# Conserved T-cell epitopes predicted by bioinformatics in SARS-COV-2 variants

## Running title: Conserved T cell epitopes in SARS-Cov-2

Feiyu Lu[1,2], Shengnan Wang[3], Ying Wang[1], Yunpeng Yao[1], Yangeng Wang[1], Shujun Liu[4], Yangyang Wang[1], Yongli Yu*[4], Liying Wang*[1]

[1]Department of Molecular Biology, College of Basic Medical Sciences and Institute of Pediatrics, The First Hospital of Jilin University, Jilin University, Changchun, 130021,China.

[2]Department of Pediatric Endocrinology, The First Hospital of Jilin University, Jilin University, Changchun,130021, China

[3]Institute of Antler Science and Product Technology, Changchun Sci-Tech University, Changchun, 130000, China

[4]Department of Immunology, College of Basic Medical Sciences, Jilin University, Changchun,130021, China

*Correspondence: Liying Wang or Yongli Yu, Department of Molecular Biology in College of Basic Medical Sciences/Institute of Pediatrics in The First Hospital of Jilin

Conflict of interest:The authors have no conflict of interest to declare.
Funding Statement:This study is financially supported by the National Nature Scientific Foundation of China (31670937).
Correspondence: Liying Wang or Yongli Yu, Department of Molecular Biology, College of Basic Medical Sciences and Institute of Pediatrics,The First Hospital of Jilin University or Department of Immunology, College of Basic Medical Sciences, Jilin University,Changchun, China.Email: wangliy@jlu.edu.cn or ylyu@jlu.edu.cn

University or Department of Immunology in College of Basic Medical Sciences, Jilin University. E-mail: wangliy@jlu.edu.cn or ylyu@jlu.edu.cn.

Word counts of the abstract: 193

Word counts of the text: 3358

## Abstract

**Background:** Finding conservative T cell epitopes in the proteome of numerous variants of SARS-COV-2 is required to develop T cell activating SARS-COV-2 capable of inducing T cell responses against SARS-COV-2 variants.

**Methods:** A computational workflow was performed to find HLA restricted $CD8^+$ and $CD4^+$ T cell epitopes among conserved amino acid sequences across the proteome of 474727 SARS-CoV-2 strains.

**Results:** A batch of covserved regions in the amino acid sequences were found in the proteome of the SARS-COV-2 strains. 2852 and 847 peptides were predicted to have high binding affinity to distint HLA class I and class II molecules. Among them, 1456 and 484 peptides are antigenic. 392 and 111 of the antigenic peptides were found in the conseved amino acid sequences. Among the antigenic-conserved peptides, 6 $CD8^+$ T cell epitopes and 7 $CD4^+$ T cell epitopes were identifed. The T cell epitopes could be presented to T cells by high-affinity HLA molecules which are encoded by the HLA alleles with high population coverage.

**Conclusions:**The T cell epitopes are conservative, antigenic and HLA presentable, and could be constructed into SARS-COV-2 vaccines for inducing protective T cell immunity against SARS-COV-2 and their variants.


**Key words:** SARS-COV-2; T cell epitopes; Bioinformatic; vaccine; mutant-resistant;

47 population coverage.

48

## Background

Severe acute respiratory syndrome-coronavirus 2 (SARS-COV-2) causing Coronavirus disease 19 (COVID-19) (1, 2) has been epidemic in the world for more than 19 months, with more than 191 million people infected and more than 4 million death, reported by World Health Organization (WHO) on June 14, 2021. To fight this, a variety vaccines against SARS-COV-2 have been developed at unprecedented speed. Among them, 108 candidate vaccines are in clinical phase and 22 vaccines have been authorized for emergency use(3, 4). Noticeably, these SARS-COV-2 vaccines mainly target the spike (S) protein which binds the receptor angiotensin-converting enzyme 2 (ACE2) of host cells for the SARS-COV-2 entry, and are believed to be able to induce the neutralizing antibodies specific to the S protein, thereby providing immune protection in the individuals who received the SARS-COV-2 vaccines (5, 6). However, variants of SARS-CoV-2 with S protein mutations have emerged around the world. Up to August 3$^{rd}$, 2021, 1,009,273 varied SARS-COV-2 genomes had been deposited into National Center for Biotechnology Information (NCBI) Virus Database (7). Among them, B.1.1.7 ,B.1.351, and P.1 were reported to be dominantly transmitted in the UK, South Africa and Brazil, respectively, and have spread to several other countries. The SARS-CoV-2 variants evolve as a result of mutation and natural selection for their favorable traits, one of which is to evade the host immunity, such as neutalizing antibodies, induced by infection or vaccines, raising concern on a growing number of

reinfection and a reduction in the effectiveness of SARS-CoV-2 vaccines in use (8-10). Evidently, the sporadic reinfection of SARS-COV-2 reported in USA, Belgium, Hong Kong of China and Ecuador (11-15) could be correlated to the SARS-COV-2 variants because of the varied genomes of SARS-COV-2 from first and second episodes of the infections. Recent studies revealed that the B.1.351 variant dodged the neutralizing antibodies induced by several SARS-COV-2 vaccines, indicating that the mutations could change the B cell epitopes in the S proteins (15-17) and thus calling for upgrading the SARS-COV-2 vaccines against the established and emerging variants.

Recently, vaccine-induced T cell responses have been noticed for up-grading SARS-COV-2 vaccines for resistant variants. Clinical studies showed that some cases of asymptomatic SARS-COV-2 exposure have been associated with cellular immune response without seroconversion, indicating that SARS-CoV-2 specific T cells could be relevant in disease control even in the absence of neutralizing antibodies. In COVID-19 patients, neutralizing antibody titers were not generally paralleled with reduced disease severity, but SARS-CoV-2 specific $CD4^+$ and $CD8^+$ T cells are each associated with milder disease (18-20). The data suggest that immunological memory of T cells, if efficiently induced by the SARS-COV-2 vaccines, could provide protection. Intrestingly, unlike B cells, which mainly target accessible proteins like S proteins or nucleocapsid phosphoprotein (N proteins) in SARS-COV-2, T cells can target all of viral proteins in SARS-COV-2 proteome. More importantly, the T cell responses could be less affected by the mutations, and some of T cell epitopes distributed in the SARS-COV-2 proteome

seem more stable than the B cell epitopes (21, 22). Therefore, SARS-CoV-2 vaccines capable of activating both $CD4^+$ and $CD8^+$ T cells are likely to induce the protection of SARS-CoV-2 variants. If so, it is required to develop the T cell vaccines, targeting the mutation-resistant T cell epitopes in the SARS-CoV-2 proteome.

Here, we systemically compare and analyze T cell epitopes in the initially identified SARS-CoV-2 and in the rapidly accumulated SARS-CoV-2 variants, with the aim of locating some of consevative T cell epitopes in the SARS-CoV-2 proteome for developing T cell vaccines to fight the SARS-CoV-2 variants.

## Methods

### Data retrieval

The complete sequence of SARS-CoV-2 isolate Wuhan-Hu-1 was retrieved from the nucleotide database available at NCBI with the accession number NC_045512.2 (23). It encodes 9860 amino acids translating into several non-structural proteins (NSP) like NSP1-16 and accessory proteins like open reading frame (ORF) 3a, 6, 7a, 7b, 8, 10, along with structural proteins including S protein, envelope (E) protein, membrane glycoprotein (M protein), and N protein. The accession number in the NCBI of these proteins are shown in Table S1.

### Selection of human leukocyte antigen (HLA) class I and HLA class II alleles

HLA class I and II alleles were selected on the basis of their occurrence worldwide. The worldwidely most frequent 27 HLA class I and 26 HLA class II alleles (24, 25), which accouting for the population coverage > 97% and >99% worldwide, rspectively . As shown in Table S2.

**Alignments of SARS-COV-2 strains**

474727 strains of SARS-CoV-2 sequences depositing to the Global Initiative of Sharing All Influenza Data (GISAID) database (26, 27) with the presenting date ranging from January 27$^{th}$, 2021 to April 27$^{th}$, 2021 on the date of April 27$^{th}$ 2021, were collected. Using the server of COVID-19 CoV Genetics (COVID CG) (27), the amino acid sequences of each protein of repertoire to the corrsponding amino acid suquences of hCoV-19/Wuhan/WIV04/2019 (MN996528.1), which sharing 100% homology with the reference sequence NC_045512.2 (23) were aligned. High frequency variations were defined as the ratio of occuring counts of amino acid mutations to 474727 greater than $10^{-3}$ (28).

**Prediction of SARS-CoV-2-derived CD8$^+$ T cell and CD4$^+$ T cell epitopes**

The sequences of 26 proteins encoded by the SARS-COV-2 genome were split into 9 amino acid-long peptides for CD8$^+$ T cell epitopes and 15 amino acid-long peptides for CD4$^+$ T cell epitopes covering the complete proteome of the SARS-COV-2. The NetMHCpan 4.1(29) server was utilized to screen the 9-mer peptides across the proteome

of the SARS-COV-2 (NC_045512.2) for their binding affinity with distinct HLA class I molecules encoded by selected 27 HLA class I alleles. The prediction of binding affinity was based on more than 850,000 quantitative binding affinity and mass-spectrometry eluted ligands peptides. The NetMHCpan 4.1 server analysis resulted in the evaluation about the binding affinity and the binding strength between 9-mer peptides and selected HLA class I molecules, expressed by cores of nanomolar IC50 and percentile rank respectively. Thresholds for high binding affinity were set ≤500 nM and strong binding strength were set at top 0.5%. The NetMHCIIPan 4.0 server (29) was used to predict the binding affinity and binding strength of the peptides (15 mer) across the proteome of the SARS-COV-2 (NC_045512.2) to the selected HLA class II molecules (Table S2). Thresholds for high binding affinity were set ≤500 nM and strong binding strength were set at top 1%. The antigenicity of screened peptides to both HLA class I and class II molecules was predicted using Vaxijen-2.0 (30, 31), which analyzed higher order interactions positions of protein sequence and exploiting the physicochemical properties (hydrophobicity, molecular size, polarity) of amino acids. Each of the peptides was scored for selecting. The peptides with the scores above 0.4 were determined antigenecity. Exclude the peptides located at the mutated positions based on the alignment of 474727 strains depositted to the GISAID database. The peptides presented by more HLA alleles (≥3 alleles) were rescreened with the hypothesis that increased HLA binding promiscuity meant broader population (32). The fraction of individuals capable of responding to each of the candidate T cell epitopes was calculated by IEDB population Coverage (33) based

on the coresponding HLA genotypic frequencies.

**Molecular docking of HLA molecules to $CD8^+$ T cell epitopes and $CD4^+$ T cell epitopes**

The spacial structure of HLA molecules were downloaded from pHLA database (34). Docking simulations between predicted $CD8^+$ T cell epitopes or $CD4^+$ T cell epitopes and presented HLA molecules were performed using the GalaxyPepDock server (35), which enables prediction of 3D protein-peptide complex structure interactions from input protein structure and peptide sequence using similar interactions found in the structure database and energy-based optimization. The approaches of graphing are python and Graphpad.

## Results

**Alignment of sequences of proteins encoded by SARS-COV-2 strains**

To develop T cell vaccines, targeting the mutation-resistant T cell epitopes in the SARS-CoV-2 proteome, we tried to screen universal T cell epitopes in the coserved amino acid sequences across SARS-COV-2 proteome. Firstly, we collected and comprehensively screened all of the 474727 SARS-COV-2 sequences with the presenting date ranging from January 27$^{th}$, 2021 to April 27$^{th}$,2021, deposited to GISAID database (26, 27). These strains were isolated from the Asia (9403/474727), Africa (961/474727),

Europe (315833/474727), Oceania (1225/474727), North America (144760/474727) and South America (2545/474727). To find the conserved amino acid sequences in the repertoire of the SARS-COV-2, using the server of COVID CG (27), we aligned the proteome of all these submitted sequences with the sequence of hCoV-19/Wuhan/WIV04/2019 (MN996528.1) (36) which shares 100% homology with the reference sequence NC_045512.2 (23). The alignment revealved that the amino acid mutations were widely distributed in the proteins across the SARS-COV-2 proteome, ranging from $2.12\times10^{-6}$ to $9.96\times10^{-1}$ in occuring frequency. The frequency higher than $10^{-3}$ was defined as relatively high frequency mutations (28). With this cut-off, we identified 899 amino acid substitutions and 20 amino acid deletions as relatively high frequency variations at 816 positions in the SARS-COV-2 proteome (Figure 1A). As for the positions, 117 are in the S protein, a target for inducing neutralizing antibodies; 153 in NSP3; 66 in NSP6 and ORF3a; 53 in NSP2 and 1 to 42 in the other proteins encoded by the SARS-COV-2 genome. Howerver, neverthless, we identifed a batch of positions which are covserved in the proteome of SARS-COV-2 (Figure 1B). Specifically, 1156 are in the S protein; 1792 and 903 in the NSP3 and NSP12 proteins, 12 to 585 in the rest proteins of SARS-COV-2 proteome.

**Prediction of $CD8^+$ T cell and $CD4^+$ T cell epitopes in the conserved amino acid sequences across SARS-CoV-2 proteome**

Next, we screened and predicted T cell epitopes in the repertoire of the SARS-COV-2

196 proteome. As for the $CD8^+$ T cell epitopes, using NetMHCpan 4.1 server (29), we
197 screened the 9-mer peptides across the proteome of the SARS-COV-2 (NC_045512.2) for
198 their binding affinity with distinct HLA class I molecules encoded by 27 HLA class I
199 alleles (Table S2). The alleles account for the population coverage > 97% worldwide (24).
200 Based on more than 850,000 quantitative binding affinity and mass-spectrometry eluted
201 ligands peptides, the NetMHCpan 4.1 server analysis resulted in the evaluation about the
202 binding affinity and the binding strength between 9-mer peptides and selected HLA class
203 I molecules, expressed by scores of nanomolar IC50 and percentile rank respectively.
204 Thresholds for high binding affinity were set ≤500 nM and strong binding strength were
205 set at top 0.5%. Eventually, 2852 of the peptides with high affinity to the distinct class I
206 molecule were selected for further analysis. As for the $CD4^+$ T cell epitopes, using the
207 NetMHCIIPan 4.0 server (29), we predicted the binding affinity and binding strength of
208 the peptides (15 mer) across the proteome of the SARS-COV-2 (NC_045512.2) to the
209 HLA class II molecules (Table S2) encoded by the elleles which account for the
210 population coverage >99% (25). Resultantly, 847 candidate HLA class II binding
211 peptides from the SARS-COV-2 proteome were predicted to have high binding affinity (≤
212 500 nM) and strong binding strength (percentile rank ≤1%).

213 Furthermore, we predicted wheather these candidate peptides paired with distinct
214 HLA molecules could efficiently bind T cell recepters (TCR) by analysing
215 hydrophobicity, molecular size, polarity of amino acids of the pepetids, using Vaxijen-2.0
216 (30, 31). Each of the peptides was scored by the server. The peptides scored above 0.4

were selected as the antigenic ones which turned out to be 1456 pepetides and 484 of pepetides, restricted by the selected HLA class I and class II molecules, respectively. Based on the above screening, we tried to locate T cell epitopes in the conserved amino acid sequences across the proteome of SARS-COV-2 by aligning the sequences of the selected T cell epitopes with the SARS-COV-2 sequences deposited to the GISAID database from January 27th, 2021 to April 27th, 2021. As shown in the Figure 2A, 392 T cell epitopes restricted to the selected HLA class I molecules were found. Specifically, 82 are in the NSP3, 69 in the NSP12, 43 in the S protein, 30 in the NSP13, 21 in the NSP4, 20 in the M protein,18 in the NSP15, 17 in the NSP2 and NSP16, 14,13 and 11 in the NSP14, NSP6 and N protein, 8 and 7 in the NSP5 and NSP8, 1 to 5 are in the ORF6, NSP10, ORF3a, NSP7, E protein, NSP1 and NSP9. And 111 candidate T cell epitopes restricted to selected HLA class II molecules were found. Among them, 34 are in the NSP3, 16 in the S protein, 15 in the NSP12, 13 in the NSP13, 11 in the NSP15, 1 to 4 in the NSP1-2,NSP4-8, NSP10, NSP14, NSP16, N and M proteins.

To find the universal T cell epitopes, we further assessed the HLA binding promiscuity of the selected T cell epitopes. Firstly, the T cell epitopes presented by one or two HLA molecules were excluded. As shown in the Figure 2B, 88 or 14 T cell epitopes restricted by more than two HLA class I or II molecules were selected as the candidate $CD8^+$ T cell epitopes or $CD4^+$ T cell epitopes, respectively. Then, based on the coresponding HLA genotypic frequencies, the fraction of individuals capable of responding to each of the candidate T cell epitopes was calculated by IEDB population

Coverage (33). The results showed that the potential coverage in populations was ranging from 5.38% to 52.34% for the 88 $CD8^+$ T cell epitopes, and 27.96% to 97.48% for 14 $CD4^+$ T cell epitopes (Figure 2B). 6 of CD8+ and 7 of CD4+ epitopes were selected finally, with the population coverage **>** 45% for $CD8^+$ T cell epitopes and **>** 90% for $CD4^+$ T cell epitopes respectively.

**The interactions between predicted T cell epitopes and presented HLA molecules**

Ultimately, we identified 6 $CD8^+$ T cell epitopes, designated as S 691-699, NSP3 950-958, NSP4 420-428, NSP12 123-131, NSP12 647-655 and NSP13 209-217, and 7 $CD4^+$ T cell epitopes, designated as S 310-324, NSP3 1134-1148, NSP12 778-792, NSP13 177-191, NSP13 413-427, NSP13 538-552, and NSP14 232-246, respectively. Specifically, among the $CD8^+$ T cell epitopes, 4, 3 or 2 can be presented by the HLA class I proteins encoded by the alleles of HLA A*02:01, A*02:03 and A*32:01, of HLA A*02:06 and B*08:01, or of HLA A*68:02; A*03:01, A*11:01, A*30:02, B*15:01 and B*35:01 (Figure 3A), respectively. All of the $CD4^+$ T cell epitopes can be presented by the HLA class II proteins encoded by the alleles of HLA DPA10103-DPB10401, DPA10201-DPB10101 and DPA10301-DPB10402. Also, among the $CD4^+$ T cell epitopes, 5 or 1 can be presented by HLA class II proteins encoded by the alleles of DPA10103-DPB10201 or DPA10201-DPB10501, respectively (Figure 3B).

At last, we conducted the docking simulations between the selected T cell epitopes and the HLA molecules using the GalaxyPepDock (35), a server which enables prediction

of 3D protein-peptide complex structure interactions from input protein structure and peptide sequence using similar interactions found in the structure database and energy-based optimization. Ten models of each epitope-HLA complex were generated on the basis of minimized energy scores. The scores of estimated accuracy represent the estimated fraction of correctly predicted binding site residues. We docked each of the selected $CD8^+$ T cell epitopes or $CD4^+$ T cell epitopes with the corresponding HLA molecules and scored the pairs. As shown in Table 1, all of the pairs between the 6 selected $CD8^+$ T cell epitopes and the distinct HLA class I molecules scored with estimated accuracy 1. The pairs between the 7 selected $CD4^+$ T cell epitopes and the distinct HLA class II molecules scored with estimated accuracy in a range from 0.792 to 0.852, which presents high grade 3D fits of the pairs. The results were further supporting that these epitopes should be strong binders to presented HLA molecules and promising candidates for vaccine development studies.

## Discussion

Using the bioinformatic tools, we screened and identified 6 $CD8^+$ T cell epitopes and 7 $CD4^+$ T cell epitopes in the conserved amino acid sequences among the 474727 SARS-COV-2 strains whose sequences are deposited to the GISAID. Notably, the inter-individual variation in the SARS-CoV-2 sequences is low, compared to many other RNA viruses (37, 38), in part because coronaviruses encode a 3'−5' exonuclease activity

(NSP14), which provides a proofreading function that enhances replication fidelity and limits viral sequence diversification (39). Thus, it is virtually possible to find T cell epitopes in the repertoire of the SARS-COV-2 proteins. Primarily, these T cell epitopes exist at the the essential amino acid sequences which are crutial for the SARS-CoV-2 life-cycle in the host cell. As shown in Fig 2B, two conserved T cell epitopes were identifed in S proteins among the 474727 SARS-CoV-2 strains. One is a $CD4^+$ T cell epitope (S310-324) which is partially overlapped with the receptor binding domain (RBD) of the S1 subunit (S319-541), and another is a $CD8^+$ T cell epitope (S691-699) which locates in the S1/S2 cleavage region (S690-709). The SARS-CoV-2 with mutations in the epitopes (S310-324, S691-699) may hamper the cell entry of the SARS-CoV-2. Two T cell epitopes were identified in the NSP3, a papain-like protease which is critical for the SARS-CoV-2 to yield mature functional proteins from a polyprotein (40, 41). One is $CD8^+$ T cell epitope (NSP3 950-958) which locates in the crutial domain of the papain-like protease. Another is $CD4^+$ T cell epitope (NSP3 1134-1148) which locates in a nucleic acid-binding domain (NAR). The SARS-CoV-2 with mutations in the epitopes (NSP3 950-958, NSP3 1134-1148) may lose its capacity to generate functional proteins. Two $CD8^+$ T cell epitops (NSP12 123-131; NSP12 647-655) and one $CD4^+$ T cell epitope (NSP12 778-792) are in the NSP12 protein, a RNA-dependent RNA polymerase (RdRp) which is required for generating SARS-CoV-2 RNAs (42). Three $CD4^+$ T cell epitopes (NSP13 177-191; NSP13 413-427; NSP13 538-552) and one $CD8^+$ T cell epitope (NSP13 209-217) is in the NSP13 protein, a helicase which can unwind DNA or RNA in

an nucleotide triphosphate (NTP)-dependent manner with a 5'>3' polarity (43), a process critical for SARS-CoV-2 replication-transcription. One $CD4^+$ T cell epitope (NSP14 232-246) is in the NSP14 protein, a 3'−5' exonuclease which provids a proofreading function that enhances replication fidelity and limits viral sequence diversification (44); One $CD8^+$ T cell epitope (NSP4 420-428) is identified in the NSP4 protein, a membrane-spanning protein containing transmembrane domain 2 which helps the SARS-CoV-2 to modify endoplasmic reticulum (ER) membrane in the host cell (45). Obviously, the 6 CD $8^+$ T cell epitopes and 7 $CD4^+$ T cell epitopes identified in this study are all located in the key proteins for SARS-CoV-2 life-cycle and therefore could be the universial T cell epotopes which are conservatively retained in all SARS-CoV-2 strains.

Hopefully, the conserved T cell epitopes presented here could be constructed into SARS-CoV-2 vaccines to induce more univerisal T cell responses aginst the SARS-CoV-2. Accumulating evidence showed that the T cell immunity, if being induced by the SARS-CoV-2 vaccines contained the T cell epitopes, could be benefitial to the patients infected with SARS-CoV-2. As reported, circulating SARS-CoV-2-specific $CD8^+$ T cells and $CD4^+$ T cells were identified in ~70% and 100% of COVID-19 convalescent patients , they are associated with the better outcomes of the COVID-19 patients (18, 46). Importantly, the protective specific T cell responses against the SARS-CoV-2, after being induced, could be more sustained compared to the neutralizing humoral responses (47). Functional SARS-CoV-2-specific T cell responses are retained at 6 months following infection (48). In contrast, SARS-CoV-2-specific antibody responses were waned after 1

month after symptom onset (49). Furthermore, SARS-CoV-2-specific memory $CD4^+$ T cells, if being induced prior to the natural infection, will help the SARS-CoV-2-specific B cell to launch rapid and robust antibody responses (46). Intrestingly, the SARS-CoV-2 needs to maintain the T cell epitopes revealed in this study for its life-cycle, even facing the evolution pressures caused by the acquired specific immunity induced by the natural infection of vaccination. Thus, the SARS-CoV-2 vaccines constructed with the T cell epitopes identified in this study could induce more univerisal T cell responses aginst the SARS-CoV-2 variants. Lastly, the revealed T cell eptitopes, due to their capacity of being presented by the HLA molecules encoded by widely distributed HLA alleles, could be ideal targets for developing novel SARS-CoV-2 vaccines which induce the protective T cell immunity in large populations worldwide.

## Acknowledgments

The authors would like to thank Zhengang Jiao for his help in the application of informatics software, as well as Mengyuan Kou, Wenting Lu, and Cuiyun Cui for their help to the authors in other aspects.

## Funding Statement

This study is financially supported by the National Nature Scientific Foundation of China

(31670937).

## Conflict of interest

The authors have no conflict of interest to declare.

## References

1. Anonymous. 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5:536-544.

2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 382:727-733.

3. WHO. 2021. COVID-19 vaccine tracker and landscape. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines. Accessed

4. Anonymous. 2021. Status of COVID-19 Vaccines within WHO EUL/PQ evaluation process.

5. Krammer F. 2020. SARS-CoV-2 vaccines in development. Nature 586:516-527.

6. The Lancet Respiratory M. 2021. Realising the potential of SARS-CoV-2 vaccines-a long shot? Lancet Respir Med 9:117.

7. Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schäffer AA, Brister JR.

2017. Virus Variation Resource - improved response to emergent viral outbreaks. Nucleic Acids Res 45:D482-d490.

8. Callaway E. 2021. Fast-spreading COVID variant can elude immune responses. Nature 589:500-501.

9. Callaway E, Ledford H. 2021. How to redesign COVID vaccines so they protect against variants. Nature 590:15-16.

10. Cohen J. 2021. South Africa suspends use of AstraZeneca's COVID-19 vaccine after it fails to clearly stop virus variant. Science doi:10.1126/science.abg9559.

11. Tillett RL, Sevinsky JR, Hartley PD, Kerwin H, Crawford N, Gorzalski A, Laverdure C, Verma SC, Rossetto CC, Jackson D, Farrell MJ, Van Hooser S, Pandori M. 2021. Genomic evidence for reinfection with SARS-CoV-2: a case study. Lancet Infect Dis 21:52-58.

12. Molina LP, Chow SK, Nickel A, Love JE. 2020. Prolonged Detection of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) RNA in an Obstetric Patient With Antibody Seroconversion. Obstet Gynecol 136:838-841.

13. Van Elslande J, Vermeersch P, Vandervoort K, Wawina-Bokalanga T, Vanmechelen B, Wollants E, Laenen L, André E, Van Ranst M, Lagrou K, Maes P. 2020. Symptomatic SARS-CoV-2 reinfection by a phylogenetically distinct strain. Clin Infect Dis doi:10.1093/cid/ciaa1330.

14. To KK, Hung IF, Ip JD, Chu AW, Chan WM, Tam AR, Fong CH, Yuan S, Tsoi HW, Ng AC, Lee LL, Wan P, Tso E, To WK, Tsang D, Chan KH, Huang JD, Kok KH, Cheng VC, Yuen KY. 2020. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis doi:10.1093/cid/ciaa1275.

15. Prado-Vivar B, Becerra-Wong M, Guadalupe JJ, Marquez S, Gutierrez B, Rojas-Silva P, Grunauer M, Trueba G, Barragan V, Cardenas P. COVID-19 Re-Infection by a Phylogenetically Distinct SARS-CoV-2 Variant, First Confirmed Event in South America. Social Science Electronic Publishing.

16. Overbaugh J. 2020. Understanding protection from SARS-CoV-2 by studying reinfection. Nat Med 26:1680-1681.

17. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE, Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, Da Silva J, Xu J, Colbert RA, Patel R, Dizon J, Unson-O'Brien C, Shimeliovich I, Gazumyan A, Caskey M, Bjorkman PJ, Casellas R, Hatziioannou T, Bieniasz PD, Nussenzweig MC. 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature doi:10.1038/s41586-021-03324-6.

18. Rydyznski Moderbacher C, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, Belanger S, Abbott RK, Kim C, Choi J, Kato Y, Crotty EG, Kim C, Rawlings SA, Mateus J, Tse LPV, Frazier A, Baric R, Peters B, Greenbaum J, Ollmann Saphire E, Smith DM, Sette A, Crotty S. 2020. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. Cell 183:996-1012.e19.

19. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, Dejnirattisai W, Rostron T, Supasa P, Liu C, Lopez-Camacho C, Slon-Campos J, Zhao Y, Stuart D, Paeson G, Grimes J, Antson F, Bayfield OW, Hawkins DE, Ker DS, Turtle L, Subramaniam K, Thomson P, Zhang P, Dold C, Ratcliff J, Simmonds P, de Silva T, Sopp P, Wellington D, Rajapaksa U, Chen YL, Salio M, Napolitani G, Paes W, Borrow P, Kessler B, Fry JW, Schwabe NF, Semple MG, Baillie KJ, Moore S, Openshaw PJ, Ansari A, Dunachie

S, Barnes E, Frater J, Kerr G, Goulder P, Lockett T, et al. 2020. Broad and strong memory CD4 (+) and CD8 (+) T cells induced by SARS-CoV-2 in UK convalescent COVID-19 patients. bioRxiv doi:10.1101/2020.06.05.134551.

20. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, Rawlings SA, Sutherland A, Premkumar L, Jadi RS, Marrama D, de Silva AM, Frazier A, Carlin AF, Greenbaum JA, Peters B, Krammer F, Smith DM, Crotty S, Sette A. 2020. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell 181:1489-1501.e15.

21. Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez SI, Dan JM, Burger ZC, Rawlings SA, Smith DM, Phillips E, Mallal S, Lammers M, Rubiro P, Quiambao L, Sutherland A, Yu ED, da Silva Antunes R, Greenbaum J, Frazier A, Markmann AJ, Premkumar L, de Silva A, Peters B, Crotty S, Sette A, Weiskopf D. 2020. Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. Science 370:89-94.

22. Jain N, Shankar U, Majee P, Kumar A. 2021. Scrutinizing the SARS-CoV-2 protein information for designing an effective vaccine encompassing both the T-cell and B-cell epitopes. Infect Genet Evol 87:104648.

23. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in China. Nature 579:265-269.

24. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, Broadwater A, Kolla RV, De Silva AD, de Silva AM, Mattia KA, Doranz BJ, Grey HM, Shresta S, Peters B, Sette A.

2013. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc Natl Acad Sci U S A 110:E2046-53.

25. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. 2011. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics 63:325-35.

26. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall 1:33-46.

27. Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill 22.

28. Xu W, Wang M, Yu D, Zhang X. 2020. Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. Front Immunol 11:565278.

29. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. 2020. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 48:W449-w454.

30. Doytchinova IA, Flower DR. 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8:4.

31. Doytchinova IA, Flower DR. 2007. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 25:856-66.

32. Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. 2020. Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. Sci Rep 10:14179.

33. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. 2006. Predicting population

448 coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 7:153.

449 34. Menezes Teles EOD, Melo Santos de Serpa Brandão R, Claudio Demes da Mata Sousa L, das
450 Chagas Alves Lima F, Jamil Hadad do Monte S, Sérgio Coelho Marroquim M, Vanildo de Sousa
451 Lima A, Gilberto Borges Coelho A, Matheus Sousa Costa J, Martins Ramos R, Socorro da Silva A.
452 2019. pHLA3D: An online database of predicted three-dimensional structures of HLA molecules.
453 Hum Immunol 80:834-841.

454 35. Ko J, Park H, Heo L, Seok C. 2012. GalaxyWEB server for protein structure prediction and
455 refinement. Nucleic Acids Res 40:W294-7.

456 36. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J,
457 Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu
458 LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new
459 coronavirus of probable bat origin. Nature 579:270-273.

460 37. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier
461 FAT, Ortiz AT, Balloux F. 2020. Emergence of genomic diversity and recurrent mutations in
462 SARS-CoV-2. Infect Genet Evol 83:104351.

463 38. Rausch JW, Capoferri AA, Katusiime MG, Patro SC, Kearney MF. 2020. Low genetic diversity may
464 be an Achilles heel of SARS-CoV-2. Proc Natl Acad Sci U S A 117:24614-24616.

465 39. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. 2011. Coronaviruses: an RNA
466 proofreading machine regulates replication fidelity and diversity. RNA Biol 8:270-9.

467 40. Báez-Santos YM, St John SE, Mesecar AD. 2015. The SARS-coronavirus papain-like protease:
468 structure, function and inhibition by designed antiviral compounds. Antiviral Res 115:21-38.

41. Korn SM, Dhamotharan K, Fürtig B, Hengesbach M, Löhr F, Qureshi NS, Richter C, Saxena K, Schwalbe H, Tants JN, Weigand JE, Wöhnert J, Schlundt A. 2020. (1)H, (13)C, and (15)N backbone chemical shift assignments of the nucleic acid-binding domain of SARS-CoV-2 non-structural protein 3e. Biomol NMR Assign 14:329-333.

42. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J, Zheng L, Zhang Y, Wang H, Zhu Y, Zhu C, Hu T, Hua T, Zhang B, Yang X, Li J, Yang H, Liu Z, Xu W, Guddat LW, Wang Q, Lou Z, Rao Z. 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science 368:779-782.

43. Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET, Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. 2020. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. Cell 182:1560-1573.e13.

44. Robson F, Khan KS, Le TK, Paris C, Demirbag S, Barfuss P, Rocchi P, Ng WL. 2020. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. Mol Cell 79:710-727.

45. Sakai Y, Kawachi K, Terada Y, Omori H, Matsuura Y, Kamitani W. 2017. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. Virology 510:165-174.

46. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, Dejnirattisai W, Rostron T, Supasa P, Liu C, López-Camacho C, Slon-Campos J, Zhao Y, Stuart DI, Paesen GC, Grimes JM, Antson AA, Bayfield OW, Hawkins D, Ker DS, Wang B, Turtle L, Subramaniam K, Thomson P, Zhang P, Dold C, Ratcliff J, Simmonds P, de Silva T, Sopp P, Wellington D, Rajapaksa U, Chen YL, Salio M, Napolitani G, Paes W, Borrow P, Kessler BM, Fry JW, Schwabe NF, Semple MG, Baillie JK, Moore SC, Openshaw PJM,

Ansari MA, Dunachie S, Barnes E, Frater J, Kerr G, Goulder P, et al. 2020. Broad and strong memory CD4(+) and CD8(+) T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nat Immunol 21:1336-1345.

47. Le Bert N, Tan AT, Kunasegaran K, Tham CYL, Hafezi M, Chia A, Chng MHY, Lin M, Tan N, Linster M, Chia WN, Chen MI, Wang LF, Ooi EE, Kalimuddin S, Tambyah PA, Low JG, Tan YJ, Bertoletti A. 2020. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. Nature 584:457-462.

48. Zuo J, Dowell AC, Pearce H, Verma K, Long HM, Begum J, Aiano F, Amin-Chowdhury Z, Hoschler K, Brooks T, Taylor S, Hewson J, Hallis B, Stapley L, Borrow R, Linley E, Ahmad S, Parker B, Horsley A, Amirthalingam G, Brown K, Ramsay ME, Ladhani S, Moss P. 2021. Robust SARS-CoV-2-specific T cell immunity is maintained at 6 months following primary infection. Nat Immunol 22:620-626.

49. Kim JY, Kwon JS, Bae S, Cha HH, Lim JS, Kim MC, Chung JW, Park SY, Lee MJ, Kim BN, Jung J, Kim MJ, Shin EC, Kim SH. 2021. SARS-CoV-2-Specific Antibody and T Cell Response Kinetics According to Symptom Severity. Am J Trop Med Hyg doi:10.4269/ajtmh.20-1594.

## Figure legends

**Figure 1. Mutated and conserved amino acid sequences of SARS-COV-2.** Frequency of mutation and homology plot based on the full-length proteome sequence of SARS-CoV-2. Proteome sequences of 474727 SARS-CoV-2 strains deposited to the GISAID database with the presenting date ranging from January 27th, 2021 to April 27th, 2021 were aligned with hCoV-19/Wuhan/WIV04/2019 (accession number MN996528.1), which shares 100% homology with the reference sequence Wuhan-Hu-1(NC_045512.2). Mutations with occurring frequency greater than $10^{-3}$ were recruited. 899 amino acid substitutions and 20 amino acid deletions were found across the proteome of SARS-COV-2. (**A**) Dots represent the amino acid of mutations. (**B**) Lines represent the amino acid of mutations. The percentage of homology 100% represent the conserved amino acid sequences. S (spike protein), E (envelope protein), M (membrane glycoprotein), N (nucleocapsid phosphoprotein). NSP (nonstructrural protein). ORF (open reading frame).

**Figure 2. The predicted conserved and universal T cell epitopes.** 9-mer peptides (in blue) derived from SARS-CoV-2 (NC_045512.2) presented by selected HLA class I molecules with high binding affinity (IC50$\leqslant$500 nM, percentile rank < 0.5%) and antigenicity predicted by the NetMHCpan 4.1 and Vaxijen-2.0 server. 15-mer Peptides (in yellow) across the proteome of the SARS-COV-2 (NC_045512.2) presented by the

selected HLA class II molecules with high binding affinity (IC50≤500 nM, percentile rank < 1%) and antigenicity predicted by the NetMHCIIPan 4.0 and Vaxijen-2.0 server. (**A**) Squares represent the predicted peptides located at the conserved amino acid sequences.(**B**) The vertical lines represent the peptides loacting at conserved amino acid sequences presented by more alleles of HLA molecules (≥3 alleles). The population coverage was calculated by IEDB population Coverage server. Transverse lines represent the cut-off of population coverage 45% or 90%, respectively. S (spike protein), E (envelope protein), M (membrane glycoprotein), N (nucleocapsid phosphoprotein). NSP (nonstructrural protein). ORF (open reading frame).

**Figure 3. The selected conservative and universal T cell epitopes and the presented HLA molecules.** The selected conservative and universal T cell epitopes with antigenicity and high binding affinity to presented HLA molecules illustrated by dots. (**A**) $CD8^+$ T cell epitopes. (**B**) $CD4^+$ T cell epitopes.

**Table 1 Docking simulations between the selected T cell epitopes and the presented HLA molecules by GalaxyPepDock**

| Epitope | Sequence of epitope | Alleles of HLA molecules | PDB ID of HLA molecules[1] | Protein structure similarity[2] | Interaction similarity score[3] | Estimated accuracy[4] |
|---|---|---|---|---|---|---|
| S691-699 | SIIAYTMSL | HLA-A*02:01 | 3UTQ:A | 0.988 | 241 | 1 |
| | | HLA-A*02:03 | 3OX8:A | 0.989 | 226 | 1 |
| | | HLA-A*02:06 | 3BH9:A | 0.98 | 238 | 1 |
| | | HLA-A*32:01 | 5E00:A | 0.983 | 205 | 1 |
| | | HLA-A*68:02 | 4HX1:A | 0.977 | 231 | 1 |
| | | HLA-B*08:01 | 4QRU:A | 0.994 | 208 | 1 |
| NSP4 420-428 | FLLNKEMYL | HLA-A*02:01 | 3UTQ:A | 0.988 | 268 | 1 |
| | | HLA-A*02:03 | 3OX8:A | 0.989 | 244 | 1 |
| | | HLA-A*02:06 | 3BH9:A | 0.98 | 265 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | HLA-B*08:01 | 4QRU:A | 0.994 | 305 | 1 |
| | | HLA-A*02:01 | 3UTQ:A | 0.986 | 239 | 1 |
| | | HLA-A*02:03 | 3OX8:A | 0.988 | 227 | 1 |
| NSP12 123-131 | TMADLVYAL | HLA-A*02:06 | 3BH9:A | 0.973 | 236 | 1 |
| | | HLA-A*32:01 | 5E00:A | 0.979 | 206 | 1 |
| | | HLA-A*68:02 | 4HX1:A | 0.979 | 221 | 1 |
| | | HLA-A*02:01 | 3UTQ:A | 0.986 | 235 | 1 |
| NSP12 647-655 | SLSHRFYRL | HLA-A*02:03 | 3OX8:A | 0.988 | 230 | 1 |
| | | HLA-B*08:01 | 4QRU:A | 0.994 | 264 | 1 |
| | | HLA-B*15:01 | 1XR9:A | 0.986 | 212 | 1 |
| | | HLA-B*35:01 | 2CIK:A | 0.991 | 226 | 1 |
| NSP13 209-217 | VVYRGTTTY | HLA-A*03:01 | 3RL1:A | 0.989 | 190 | 1 |
| | | HLA-A*11:01 | 1X7Q:A | 0.969 | 181 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | HLA-A*30:02 | 1X7Q:A | 0.973 | 181 | 1 |
| | | HLA-A*32:01 | 5E00:A | 0.989 | 187 | 1 |
| | | HLA-A*03:01 | 3RL1:A | 0.989 | 196 | 1 |
| | | HLA-A*11:01 | 1X7Q:A | 0.99 | 186 | 1 |
| NSP3 950-958 | VMYMGTLSY | HLA-A*30:02 | 1X7Q:A | 0.984 | 196 | 1 |
| | | HLA-A*32:01 | 5E00:A | 0.979 | 199 | 1 |
| | | HLA-B*15:01 | 1XR9:A | 0.986 | 219 | 1 |
| | | HLA-B*35:01 | 2CIK:A | 0.991 | 234 | 1 |
| S 310-324 | KGIYQTSNFRVQPTE | HLA-DPA10103-DPB10401<br>HLA-DPA10201-DPB10101<br>HLA-DPA10301-DPB10402 | 3WEX:A;<br>3WEX:B | 0.95 | 96 | 0.792 |
| NSP12 778-792 | SIKNFKSVLYYQNNV | HLA-DPA10103-DPB10401<br>HLA-DPA10201-DPB10101 | 3WEX:A;<br>3WEX:B | 0.962 | 98 | 0.808 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | HLA-DPA10301-DPB10402 | | | | |
| | | HLA-DPA10103-DPB10201 | | | | |
| | | HLA-DPA10103-DPB10201 | | | | |
| NSP13 177-191 | NRNYVFTGYRVTKNS | HLA-DPA10201-DPB10101<br>HLA-DPA10301-DPB10402 | 3WEX:A;<br>3WEX:B | 0.95 | 106 | 0.814 |
| | | HLA-DPA10103-DPB10401 | | | | |
| | | HLA-DPA10103-DPB10401 | | | | |
| NSP13 413-427 | TKGTLEPEYFNSVCR | HLA-DPA10103-DPB10201<br>HLA-DPA10201-DPB10101 | 3WEX:A;<br>3WEX:B | 0.95 | 106 | 0.814 |
| | | HLA-DPA10301-DPB10402 | | | | |
| | | HLA-DPA10103-DPB10401 | | | | |
| NSP13 538-552 | GSEYDYVIFTQTTET | HLA-DPA10103-DPB10201<br>HLA-DPA10201-DPB10101 | 3WEX:A;<br>3WEX:B | 0.95 | 116 | 0.837 |

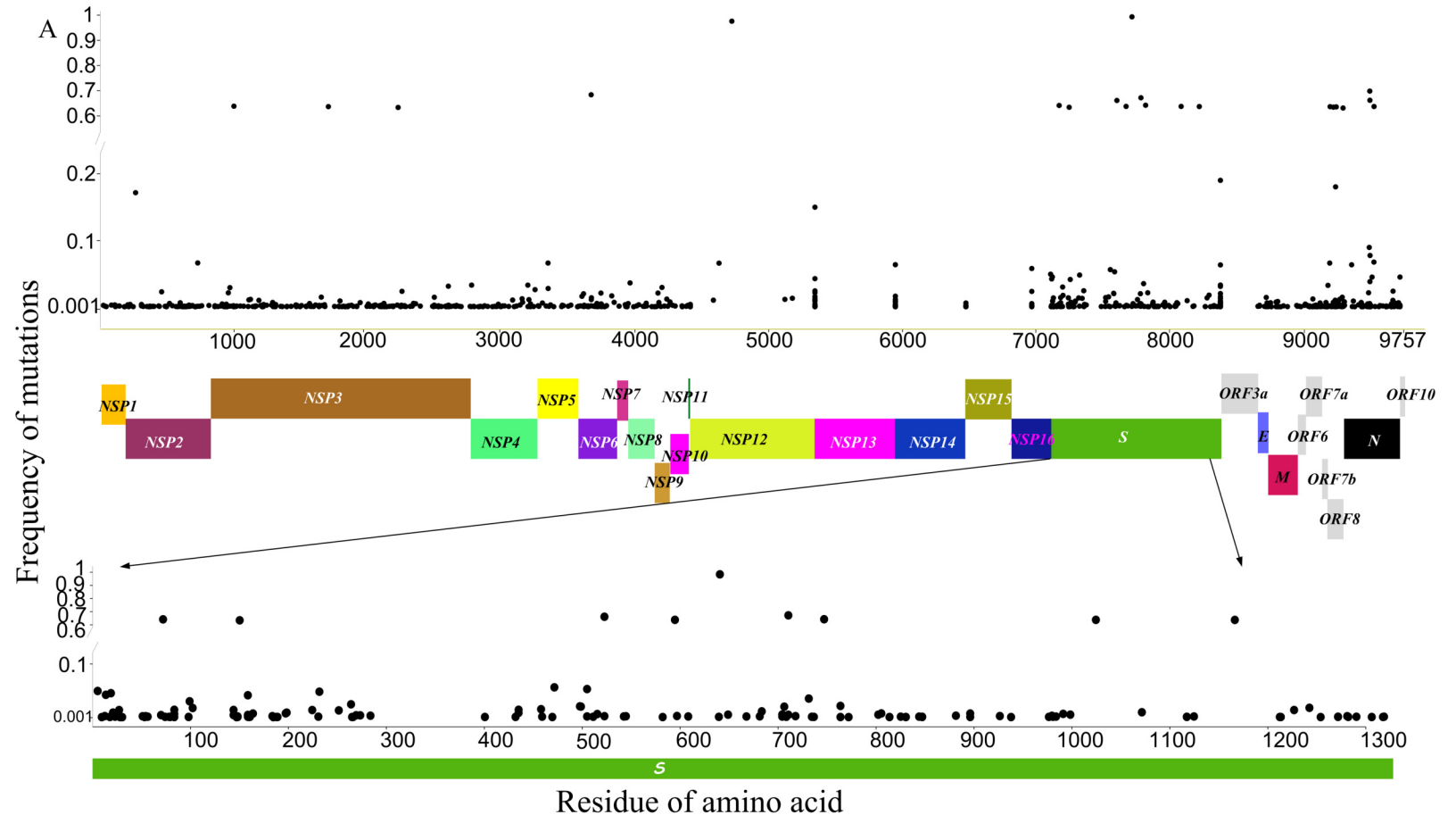
| | | | | | | |
|---|---|---|---|---|---|---|
| | | HLA-DPA10301-DPB10402 | | | | |
| | | HLA-DPA10103-DPB10201 | | | | |
| | | HLA-DPA10201-DPB10101 | | | | |
| NSP14 232-246 | GFDYVYNPFMIDVQQ | HLA-DPA10201-DPB10501 | 3WEX:A; 3WEX:B | 0.963 | 117 | 0.852 |
| | | HLA-DPA10301-DPB10402 | | | | |
| | | HLA-DPA10103-DPB10401 | | | | |
| | | HLA-DPA10103-DPB10401 | | | | |
| NSP3 1134-1148 | SRELKVTFFPDLNGD | HLA-DPA10201-DPB10101 | 3WEX:A; 3WEX:B | 0.943 | 112 | 0.82 |
| | | HLA-DPA10301-DPB10402 | | | | |

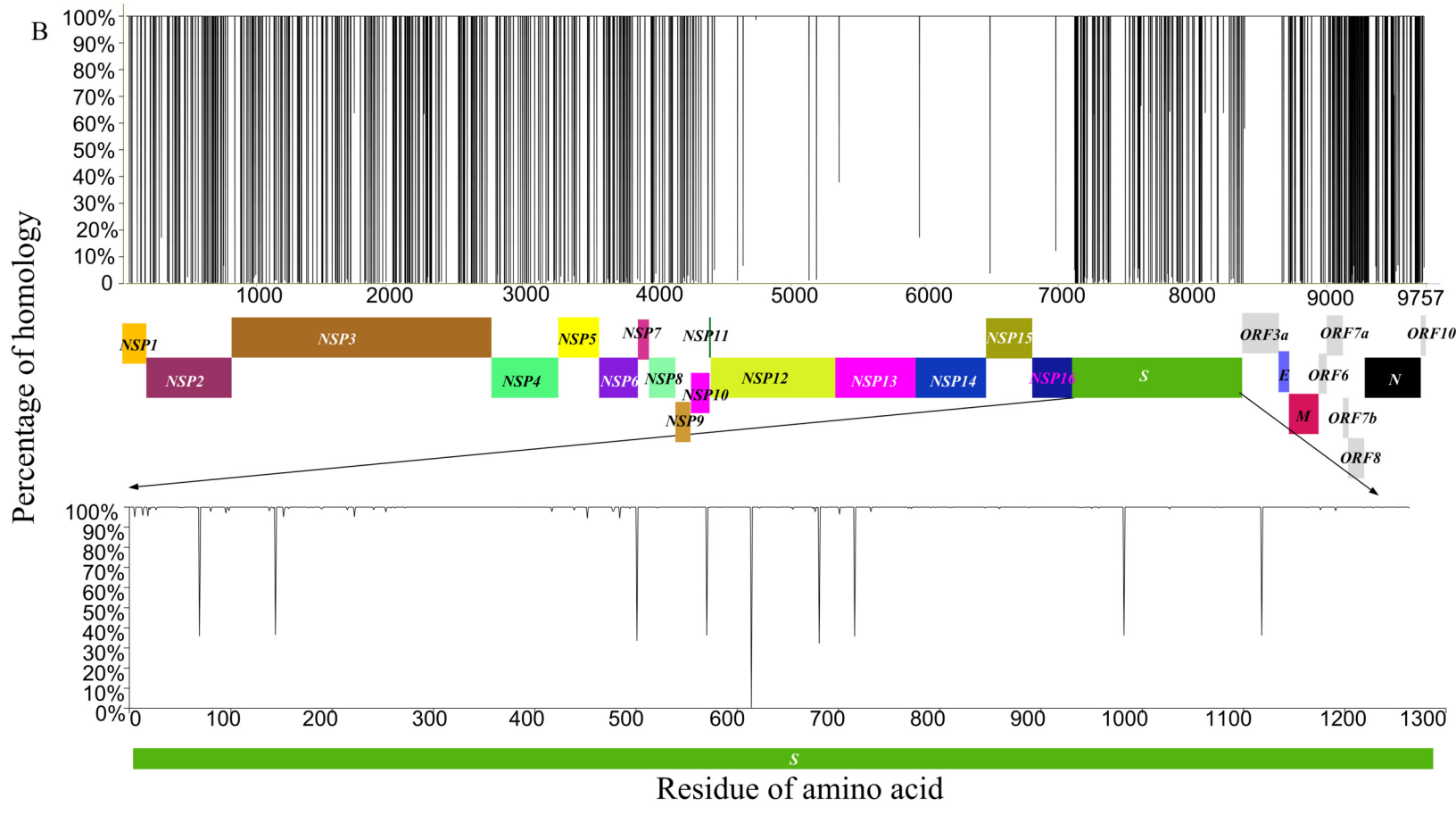
[1] PDB ID of the 3D structure used as template for the modeling derived from pHLA database .

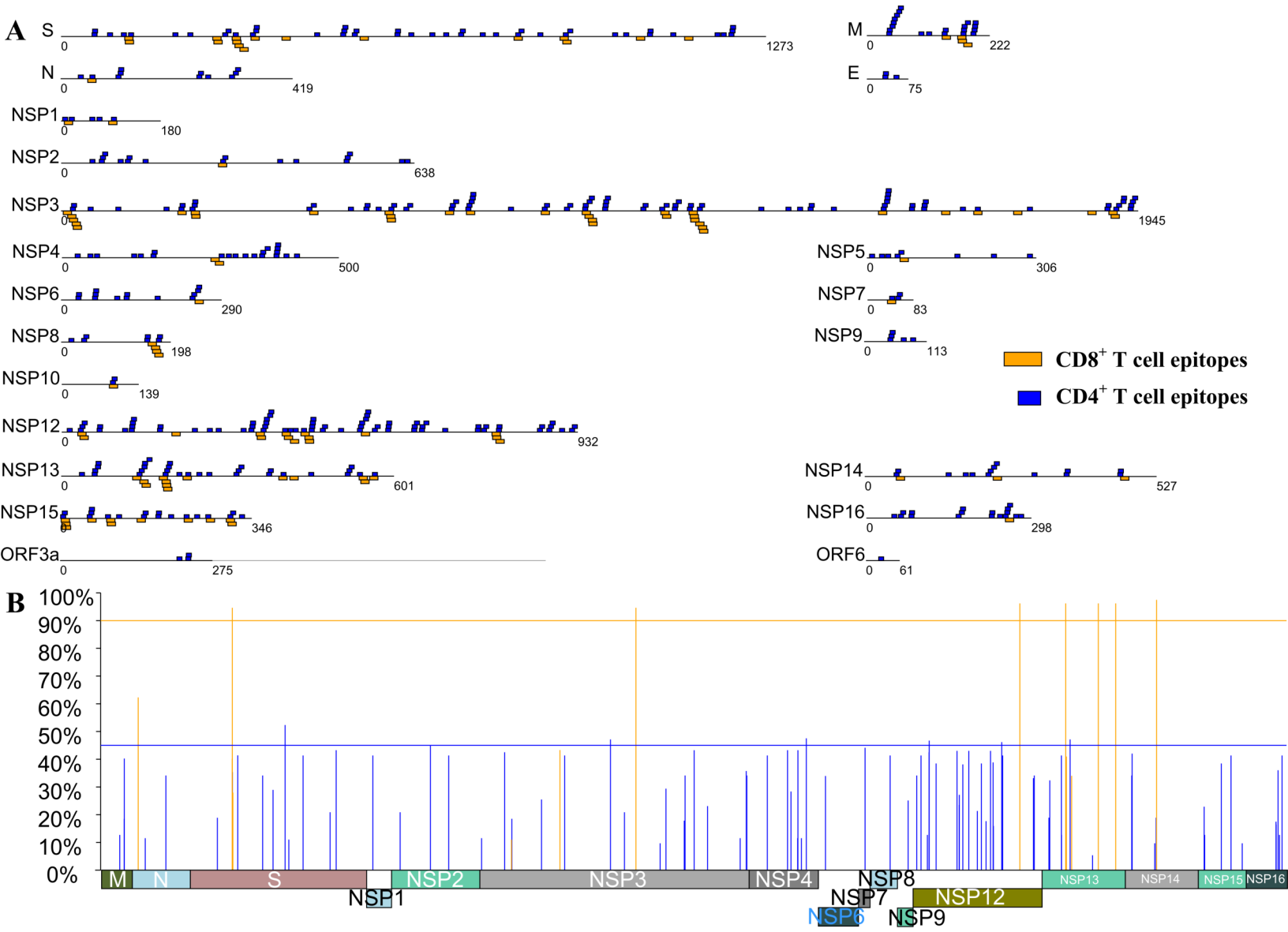
[2]Similarity of the input protein structure to the selected template structure.

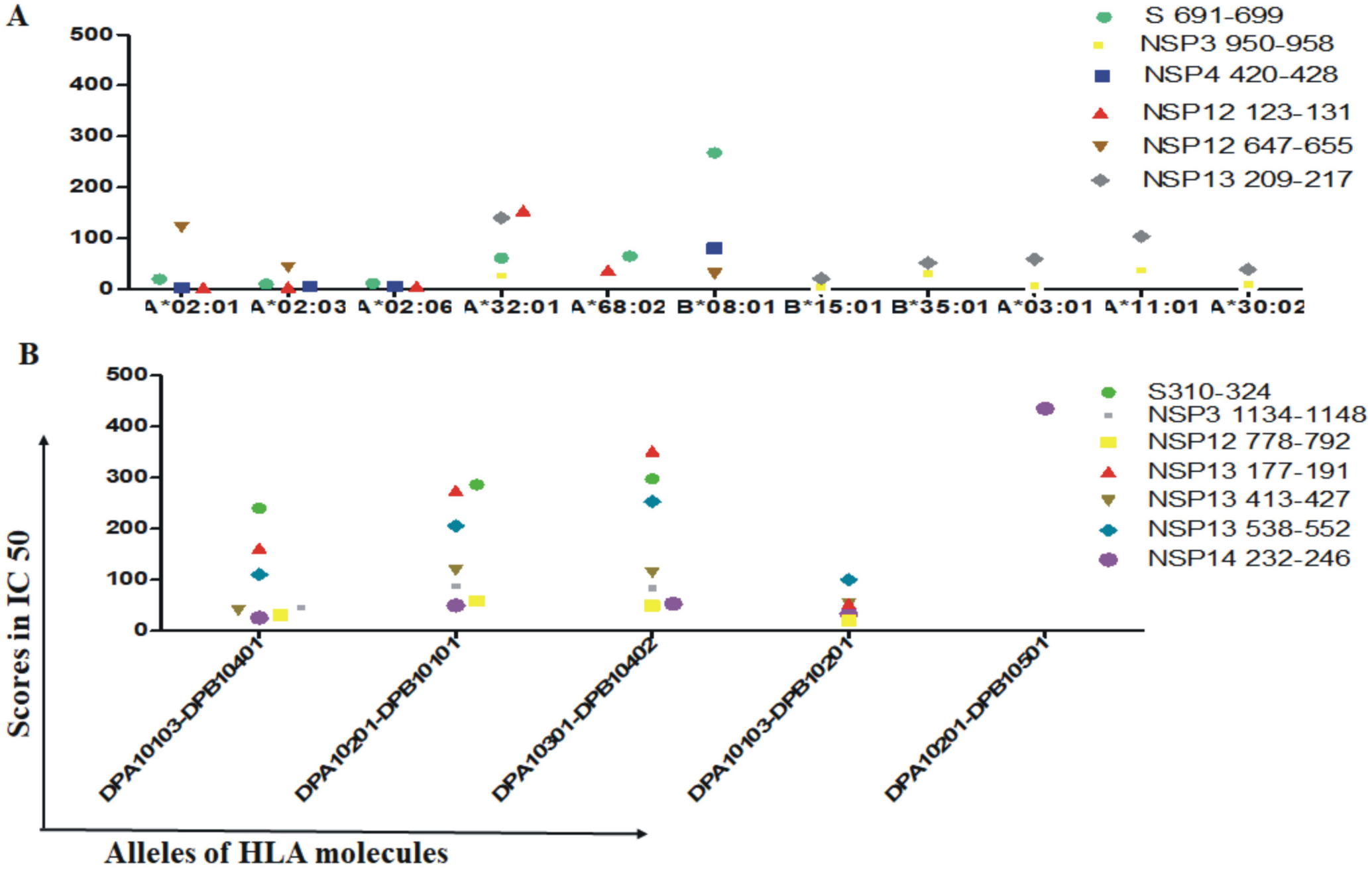
[3]Similarity of the amino acids of the target complex aligned to the contacting residues in the template structure to the template amino acids.

[4]The estimated fraction of correctly predicted binding site residues.






A
500
400
300
200
100
0
S 691-699
NSP3 950-958
NSP4 420-428
NSP12 123-131
NSP12 647-655
NSP13 209-217
A*02:01
A*02:03
A*02:06
A*32:01
A*68:02
B*08:01
B*15:01
B*35:01
A*03:01
A*11:01
A*30:02
B
500
400
300
200
100
0
S310-324
NSP3 1134-1148
NSP12 778-792
NSP13 177-191
NSP13 413-427
NSP13 538-552
NSP14 232-246
Scores in IC 50
DPA10103-DPB10401
DPA10201-DPB10101
DPA10301-DPB10402
DPA10103-DPB10201
DPA10201-DPB10501
Alleles of HLA molecules