01	TRAJ11*01	CASSSTGGGEKDQPQHF	TRBV5-4*01	TRBJ1-5*01
	M	CAVNRGTALIF	TRAV12-2*01	TRAJ15*01	CASTPENQETQYF	TRBV28*01	TRBJ2-5*01

prediction of native peptide conformations in aqueous solution [40]. Then, the predicted epitopes were subjected to flexible ligand docking in AutoDock Vina using default docking parameters (e.g., exhaustiveness = 8). A 33.75×18.75×15.0 Å^3 (HLA-A*02:01 and HLA-A*24:02) or 20.25×20.25×47.25 Å^3 (HLA-DRB1*12:02 and HLA-DRB1*24:02) gridbox spanning the whole peptide-binding cleft was used for HLA class I or class II alleles, respectively.

Template-based molecular docking

The predicted epitopes were subjected to preliminary docking in GalaxyPepDock, a web server that performs template-based protein-peptide docking [41]. For each predicted epitope, the top model based on estimated prediction accuracy was subsequently re-docked in AutoDock Vina using default docking parameters [42]. The ligand backbone was set as rigid prior to docking, allowing only its side chain bonds to rotate. The gridbox was defined with the same size as previously described, oriented along the peptide-binding cleft, and centered on the ligand.

Prediction of T-cell receptor binding

Prospect T cell receptors (TCRs) that can potentially recognize the generated peptide-MHC (pMHC) complexes were obtained from VDJdb [43]. Among the TCRs restricted to the chosen HLA-A alleles, those whose epitopes shared the highest sequence identity with the SARS-CoV-2 peptides were selected (Table 2). There were no available TCRs for the HLA class II alleles of interest hence the predicted epitopes for HLA class II were excluded in the subsequent analyses. The sequence of each TCR was obtained from the TCRmodel web server [44]. Then, sequences of the TCR, peptide, and MHC were submitted to TCRpMHCmodels for homology modeling of TCR-pMHC complex structures [45]. In the resulting model, the pMHC complex was replaced with that obtained from molecular docking and the new TCR-pMHC complex was

refined using the GalaxyRefineComplex web server [46]. The free energy of binding of the pMHC to the TCR at 25°C was predicted using the PRODIGY web server [47].

Results

For HLA class I, the highest-scoring epitope predicted by SYFPEITHI per viral protein per allele was selected for analysis (Table 3). HLA class I molecules, which have close-ended peptide-binding clefts, are only able to present epitopes 8-15 residues long [48,49]. Optimal epitope length varies between HLA alleles, and is 9-mer for A*02:01 and A*24:02 [48]. Unsurprisingly, the top HLA-A*02:01 epitopes for all protein targets have hydrophobic residues at P2 and the C-terminus, which are known anchors for this allele [50]. Meanwhile, HLA-A*24:02 epitopes have either F or Y at P2 and I or L at the C-terminus, consistent with its known anchor residues [51]. Notably, the scores for HLA-A*02:01 were generally higher than for HLA-A*24:02, possibly due to the different maximal SYFPEITHI scores for different alleles [52]. While the highest possible score for each allele is not published, the disparity could potentially arise from differences in number of anchor positions, unfavorable residues, and other such factors that may vary between alleles that could affect the scoring.

On the other hand, the peptide-binding groove of HLA class II molecules have open ends, allowing epitopes 12-25 amino acids long where only 9 residues typically comprise the binding core [53,54]. Because of the degeneracy of HLA class II anchor positions that result from this, epitope prediction for class II tends to be trickier than for class I [55]. For this analysis, four S protein epitopes predicted by IEDB to bind to multiple HLA class II alleles were chosen, although subsequent analyses were performed using only the two most common alleles among Filipinos, i.e. HLA-DRB1*12:02 and HLA-DRB1*15:02. While data on optimal epitope length for each HLA-DR allele is not available, HLA class II epitope length distribution has been said to peak at 15-mer [56]. The chosen HLA class II epitopes are as follows: VEGFNCYFPLQSYGF, YQTQTNSPRRARSVA, VGGNYNYLYRLFRKS, and CGSCCKFDEDDSEPV.

Next, molecular docking was used to provide structure-based insights into the potential immunogenicity of the selected SARS-CoV-2 peptides. For the purposes of this study, potential binding to the HLA alleles of interest will be defined as greater predicted binding free energy compared to the corresponding stuffer and negative control peptides. A higher affinity for the predicted epitope than for the stuffer peptide suggests that the former may be able to displace the stuffer peptides and be loaded onto the HLA molecule for antigen presentation. Out of the four predicted epitopes per allele, two for A*02:01, four for A*24:02, three for DRB1*12:02, and two for DRB1*15:02 potentially bind the HLA molecule according to de novo or template-based docking (Table 4). The rest of the peptides were predicted to have greater binding energy than the stuffer peptide by de novo and/or template-based docking but could not be considered to have potential HLA binding due to equal or lower predicted binding energy than the negative control. This was because the stuffer peptide, which is supposedly able to bind to HLA, surprisingly had a lower predicted binding energy than the negative control in many cases. This demonstrates the limitations of the methods used in this study, which shall be discussed further in the succeeding section.

The results differed greatly between the two methods of molecular docking - both in the absolute and more importantly in the relative values. Template-based docking predicted much greater binding energies for all peptides. Additionally, the trends in the predicted binding energy for each allele differed between the two methods. Both methods agreed on the top epitope only for HLA-DRB1*15:02. VEGFNCYFPLQSYGF was the top epitope predicted by both methods for DRB1*15:02 and by template-based docking for HLA-DRB1*12:02. Meanwhile, template-based docking ranked QYIKWPWYI as the top epitope for HLA-A*24:02 which coincides with the results of sequence-based algorithm of SYFPEITHI (Table 3). It is interesting to note that no single peptide was predicted to have potential HLA binding by both de novo and template-based molecular docking. Since the docking methods produced conflicting results, structural information of the docking outputs may be used to determine which is more likely to be correct.

The interactions between the epitopes with the greatest binding energies and the HLA peptide-binding domain were predominated by Van der Waals forces (Figure 1). Hydrogen bonds mostly involved terminal residues for the HLA class I epitopes and was distributed along the peptide for the HLA class II epitopes. More peptide-MHC contacts and hydrogen bonds were determined for complexes generated by template-based docking, which possibly explains their difference in the obtained binding energy values.

The docking outputs generally resembled typical MHC class I and class II epitopes in terms of overall ligand conformation (Figure 2). That is, buried ends and raised center for class I epitopes and a more linear, fully extended conformation for class II epitopes [57]. However, de novo docking yielded ligand structures that were generally raised

Table 3. *SARS-CoV-2 peptides predicted by SYFPEITHI to bind to HLA-A*02:01 and HLA-A*24:02.*

HLA Allele	Viral Protein Target	Epitope Sequence	SYFPEITHI Score
A*02:01	S (spike protein) ORF1ab (polyprotein) E (envelope protein) M (membrane protein)	FIAGLIAIV FLLPSLATV FLLVTLAIL KLLEQWNLV	30 33 29 26
A*24:02	S (spike protein) ORF1ab (polyprotein) E (envelope protein) M (membrane protein)	QYIKWPWYI MYASAVVLL VFLLVTLAI LYIIKLIFL	24 23 21 23

Table 4. *Predicted binding energies (ΔGbinding) of the SARS-CoV-2 peptides to HLA molecules based on de novo and template-based molecular docking, including stuffer peptides (phogrin, DPY19L4, and CLIP) and negative controls (HNTV-NP, Gag-Pol, and VP11/12). The values in bold indicate the top predicted epitopes per docking method per HLA allele.*

HLA Allele	Protein	Epitope Sequence	De novo	Template-based	
			$\Delta G_{binding}$ (kcal/mol)	$\Delta G_{binding}$ (kcal/mol)	Template
A*02:01	S	FIAGLIAIV[b]	-7.7	**-16.0**	1QEW
	ORF1ab	FLLPSLATV[b]	-7.6	-15.5	1QEW
	E	FLLVTLAIL	**-8.1**	-13.4	2CLR
	M	KLLEQWNLV	-7.1	-11.5	2CLR
	Phogrin	GMAELMAGL	-7.6	-10.3	1A9E
	HNTV-NP	VPILLKALY	-8.1	-13.4	2X4S
A*24:02	S	QYIKWPWYI[b]	-7.3	**-16.8**	2BCK
	ORF1ab	MYASAVVLL[a]	-7.7	-8.1	2BCK
	E	VFLLVTLAI[a]	-7.7	-8.7	2BCK
	M	LYIIKLIFL[a]	**-7.9**	-13.1	2BCK
	DPY19L4	LYPELIASI	-7.5	-14.7	2BCK
	Gag-Pol	VPLDKDFRKY	-6.6	-11.3	2AXG
DRB1*12:02	S	VEGFNCYFPLQSYGF[b]	-6.4	**-14.4**	1A6A
	S	YQTQTNSPRRARSVA[b]	**-6.9**	-13.1	1H15
	S	VGGNYNYLYRLFRKS[b]	-6.2	-12.6	1H15
	S	CGSCCKFDEDDSEPV	-6.5	-9.7	1PYW
	CLIP	PVSKMRMATPLLMQA	-5.6	-11.1	1A6A
	VP11/12	YTHMGEVPPRLPARN	-7.2	-10.9	1A6A
DRB1*15:02	S	VEGFNCYFPLQSYGF[b]	**-6.8**	**-13.8**	1YMM
	S	YQTQTNSPRRARSVA	-6.2	-9.7	1H15
	S	VGGNYNYLYRLFRKS[b]	-5.5	-12.7	1H15
	S	CGSCCKFDEDDSEPV	-6.0	-11.6	1YMM
	CLIP	PVSKMRMATPLLMQA	-5.6	-10.9	1A6A
	VP11/12	YTHMGEVPPRLPARN	-6.9	-12.6	1A6A

[a]Predicted by de novo molecular docking to have higher affinity to the HLA molecule than both the stuffer peptide and the negative control.
[b]Predicted by template-based molecular docking to have higher affinity to the HLA molecule than both the stuffer peptide and the negative control.

farther from the HLA 'floor' and many were folded upon themselves which is uncharacteristic of MHC class II epitopes. Furthermore, template-based docking-derived ligands generally showed a snugger fit in the peptide-binding cleft than those obtained from de novo docking (Figures 3-6). This difference was more apparent for de novo-predicted HLA class II complexes which had folded epitope conformations. Overall, these indicate the template-based docking method used in the study may have yielded more accurate results over the de novo method.

TCR-pMHC binding was additionally attempted for HLA class I alleles as a potential predictor of epitope immunogenicity. While the TCR sequences were chosen on the basis of the shared sequence identity between the original TCR epitopes and each SARS-CoV-2 epitope, there was generally little similarity between the available TCR epitopes and the candidate epitopes (Table 5). This was more so for HLA-A*24:02, presumably because of the smaller set of known TCR epitopes for this allele. As expected, most matching amino acids correspond to each allele's anchor residues.

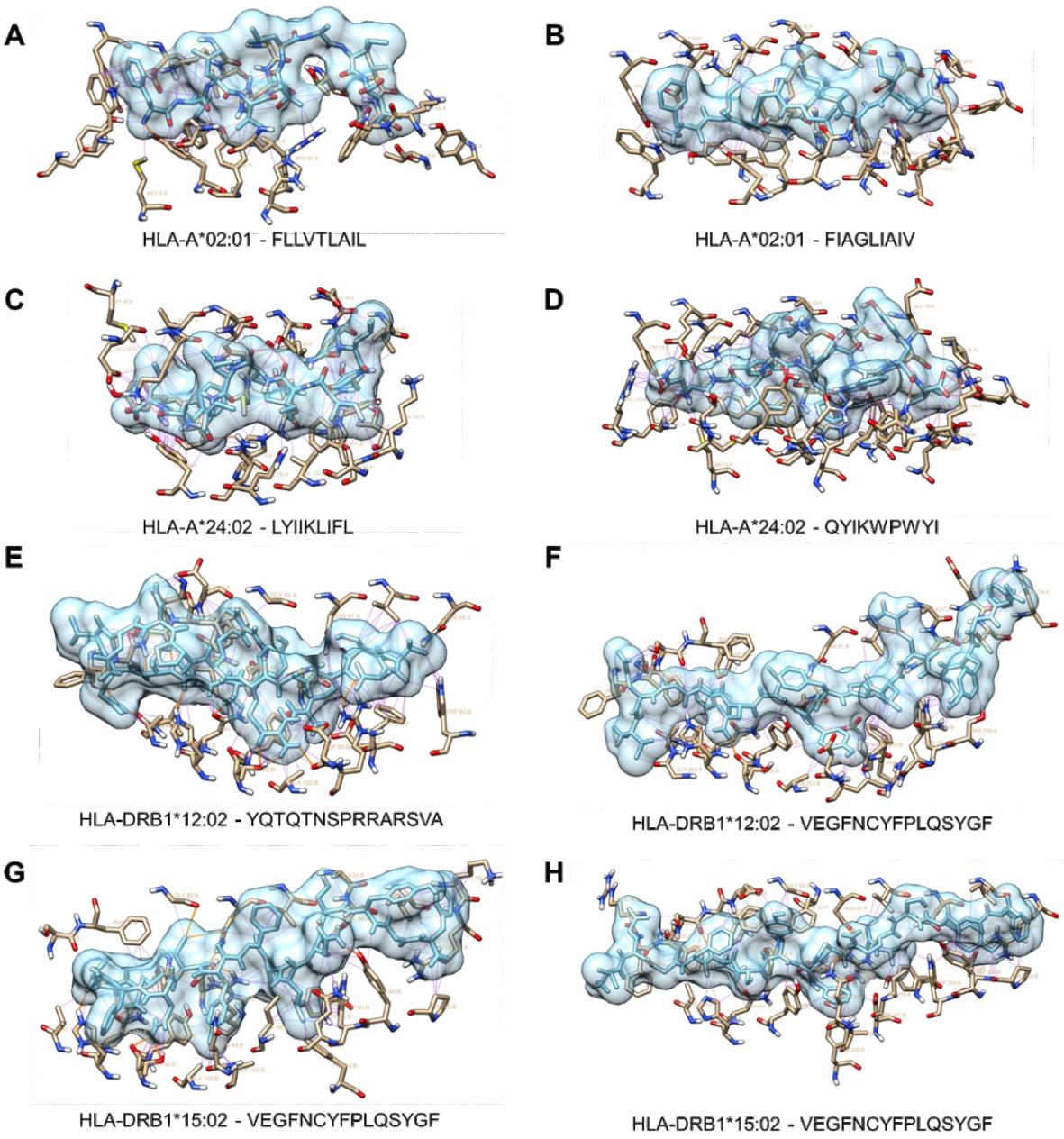

Figure 1. *Interaction diagrams of the top SARS-CoV-2 epitopes per HLA allele according to de novo (A,C,E,G) and template-based (B,D,F,H) molecular docking with the HLA peptide-binding domain. Close contacts are shown as purple lines while hydrogen bonds are shown as orange lines.*

Proceeding with the analysis despite this limitation, binding energies were determined for a homology model (TCRpMHCmodels), as well as models that use the TCR structures from this initial model and pMHC structures from the de novo and template-based peptide-HLA docking steps, where binding affinity trends differed among these three models for both alleles (Table 6). For HLA-A*02:01, the top epitope from the homology model agreed with the de novo model whereas it agreed with the template-based model for HLA-A*24:02. Apart from this no other similarities in the predictions between models were observed, making it difficult to hypothesize as to which might be the most reliable method based on this data alone. Notably, the strongest binders predicted for HLA-A*02:01 in the peptide-HLA docking also had the highest affinity to their corresponding TCR in this analysis. That is, the E epitope for the de novo model and the S epitope for the template-based model, suggesting that these might be the most promising epitopes for this allele. There is no such consensus in the HLA and TCR docking results for HLA-A*24:02, although the

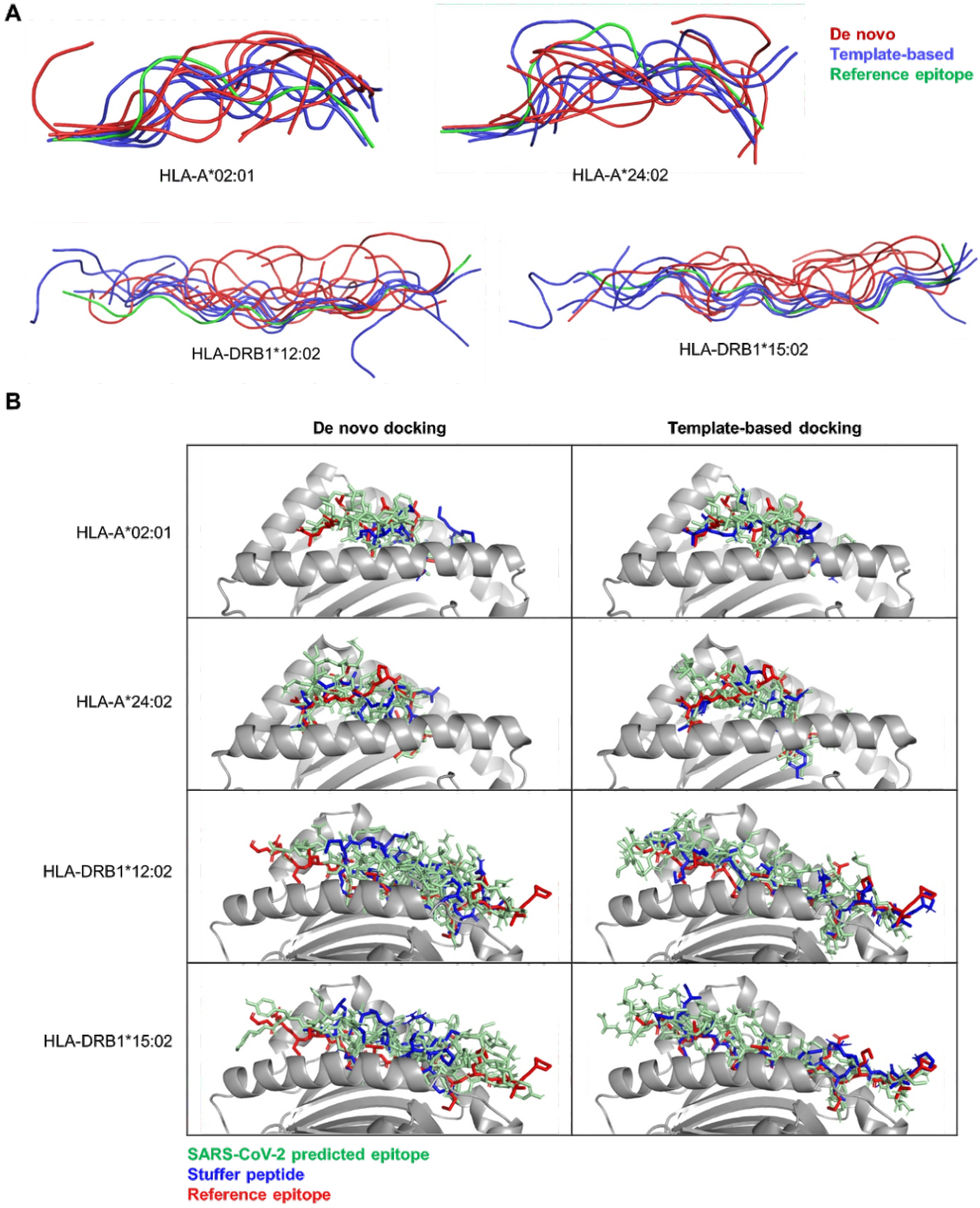

Figure 2. *Comparison of ligand conformations of SARS-CoV-2 predicted epitopes. Reference epitopes were obtained from experimental pMHC structures. For the HLA-A alleles, the original ligands of the HLA structures (PDB ID: 3D25, 3WL9) were used as the reference. Experimental structures of ligands bound on HLA-DRB1*03:01 (PDB ID: 1A6A) and HLA-DRB1*15:01 (PDB ID: 1BX2) were used for HLA-DRB1*12:02 and HLA-DRB1*15:02, respectively. There are currently no available pMHC structures for these two alleles. (A) Side view of ligand backbone conformations. (B) Ligand structures on the MHC molecule.*

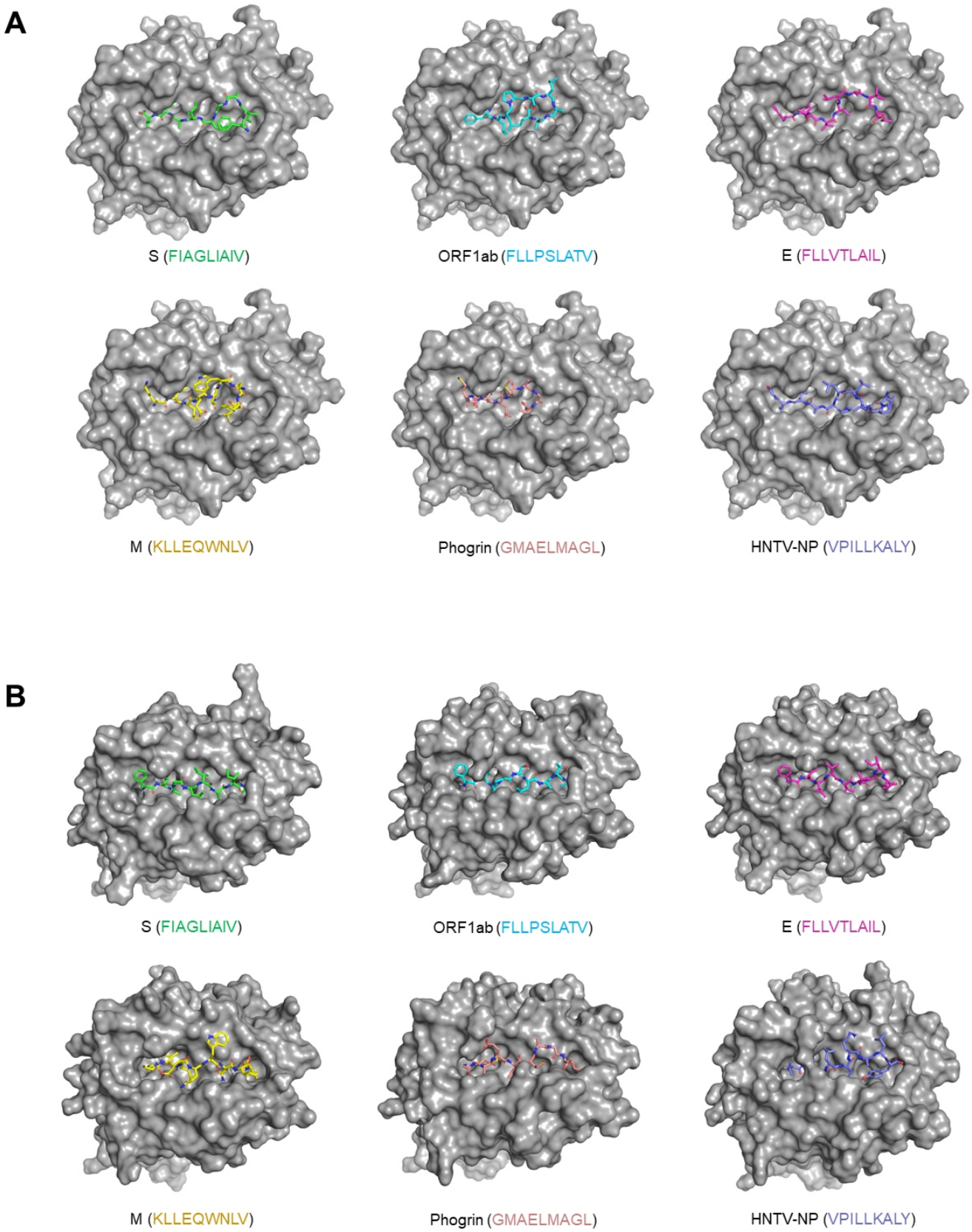

Figure 3. *Conformations of the predicted epitopes in the HLA-A*02:01 peptide-binding cleft predicted by de novo (A) and template-based (B) molecular docking. The stuffer (phogrin) and negative control (HNTV-NP) peptides are also shown.*

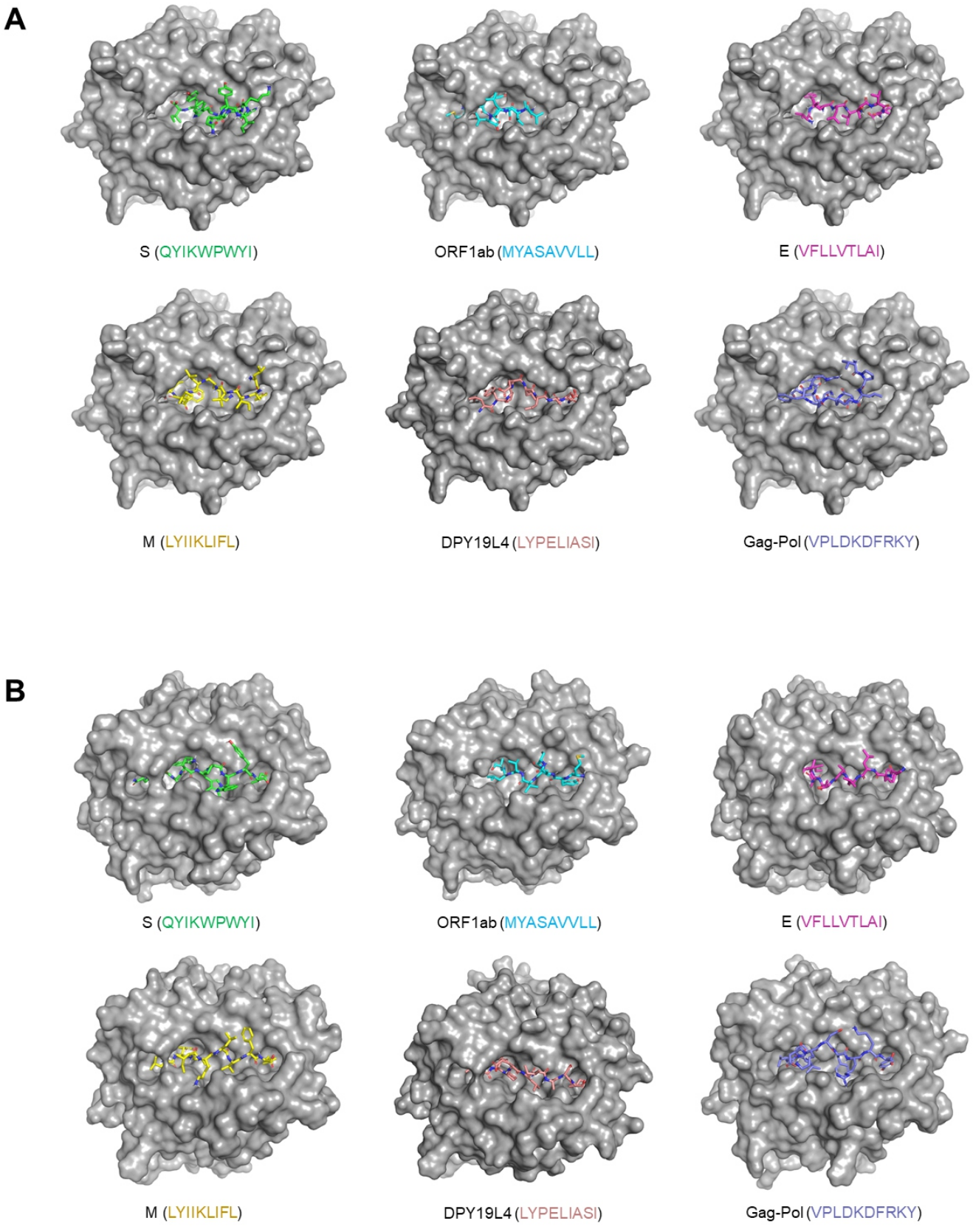

Figure 4. *Conformations of the predicted epitopes in the HLA-A*24:02 peptide-binding cleft predicted by de novo (A) and template-based (B) molecular docking. The stuffer (DPY19L4) and negative control (Gag-Pol) peptides are also shown.*

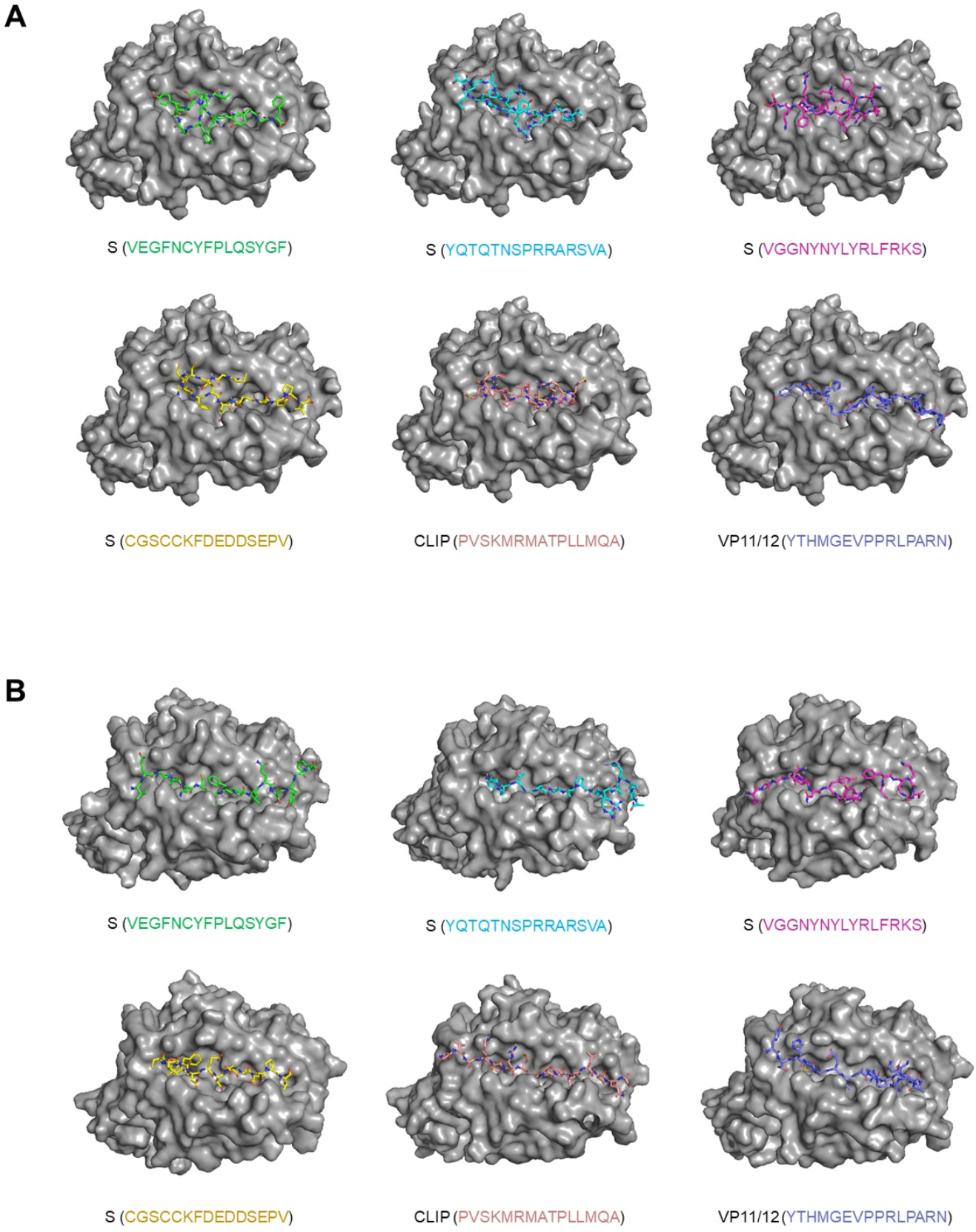

Figure 5. *Conformations of the predicted epitopes in the HLA-DRB1*12:02 peptide-binding cleft predicted by de novo (A) and template-based (B) molecular docking. The stuffer (CLIP) and negative control (VP11/12) peptides are also shown.*

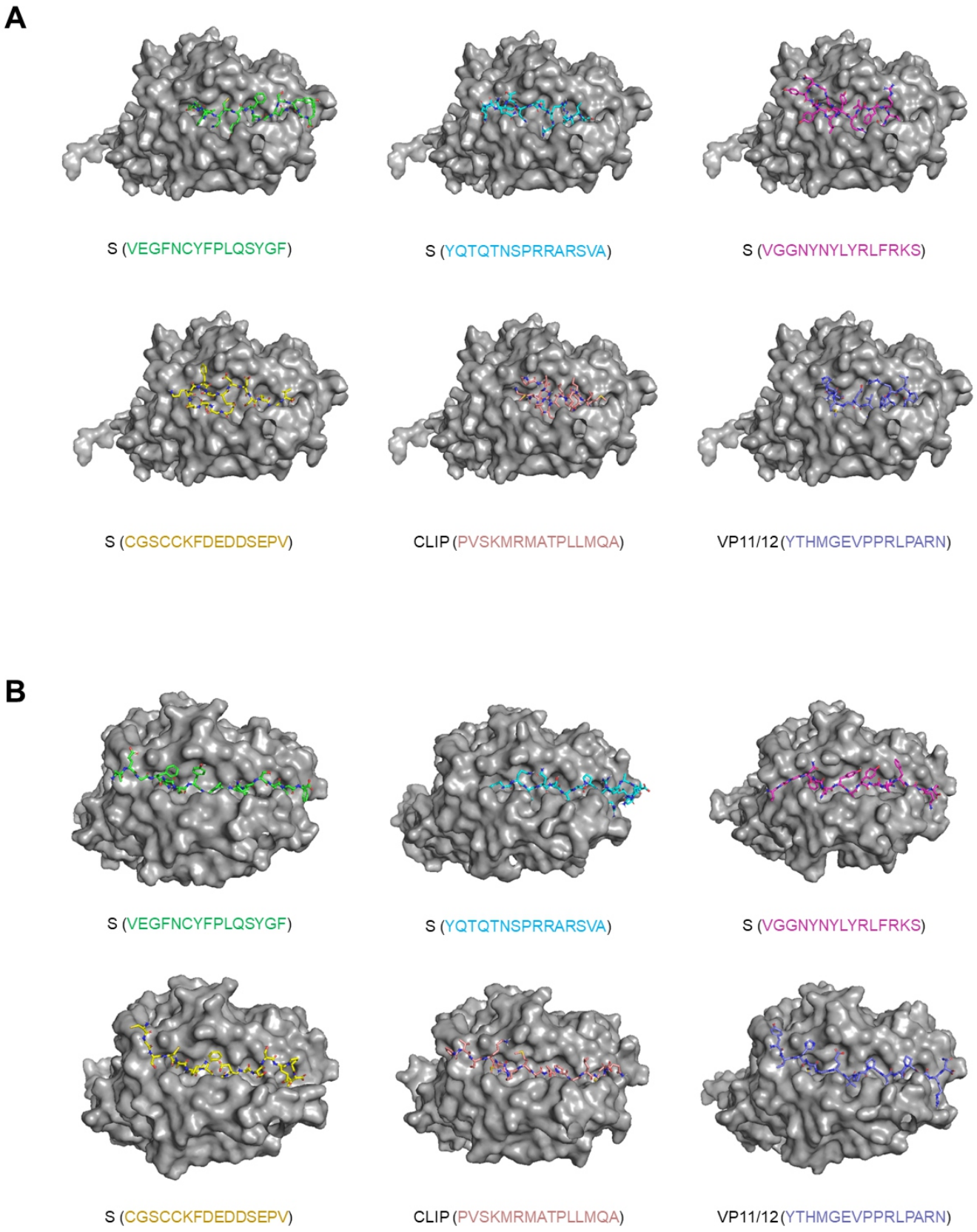

Figure 6. *Conformations of the predicted epitopes in the HLA-DRB1*15:02 peptide-binding cleft predicted by de novo (A) and template-based (B) molecular docking. The stuffer (CLIP) and negative control (VP11/12) peptides are also shown.*

Table 5. *Percent identity shared between the sequences of each SARS-CoV-2 peptide and its most similar TCR epitope for HLA class I alleles. The amino acids in bold indicate matching residues.*

HLA Allele	Protein	SARS-CoV-2 Epitope Sequence	TCR Epitope Sequence	% Identity
A*02:01	S	**F**IAG**L**IAIV	**F**LYA**L**ALLL	22.22
	ORF1ab	F**LLP**SLA**TV**	NL**VP**MVA**TV**	55.56
	E	F**L**LV**TL**A**IL**	S**L**FN**TV**A**TL**	44.44
	M	**KLLE**QWNL**V**	**KVLE**YVIK**V**	44.44
A*24:02	S	Q**Y**IKWPWY**I**	A**Y**AQKIFK**I**	22.22
	ORF1ab	M**Y**ASAVV**L**L	Q**Y**DPVAA**L**F	22.22
	E	VFLL**V**TLAI	QYDP**V**AALF	11.11
	M	L**Y**II**K**LIFL	A**Y**AQ**K**IFKI	22.22

Table 6. *Predicted binding energies ($\Delta G_{binding}$) of the SARS-CoV-2 peptide-MHC complexes from homology (TCRpMHCmodels), de novo, and template-based docking methods to TCR. The values in bold indicate the top predicted epitopes per method per HLA class I allele.*

HLA Allele	Protein	SARS-CoV-2 Epitope Sequence	$\Delta G_{binding}$ (kcal/mol)		
			TCRpMHCmodels	De novo	Template-based
A*02:01	S	FIAGLIAIV	-14.0	-15.1	**-15.4**
	ORF1ab	FLLPSLATV	-12.4	-13.2	-15.0
	E	FLLVTLAIL	**-14.9**	**-15.9**	-14.4
	M	KLLEQWNLV	-14.8	-13.7	-13.1
A*24:02	S	QYIKWPWYI	-11.5	-14.1	-13.0
	ORF1ab	MYASAVVLL	-10.4	**-15.5**	-11.4
	E	VFLLVTLAI	-8.3	-15.4	-11.9
	M	LYIIKLIFL	**-11.9**	-13.5	**-13.1**

strongest HLA binder predicted from the de novo method showed the highest affinity to its TCR based on the homology and template-based models. That is, the M epitope.

Structurally, the initial TCR-pMHC models, which were derived through homology modelling, resembled their corresponding templates (Figures 7-8). As the structures were refined for the de novo and template-based models, the orientation of the TCR relative to the pMHC changed for some, most noticeably in both the de novo and template-based models for the S and M epitopes for HLA-A*24:02 (Figure 8). The binding geometry of TCR to pMHC has been shown to affect induction of TCR signalling, so developing a method that reliably predicts TCR-pMHC structure could potentially be useful in predicting peptide immunogenicity [58].

Discussion

Immunogenicity of an antigen requires binding with sufficient affinity to both the MHC and the TCR. Candidate SARS-CoV-2 epitopes identified by sequence-based algorithms were subjected to molecular docking on HLA molecules – de novo docking using PEP-FOLD3 and AutoDock Vina and template-based docking using GalaxyPepDock followed by semi-flexible redocking in AutoDock Vina. The binding affinity of the TCR-peptide-MHC complexes were also predicted using homology modeling. Multiple methods for predicting HLA binding were used in order to provide a more reliable prediction. Positive results across the different methods for a candidate peptide would reinforce its potential for vaccine development. VEGFNCYFPLQSYGF showed the most promising HLA binding among the predicted class II epitopes. For HLA class I epitopes, different peptides were predicted to have the highest affinity for the HLA molecules and, in the case of HLA-A*24:02, for the TCRs.

An important foundation of epitope prediction is that MHC binding is mostly determined by anchor residues, although other residues contribute as well [59,60]. SYFPEITHI's epitope prediction algorithm involves generating all possible oligomers of specified length, in this case nonamers, for the input sequence and scoring all amino acids in each oligomer based on criteria that includes the prevalence of that amino acid in ligands for that HLA allele, how frequently the amino acid occurs in anchor positions, whether it's considered unfavorable for binding, and such [30]. Meanwhile, IEDB employs multiple available methods to predict HLA class II epitopes, including an IEDB recommended method that currently uses a consensus

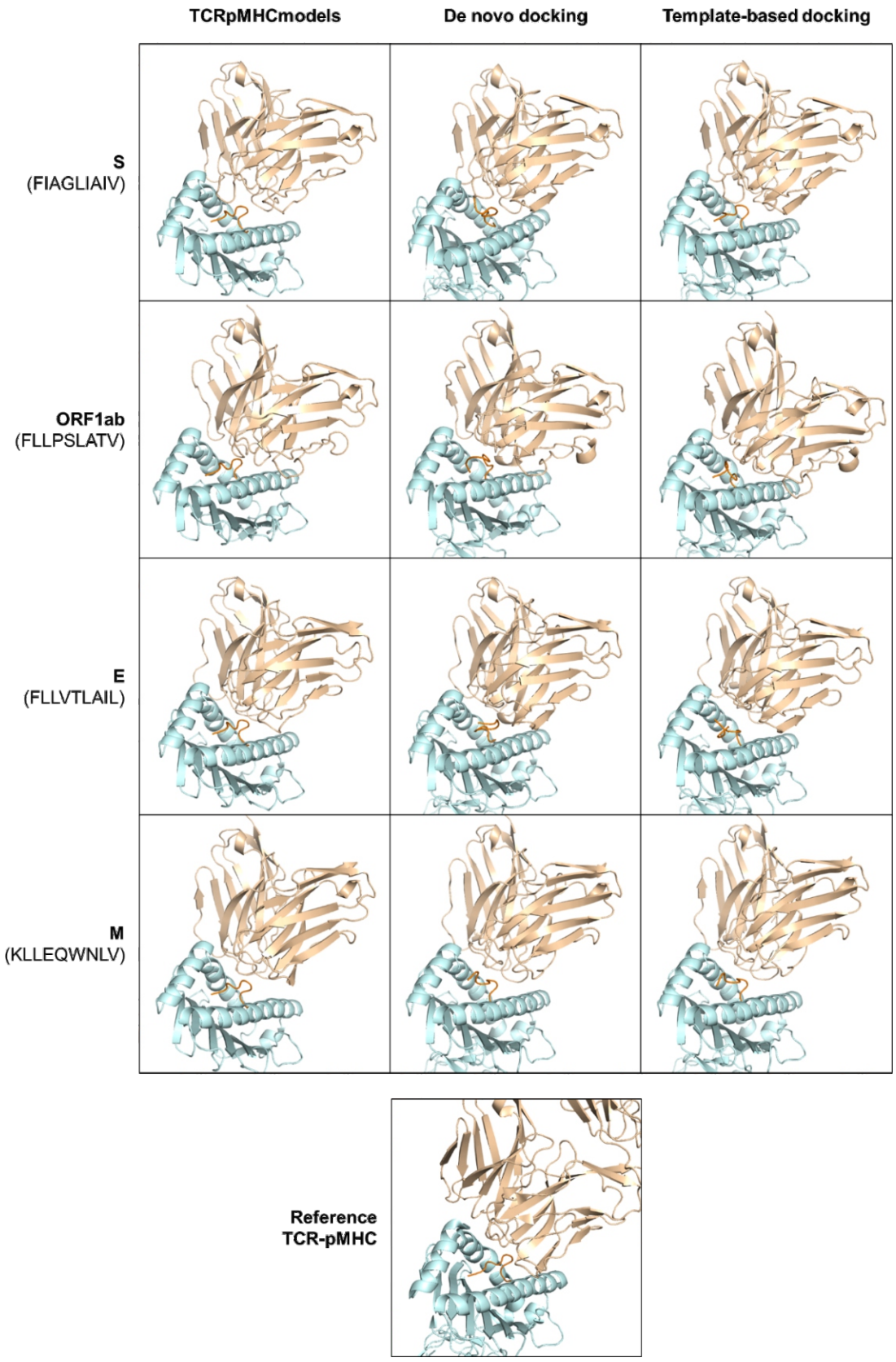

Figure 7. *TCR-pMHC complexes formed by SARS-CoV-2 peptides for HLA-A*02:01 predicted by homology modeling (TCRpMHCmodels), and de novo and template-based molecular docking. A reference structure for TCR-bound HLA-A*02:01 was obtained from the RCSB PDB (PDB ID: 2VLJ).*

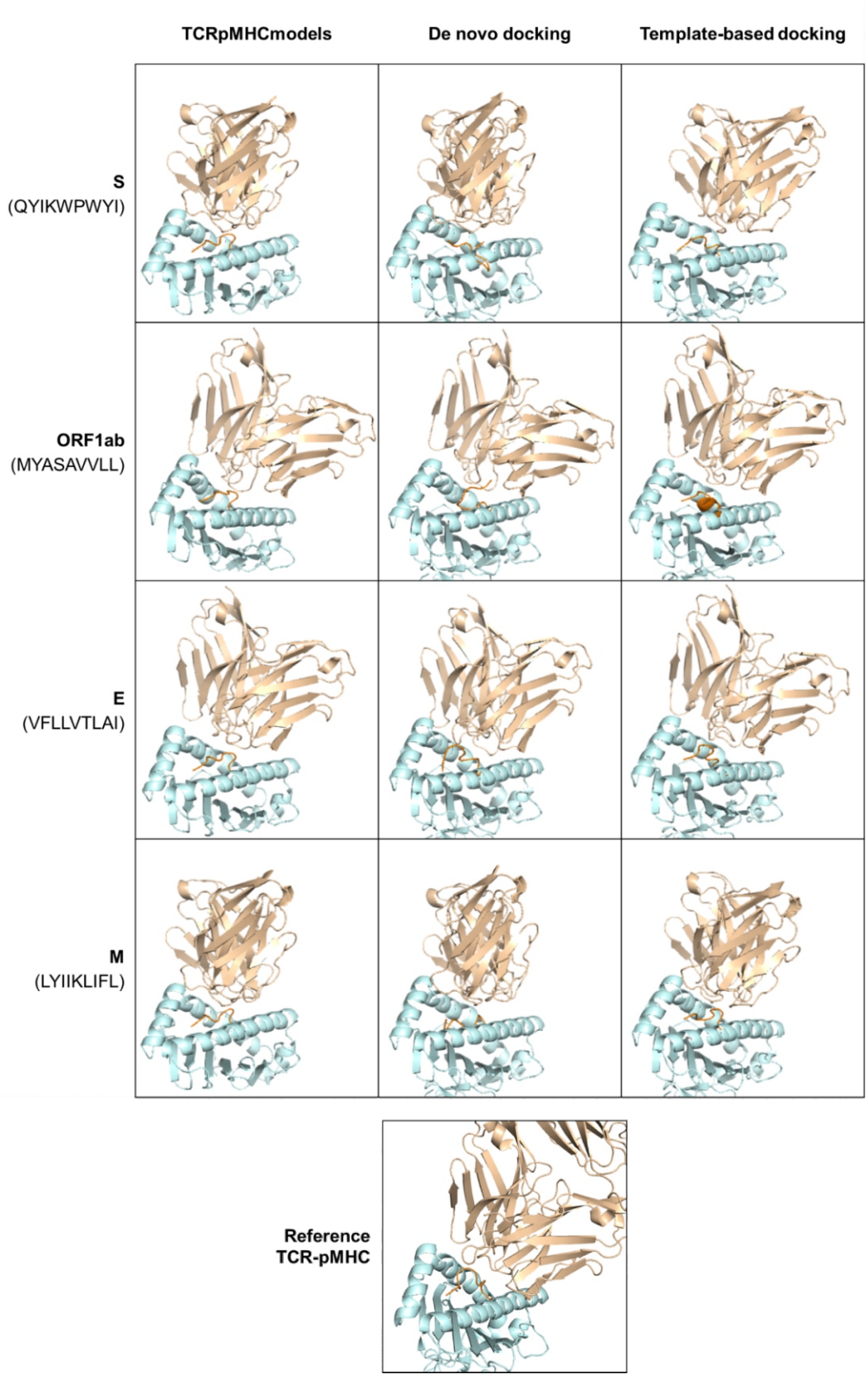

Figure 8. *TCR-pMHC complexes formed by SARS-CoV-2 peptides for HLA-A*24:02 predicted by homology modeling (TCRpMHCmodels), and de novo and template-based molecular docking. A reference structure for TCR-bound HLA-A*24:02 was obtained from the RCSB PDB (PDB ID: 3VXS).*

approach that combines several molecule-specific predictors when these are available for the specified allele, and uses pan-specific predictors otherwise [61-65]. While structure-based methods are available, sequence-based epitope prediction tools such as those in SYFPEITHI and IEDB are generally preferred due to the poorer predictive performance and longer computation time associated with structure-based approaches [66]. Nevertheless, structural information may still guide sequence-based epitope prediction. For example, molecular docking has been used to investigate peptide binding to MHC class I and II [67,68].

There are two general methods for protein-peptide docking – template-based and template-free or de novo docking. Template-based docking uses known structures of complexes or complex interfaces with similar sequences to model the complex formed between two target molecules. While template-based docking generally provides good predictive performance when template structures are available, it is limited by the availability of template structures and by the fact that interface architecture is not always similar for similar interactions. On the other hand, de novo docking has a wider applicability since it is not dependent on available complex structures but generally has poorer predictive power [69].

In the peptide-HLA molecular docking performed, the de novo and template-based methods yielded conflicting results. The two methods did not agree as to the absolute and even relative affinities of the candidate epitopes for all HLA alleles. Further, no peptide was predicted to bind to its respective HLA by both methods. In the de novo method, PEP-FOLD3 first generates the native conformation of the peptide by free modelling, i.e. no input reference or receptor structure, and the top model is docked as a fully flexible ligand onto the rigid HLA molecule using AutoDock Vina. AutoDock Vina performs template-free docking using a scoring function derived from knowledge-based potentials and empirical information on conformation preferences of protein-ligand complexes and affinity measurements [42]. On the other hand, the template-based method initially uses GalaxyPepDock, which models peptide-HLA binding by using similar experimental structures as template while still allowing for structural flexibility for both the peptide and HLA molecule, and secondly re-docks the top output as a semi-flexible (i.e. rigid backbone) ligand and a rigid receptor using AutoDock Vina to allow readjustment of peptide side chains for the most stable bound conformation. This means that only the template-based method permits HLA flexibility, which might have allowed for more interactions to be established between the peptide and the HLA peptide-binding cleft, increasing the predicted affinities. Expectedly, hydrogen bonds were mainly predicted to hold the termini of the peptides to the HLA class I groove, but were found along the length of the peptide for HLA class II [70]. However, protein-peptide docking using AutoDock Vina has been shown to perform generally poorly for peptides longer than four residues so the de novo method, which relies solely on AutoDock Vina to find peptide-HLA interactions, might be less preferable to the template-based method that uses AutoDock Vina strictly for docking refinement [71]. Additionally, the epitope conformations and pMHC structures generated by template-based docking were more reminiscent of typical epitopes presented on HLA molecules. This is expected as this method used experimental HLA epitopes as scaffolds for peptide modelling, whereas de novo-docked peptides were free to take on conformations not likely to be taken by HLA epitopes.

While this suggests that the template-based method might be a better predictor of pMHC binding, some results for the stuffer and negative controls, even for the seemingly more reliable template-based docking, conflicted with what was expected, i.e. stuffer peptides ought to have higher affinity for the HLA molecule compared to the negative control. Poor epitope selection cannot be ruled out, as this is limited by data availability especially for certain HLA alleles. For example, while CLIP seems to be the intuitive choice as stuffer for HLA class II, it is known to bind to different alleles with varying affinity [72]. Hence, low-affinity CLIP binding, to a point where the negative control might exhibit stronger affinity for the HLA molecule, is conceivable. Ultimately, experimental T cell response data will be necessary to confirm the potential of either method as reliable epitope prediction strategies.

Lastly, there were several assumptions in the prediction of TCR binding: the TCRs are able to cross-react with epitopes of similar sequences, TCR-pMHC complexes adopt similar structures as those with similar sequences, and pMHC binding affinity for the TCR is a predictor of T cell response. TCR promiscuity is one of the factors that broadens the range of antigens that the host immune system can recognize [73]. However, epitope sequence similarity alone may not be sufficient basis for prediction of TCRs that can bind a certain pMHC complex since even a single amino acid polymorphism in the epitope can result in non-binding of a TCR in some cases [74]. Even if, optimistically, the TCRs used in this study were not as selective, the mere 11-56% sequence identity shared by the candidate peptides with the experimental TCR

epitopes may not be enough to result in actual binding to the selected TCRs. In addition to the limited data on epitope-specific TCRs, these make the selection of appropriate TCR sequences for TCR-pMHC modeling a major limitation of this study [75]. The development of computational tools for the prediction of TCRs that can recognize a given pMHC complex, perhaps by incorporating TCR epitope sequence data as well as determinants of TCR promiscuity, could improve in silico approaches of vaccine design.

Other factors that could influence the T cell response to the predicted SARS-CoV-2 epitopes include conformational dynamics and kinetic stability of the pMHC and TCR-pMHC complexes [76-79]. Furthermore, multiple TCRs can recognize a given epitope [80,81]. The study did not account for TCR avidity since only one TCR was evaluated per peptide. Experimental studies, such as in vitro IFN-γ production of T cells, would have to be conducted to validate whether these peptides indeed elicit immune responses. It would be interesting to determine whether experimental data correlate well with the results presented in this study, and which among the methods used produced more accurate results.

Conclusions

In this study, sequence-based and structure-based computational tools were used to identify potentially immunogenic SARS-CoV-2 epitopes for vaccine development. However, results varied considerably between the different prediction methods used, likely due to the different assumptions and limitations associated with each method. Template-based molecular docking may have produced the more accurate results than de novo docking based on the structures of the generated models but this will have to be validated by experimental data. As part of a continuing project on the development of a dendritic cell vaccine for COVID-19, the next stage would be the in vitro validation of the immunogenicity of each candidate peptide.

References

1. World Health Organization. (2020) WHO Coronavirus Disease 2019 (COVID-19) Dashboard.
2. Morawska L, Milton DK. (2020) It is Time to Address Airborne Transmission of COVID-19. Clinical Infectious Diseases, (ciaa939).
3. World Health Organization. (2020) Transmission of SARS-CoV-2: implications for infection prevention precautions.
4. Santarpia JL, Rivera DN, Herrera V, *et al*. (2020) Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center. MedRxiv, 2020.03.23.20039446.
5. Sungnak W, Huang N, Bécavin C. *et al.* (2020) SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nature Medicine, 26(5):681–687.
6. Zhou P, Yang XL, Wang XG. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798): 270–273.
7. Hoffmann M, Kleine-Weber H, Schroeder S. *et al*. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell, 181(2):271-280.e8.
8. Yuki K, Fujiogi M, Koutsogiannaki S. (2020). COVID-19 pathophysiology: A review. Clinical Immunology, 215;108427.
9. Gupta A, Madhavan MV, Sehgal K, Nair N, Mahajan S, *et al*. (2020) Extrapulmonary manifestations of COVID-19. Nature Medicine, 1–16.
10. Zhou F, Yu T, Du R. *et al*. (2020) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet, 395(10229):1054–1062.
11. Dong E, Du H, Gardner L. (2020) An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases, 20:533–534.
12. Center for Strategic and International Studies. (2020) Southeast Asia COVID-19 Tracker.
13. Nicola M, O'Neill N, Sohrabi C, Khan M, Agha M, Agha R. (2020) Evidence based management guideline for the COVID-19 pandemic - Review article. International Journal of Surgery, 77: 206–216.
14. World Health Organization. (2020). Clinical management of COVID-19: interim guidance.
15. Cunningham AC, Goh HP, Koh D. (2020) Treatment of COVID-19: Old tricks for new challenges. Critical Care, 24:91.
16. Fragkou PC, Belhadi D, Peiffer-Smadja N, *et al*. (2020) Review of trials currently testing treatment and prevention of COVID-19. Clinical Microbiology and Infection.
17. Cowling BJ, Aiello AE. (2020) Public Health Measures to Slow Community Spread of Coronavirus Disease 2019. The Journal of Infectious Diseases, 221(11):1749–1751.
18. World Health Organization. (2020). Draft landscape of COVID-19 candidate vaccines.
19. Amanat F, Krammer F. (2020) SARS-CoV-2 Vaccines: Status Report. Immunity, 52:583–589.

20. Kato T, Matsuda T, Ikeda Y. *et al*. (2018) Effective screening of T cells recognizing neoantigens and construction of T-cell receptor-engineered T cells. Oncotarget, 9(13):11009–11019.
21. Kumar BV, Connors TJ, Farber DL. (2018) Human T Cell Development, Localization, and Function throughout Life. Immunity, 48:202–213.
22. Chen M, Li YG, Zhang DZ. *et al*. (2005). Therapeutic effect of autologous dendritic cell vaccine on patients with chronic hepatitis B: A clinical study. World Journal of Gastroenterology, 11(12): 1806–1808.
23. Echeverria I, Pereboev A, Silva L. *et al*. (2011) Enhanced T cell responses against hepatitis C virus by *ex vivo* targeting of adenoviral particles to dendritic cells. Hepatology, 54(1):28–37.
24. Gay CL, Debenedette MA, Tcherepanova IY. *et al*. (2018). Immunogenicity of AGS-004 Dendritic Cell Therapy in Patients Treated during Acute HIV Infection. AIDS Research and Human Retroviruses, 34(1): 111–122.
25. Leplina O, Starostina N, Zheltova O, Ostanin A, Shevela E, Chernykh E. (2016) Dendritic cell-based vaccines in treating recurrent herpes labialis: Results of pilot clinical study. Human Vaccines and Immunotherapeutics, 12(12): 3029–3035.
26. Wieczorek M, Abualrous ET, Sticht J. et al. (2017). Major histocompatibility complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation. Frontiers in Immunology, 8:1.
27. Grifoni A, Weiskopf D, Ramirez SI. *et al*. (2020). Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell, 181(7):1489-1501.e15.
28. Woo PCY, Huang Y, Lau SKP, Yuen KY. (2010) Coronavirus genomics and bioinformatics analysis. Viruses, 2:1805–1820.
29. Gonzalez-Galarza FF, McCabe A, Santos EJM. *et al.* (2020) Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Research, 48(D1),D783–D788.
30. Rammensee HG, Bachmann J, Emmerich NPN, Bachor OA, Stevanović S. (1999) SYFPEITHI: Database for MHC ligands and peptide motifs. Immunogenetics, 50:213–219.
31. Vita R, Overton JA, Greenbaum JA. *et al.* (2015) The immune epitope database (IEDB) 3.0. Nucleic Acids Research, 43(D1), D405–D412.
32. Waterhouse A, Bertoni M, Bienert S. *et al*. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research, 46(W1), W296–W303.
33. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. (2020) IPD-IMGT/HLA Database. Nucleic Acids Research, 48(D1), D948–D955.
34. Ouyang Q, Standifer NE, Qin H, *et al*. (2006) Recognition of HLA class I-restricted β-cell epitopes in type 1 diabetes. Diabetes, 55(11):3068–3074.
35. Ghosh P, Amaya M, Mellins E, Wiley DC. (1995) The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. Nature, 378(6556):457–462.
36. Ma Y, Cheng L, Yuan B. *et al*. (2016) Structure and function of HLA-A*02-restricted hantaan virus cytotoxic T-cell epitope that mediates effective protective responses in HLA-A2.1/Kb transgenic mice. Frontiers in Immunology, 7(AUG).
37. Han C, Kawana-Tachikawa A, Shimizu A, *et al*. (2014) Switching and emergence of CTL epitopes in HIV-1 infection. Retrovirology, 11(1): 38.
38. Srivastava R, Coulon P-GA, Prakash S, *et al*. (2020) Human Epitopes Identified from Herpes Simplex Virus Tegument Protein VP11/12 (UL46) Recall Multifunctional Effector Memory CD4 + T EM Cells in Asymptomatic Individuals and Protect from Ocular Herpes Infection and Disease in "Humanized" HLA-DR Transgenic M. Journal of Virology, 94(7).
39. Pettersen EF, Goddard TD, Huang CC. *et al.* (2004) UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25(13), 1605–1612.
40. 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.
41. Lee H, Heo L, Lee MS, Seok C. (2015) GalaxyPepDock: A protein-peptide docking tool based on interaction similarity and energy optimization. Nucleic Acids Research, 43(W1), W431–W435.
42. Trott O, Olson AJ. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2), 455–461.
43. Shugay M, Bagaev DV, Zvyagin IV, *et al*. (2018). VDJdb: A curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Research, 46(D1), D419–D427.
44. Gowthaman R, Pierce BG. (2018) TCRmodel: High resolution modeling of T cell receptors from sequence. Nucleic Acids Research, 46(W1), W396–W401.
45. Jensen KK, Rantos V, Jappe EC, *et al*. (2019) TCRpMHCmodels: Structural modelling of TCR-pMHC class I complexes. Scientific Reports, 9(1): 1–12.

46. Heo L, Lee H, Seok C. (2016) GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. Scientific Reports, 6(1), 1–10.
47. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. (2016) PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. Bioinformatics, 32(23):3676–3678.
48. Trolle T, McMurtrey CP, Sidney J. *et al*. (2016) The Length Distribution of Class I–Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele–Specific Binding Preference. The Journal of Immunology, 196(4):1480–1487.
49. Apostolopoulos V, Yuriev E, Lazoura E, Yu M, Ramsland PA. (2008) MHC and MHC-like molecules: Structural perspectives on the design of the molecular vaccines. Human Vaccines, 4(6): 400–409.
50. Ruppert J, Sidney J, Celis E, Kubo RT, Grey HM, Sette A. (1993) Prominent role of secondary anchor residues in peptide binding to HLA-A2.1 molecules. Cell, 74(5): 929–937.
51. Ibe M, Ikeda Moore Y, Miwa K, Kaneko Y, Yokota S, Takiguchi M. (1996) Role of strong anchor residues in the effective binding of 10-mer and 11-mer peptides to HLA-A(*)2402 molecules. Immunogenetics, 44(4): 233–241.
52. Vaz Galvão da Silva BA, Chudzinski-Tavassi AM, Mesquita Pasqualoto KF. (2019) A Combined Computer-Aided Approach to Drive the Identification of Potential Epitopes in Protein Therapeutics. Journal of Pharmacy and Pharmaceutical Sciences, 21(1): 268–285.
53. Sercarz EE, Maverakis E. (2003) MHC-guided processing: Binding of large antigen fragments. Nature Reviews Immunology, 3:621–629.
54. Jones EY. (1997) MHC class I and class II structures. Current Opinion in Immunology, 9(1):75–79.
55. Fleri W, Paul S, Dhanda SK. *et al.* (2017) The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Frontiers in Immunology, 8:278.
56. Gfeller D, Bassani-Sternberg M. (2018) Predicting antigen presentation-What could we learn from a million peptides? Frontiers in Immunology, 9: 1.
57. Engelhard VH. (1994) Structure of Peptides Associated with Class I and Class II MHC Molecules. Annual Review of Immunology, 12(1): 181–207.
58. Adams JJ, Narayanan S, Liu B, *et al*. (2011) T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. Immunity, 35(5): 681–693.
59. Johansen TE, McCullough K, Catipovic B, Su XM, Amzel M, Schneck JP. (1997) Peptide Binding to MHC Class I is Determined by Individual Pockets in the Binding Groove. Scandinavian Journal of Immunology, 46(2): 137–146.
60. Sinigaglia F, Hammer J. (1994) Defining rules for the peptide-MHC class II interaction. Current Opinion in Immunology, 6(1): 52–56.
61. Kim Y, Ponomarenko J, Zhu Z, *et al*. (2012) Immune epitope database analysis resource. Nucleic Acids Research, 40(W1): 525–530.
62. Nielsen M, Lundegaard C, Lund O. (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics, 8:238.
63. Wang P, Sidney J, Kim Y, *et al*. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics, 11, 568.
64. Sturniolo T, Bono E, Ding J, *et al*. (1999) Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nature Biotechnology, 17(6):555–561.
65. Nielsen M, Lundegaard C, Blicher T, *et al*. (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. PLoS Computational Biology, 4(7), 1000107.
66. Toussaint NC, Kohlbacher O. (2009) Towards in silico design of epitope-based vaccines. Expert Opinion on Drug Discovery, 4:1047–1060.
67. Rosenfeld R, Zheng Q, Vajda S, DeLisi C. (1995) Flexible docking of peptides to class I major-histocompatibility-complex receptors. Genetic Analysis: Biomolecular Engineering, 12(1):1–21.
68. Tzakos AG, Fuchs P, Van Nuland NAJ, *et al.* (2004) NMR and molecular dynamics studies of an autoimmune myelin basic protein peptide and its antagonist: Structural implications for the MHC II (I-A u)-peptide complex from docking calculations. European Journal of Biochemistry, 271(16):3399–3413.
69. Ohue M, Matsuzaki Y, Shimoda T, Ishida T, Akiyama Y. (2013) Highly precise protein-protein interaction prediction based on consensus between template-based and de novo docking methods. BMC Proceedings, 7(7):1–10.
70. Liu J, Gao GF. (2011) Major Histocompatibility Complex: Interaction with Peptides. In eLS.
71. Rentzsch R, Renard BY. (2015) Docking small peptides remains a great challenge: An assessment using AutoDock Vina. Briefings in Bioinformatics, 16(6):1045–1056.

72. Honey K, Forbush K, Jensen PE, Rudensky AY. (2004) Effect of Decreasing the Affinity of the Class II-Associated Invariant Chain Peptide on the MHC Class II Peptide Repertoire in the Presence or Absence of H-2M1. The Journal of Immunology, 172(7):4142–4150.
73. Sewell AK. (2012) Why must T cells be cross-reactive? Nature Reviews Immunology, 12:669–677.
74. Nicholls S, Piper KP, Mohammed F, *et al*. (2009) Secondary anchor polymorphism in the HA-1 minor histocompatibility antigen critically affects MHC stability and TCR recognition. Proceedings of the National Academy of Sciences of the United States of America, 106(10): 3889–3894.
75. Meysman P, De Neuter N, Gielis S, Bui Thi D, Ogunjimi B, Laukens K. (2019) On the viability of unsupervised T-cell receptor sequence clustering for epitope preference. Bioinformatics, 35(9):1461–1468.
76. Natarajan K, Jiang J, May NA, *et al*. (2018) The role of molecular flexibility in antigen presentation and T cell receptor-mediated signaling. Frontiers in Immunology, 9,1657.
77. Lazarski CA, Chaves FA, Jenks SA, *et al*. (2005) The kinetic stability of MHC class II:Peptide complexes is a key parameter that dictates immunodominance. Immunity, 23(1):29–40.
78. Rasmussen M, Fenoy E, Harndahl M, *et al*. (2016) Pan-Specific Prediction of Peptide–MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity. The Journal of Immunology, 197(4):1517–1524.
79. Chen JL, Stewart-Jones G, Bossi G, *et al*. (2005) Structural and kinetic basis for heightened immunogenicity of T cell vaccines. Journal of Experimental Medicine, 201(8):1243–1255.
80. Chen G, Yang X, Ko A, *et al*. (2017) Sequence and Structural Analyses Reveal Distinct and Highly Diverse Human CD8+ TCR Repertoires to Immunodominant Viral Antigens. Cell Reports, 19(3):569–583.
81. Miconnet I, Marrau A, Farina A, *et al*. (2011) Large TCR Diversity of Virus-Specific CD8 T Cells Provides the Mechanistic Basis for Massive TCR Renewal after Antigen Exposure. The Journal of Immunology, 186(12):7039–7049.