


Immunoinformatic approach to assess SARS-CoV-2 protein S epitopes recognised by the most frequent MHC-I alleles in the Brazilian population

Ronald Rodrigues de Moura ,¹ Almerinda Agreli,² Carlos André Santos-Silva,³ Natália Silva,² Bruno Rodrigo Assunção,² Lucas Brandão,² Ana Maria Benko-Iseppon,³ Sergio Crovella¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jclinpath-2020-206946>).

¹Department of Advanced Diagnostics, IRCCS Materno Infantile Burlo Garofolo, Trieste, Friuli Venezia Giulia, Italy
²Department of Pathology, Federal University of Pernambuco, Recife, Brazil
³Department of Genetics, Federal University of Pernambuco, Recife, Brazil

Correspondence to

Dr Ronald Rodrigues de Moura, Department of Advanced Diagnostics, IRCCS Materno Infantile Burlo Garofolo, 34137 Trieste, Friuli Venezia Giulia, Italy; ronaldmoura1989@gmail.com

Received 14 July 2020
Revised 21 July 2020
Accepted 23 July 2020



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Moura RR, Agreli A, Santos-Silva CA, *et al.* *J Clin Pathol* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jclinpath-2020-206946

ABSTRACT

Aims Brazil is nowadays one of the epicentres of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic and new therapies are needed to face it. In the context of specific immune response against the virus, a correlation between Major Histocompatibility Complex Class I (MHC-I) and the severity of the disease in patients with COVID-19 has been suggested. Aiming at better understanding the biology of the infection and the immune response against the virus in the Brazilian population, we analysed SARS-CoV-2 protein S peptides in order to identify epitopes able to elicit an immune response mediated by the most frequent MHC-I alleles using *in silico* methods.

Methods Our analyses consisted in searching for the most frequent Human Leukocyte Antigen (HLA)-A, HLA-B and HLA-C alleles in the Brazilian population, excluding the genetic isolates; then, we performed: molecular modelling for unsolved structures, MHC-I binding affinity and antigenicity prediction, peptide docking and molecular dynamics of the best fitted MHC-I/protein S complexes.

Results We identified 24 immunogenic epitopes in the SARS-CoV-2 protein S that could interact with 17 different MHC-I alleles (namely, HLA-A*01:01; HLA-A*02:01; HLA-A*11:01; HLA-A*24:02; HLA-A*68:01; HLA-A*23:01; HLA-A*26:01; HLA-A*30:02; HLA-A*31:01; HLA-B*07:02; HLA-B*51:01; HLA-B*35:01; HLA-B*44:02; HLA-B*35:03; HLA-C*05:01; HLA-C*07:01 and HLA-C*15:02) in the Brazilian population.

Conclusions Being aware of the intrinsic limitations of *in silico* analysis (mainly the differences between the real and the Protein Data Bank (PDB) structure; and accuracy of the methods for simulate proteasome cleavage), we identified 24 epitopes able to interact with 17 MHC-I more frequent alleles in the Brazilian population that could be useful for the development of strategic methods for vaccines against SARS-CoV-2.

INTRODUCTION

In December 2019, a novel coronavirus strain was detected and isolated in the city of Wuhan, Hubei province, China.¹ This emerging viral infection was associated with severe human respiratory disease with a fatality rate of ~2%–3%.² Despite the strict containment measures, severe acute respiratory syndrome coronavirus 2

(SARS-CoV-2) has spread rapidly worldwide, with cases now confirmed in multiple countries and propagation still ongoing. On 30 January 2020, the WHO declared COVID-19 a public health emergency of international concern.³

Brazil is actually one of the epicentres of the pandemic.⁴ Day by day, the epidemic advances with a high daily rate of cases per million people. Nonetheless, these data may be underestimated because of the reduced number of SARS-CoV-2 molecular detection tests being made.⁵

The SARS-CoV-2 is an enveloped positive-strand RNA virus belonging to the *Betacoronavirus* genus of the *Coronaviridae* family of the *Nidovirales* order.⁶ SARS-CoV-2 genome has ~30 kilobases that encode structural and non-structural proteins.⁷ The 5' encodes two large genes (ORF1a and ORF1b), which are translated into two polyproteins (pp1a and pp1ab), which are cleaved in a set of non-structural proteins essential for virus replication.⁸ The 3'-terminal region of the genome encodes four structural proteins, namely spike (S), nucleocapsid (N) envelope (E) and membrane (M).⁹

The protein S on the virion surface mediates receptor binding and membrane fusion.⁶ For SARS-CoV-2 to infect a host cell, it is mandatory that protein S is cleaved (S1/S2 cleavage) by host cell proteases into two units: the N-terminal ectodomain (S1) subunit and the C-terminal membrane-anchored (S2) subunit.¹⁰ During the receptor binding process, S1 binds via its receptor binding domain to the extracellular peptidase domain of the host receptor ACE-2, which is mainly expressed in the lung, gastrointestinal tract, kidney and heart tissues, although it is also present in other tissues, dictating viral tropism.^{10 11} After that, S2 subunit mediates membrane fusion. Although cell entry is not yet fully understood, it is likely that a second proteolytic site (S2') at subunit S2 is required for viral entry.¹² Viral entry triggers the host's immune system and initiates an inflammatory cascade that starts with the mechanism of antigen presentation.⁶

During the process of presentation of antigenic peptides to CD8⁺ (cytotoxic) T cells by Class I Major Histocompatibility Complex (MHC-I) molecules, peptides are generated by proteasomal cleavage in the cytosol and transported to the lumen of the endoplasmic reticulum by

Table 1 Average allele frequency and its respective Protein Data Bank (PDB) structure identifiers (when available) of the top 10 most frequent Class I Major Histocompatibility Complex HLA-A, HLA-B and HLA-C alleles in the general Brazilian population

HLA-A			HLA-B			HLA-C		
Allele	Frequency	PDB	Allele	Frequency	PDB	Allele	Frequency	PDB
A*02:01	0.2145	4L29	B*07:02	0.0870	5E00	C*07:01	0.156	–
A*24:02	0.1010	2BCK	B*51:01	0.0840	5VUE	C*04:01	0.1370	1QQD
A*01:01	0.0850	4NQX	B*08:01	0.0585	3×13	C*07:02	0.0800	5VGE
A*03:01	0.0755	609B	B*18:01	0.0570	6MT3	C*08:02	0.0750	–
A*23:01	0.0457	–	B*35:03	0.0470	–	C*05:01	0.0610	–
A*11:01	0.0435	1X7Q	B*35:01	0.0465	1XH3	C*03:03	0.0470	–
A*31:01	0.0330	–	B*44:03	0.0445	4JQX	C*12:03	0.0380	–
A*26:01	0.0310	–	B*14:02	0.0375	3BVN	C*15:02	0.0330	–
A*30:02	0.0285	–	B*44:02	0.0355	1M60	C*17:01	0.0330	–
A*68:01	0.0285	6PBH	B*38:01	0.0320	–	C*03:04	0.0280	1EFX
Total	0.6862			0.5295			0.6880	

HLA, Human Leukocyte Antigen.

a transporter associated with antigen processing protein before they can bind MHC groove and trigger an immune response.¹³ The correlation between MHC-I and the severity of the disease in patients with COVID-19 has been previously hypothesised.¹⁴

The understanding of which SARS-CoV-2 epitopes are immunogenic may help advances in the development of diagnostic kits and prophylactic vaccines. An immunoinformatic approach is suitable for an initial screening, since it may predict algorithms and test for the immunogenicity of a vast quantity of peptides and alleles in a cost-effective manner, reducing the costs and time on workbench. Here, we analysed SARS-CoV-2 protein S peptides aiming to prospect epitopes and their ability to elicit an immune response mediated by the most frequent MHC-I alleles in the Brazilian population.

METHODS

Selection of the MHC-I alleles

We searched for the most frequent MHC-I alleles in the Brazilian population through the Human Leukocyte Antigen (HLA) Allele Frequency Database.¹⁵ We selected allele frequency data from overall Brazilian populations, excluding genetic isolates, such as indigenous tribes or *Quilombos*. After that, we calculated the average allelic frequency for each HLA-A, HLA-B and HLA-C allele with data from more than one population data. Finally, we chose the 10 most frequent class I alleles.

Molecular modelling

The Protein Data Bank (PDB) structure of some alleles was not available (table 1). Therefore, we used SWISS-MODEL¹⁶ to perform modelling by homology. The best model was chosen using the default parameters. Model refinement was made using 3Drefine software.¹⁷ The quality of the refinement was evaluated by the MolProbity score. For model validation, ERRAT¹⁸ and RAMPAGE¹⁹ softwares were used with their default parameters.

MHC binding affinity and antigenicity prediction

For immunogenicity inference, we used the NetMHCpan V.4.0²⁰ server, by importing the FASTA sequence of the SARS-CoV-2 spike protein retrieved from PDB (PDB id: 6VSB) that will be subsequently cleaved in peptides with a length of 10-mer. The interrogation of each 10-mer peptide

was made querying the selected alleles ranked in the previous section. We accepted as possible candidates, those allele:peptide complexes with strong binding affinity, that is, those with %Rank score <0.50.

After that, we tested the strong binding peptides for probable antigenicity using Vaxijen V.2.0,²¹ using a threshold of score <0.5 to filter out possible non-antigenic peptides.

Peptide docking

Peptide docking was carried out using the CABS-dock server.²² We submitted to a docking simulation, those peptides that showed strong binding affinity and probable antigenicity to their specific MHC-I allele using the default parameters. After the docking, we excluded the complexes that showed root mean squared deviation (RMSD) values above 3 Å to follow the molecular dynamics (MD).²³

Molecular dynamics

The complexes from the docking simulations (RMSD values below 3 Å) were further investigated by MD simulations using GROMACS 2020.1.²⁴ The complex was inserted in a 0.7 nm cubic box, filling the voided space with SPC-E water molecules. Calcium and sodium ions were added for reaching the system electronic equilibrium. Energy minimisation was carried out using default parameters. The equilibrium phase was made during 10 ns (for temperature and pressure), while the production phase lasted 100 ns. These steps were also performed using default parameters.

In order to investigate the stability of the complexes, we calculated RMSD, root mean squared fluctuation, radius of gyration (Rg), solvent accessible surface area and the determination of hydrogen bonds.

RESULTS AND DISCUSSION

Table 1 shows the most frequent HLA-A, HLA-B and HLA-C alleles, according to the HLA Allele Frequency Database. The top 10 HLA-A alleles correspond to 68.7% of the A* alleles, whereas the top 10 HLA-B alleles represent 53% and the HLA-C alleles 68.8%. These values may not be sufficient to depict the MHC-I allelic variation, but it can give an estimation of the major diversity of the class I genes.

We modelled the MHC-I alleles, whose PDB structures were not available. The templates used for model prediction, refinement and validations are reported in online supplementary table 1.

Table 2 Best protein S epitope is bound to its respective Major Histocompatibility Complex Class I alleles according to molecular dynamics results

Allele	PDB	Peptide	Code	%Rank	OPPA prediction	Average RMSD (Å)
HLA-A*01:01	4NQX	NLDSKVGNGNY	pep2	0.1041	0.7882	2.5624
HLA-A*01:01	4NQX	GWTAGAAAYY	pep3	0.2305	0.5358	2.6896
HLA-A*02:01	4L29	KLNDLCFTNV	pep6	0.2287	2.6927	1.1479
HLA-A*02:01	4L29	FELLHAPATV	pep7	0.3513	0.5982	1.1796
HLA-A*11:01	1X7Q	VTLADAGFIK	pep15	0.1366	0.8702	2.1979
HLA-A*11:01	1X7Q	SLIDLQELGK	pep18	0.2777	1.0275	1.9232
HLA-A*24:02	2BCK	TYVPAQEKNF	pep24	0.0424	0.7276	2.0925
HLA-A*68:01	6PBH	HVTYVPAQEK	pep65	0.2517	1.0786	1.1007
HLA-A*23:01	swissmodel	CYFPLQSYGF	pep83	0.3803	0.7776	2.8308
HLA-A*26:01	swissmodel	EILDITPCSF	pep85	0.1991	1.2698	1.7413
HLA-A*26:01	swissmodel	GWTAGAAAYY	pep87	0.2372	0.5358	1.9260
HLA-A*30:02	swissmodel	IGAEHVNNNSY	pep92	0.4353	1.2671	1.5989
HLA-A*31:01	swissmodel	RKSNLKPFR	pep96	0.4348	0.8187	2.4331
HLA-B*07:02	5E00	TPCSFGGVSV	pep59	0.3909	1.7080	2.7139
HLA-B*51:01	5VUE	VAYSNNNSIAI	pep70	0.4565	1.0545	1.2241
HLA-B*35:01	1XH3	LPIGINITRF	pep72	0.0959	1.3027	2.2773
HLA-B*35:01	1XH3	QIPFAMQMAY	pep73	0.2719	1.2149	1.6030
HLA-B*35:01	1XH3	FPNITNLCPF	pep75	0.4307	1.3964	2.4222
HLA-B*44:02	1M60	TESNKKFLPF	pep79	0.4851	1.0405	1.4687
HLA-B*35:03	swissmodel	LPIGINITRF	pep97	0.1519	1.3027	2.0459
HLA-B*35:03	swissmodel	LPFFSNVTWF	pep98	0.4127	0.6567	0.4809
HLA-C*05:01	swissmodel	LVDLPIGINI	pep103	0.1770	1.3499	0.2044
HLA-C*07:01	swissmodel	IVRFPNITNL	pep105	0.1311	0.5548	4.0092
HLA-C*15:02	swissmodel	AVRDPQTLEI	pep109	0.4849	0.7083	2.5140

It is possible to see the Protein Data Bank (PDB) ID for the structures used (or the indication that the structure was modelled using Swiss-model); the amino acid sequence; the code that we used during the simulations; %Rank value from NetMHC Pan; Vaxijen's OPPA score and average root mean squared deviation (RMSD) for the docking simulations according to CABS dock. HLA, Human Leukocyte Antigen.

Overall, the models were considered suitable for docking analysis since they presented significant amino acid sequence similarity to the templates and were above RAMPAGE and ERRAT thresholds (98% and 80%, respectively).

Two hundred twenty-eight peptides were considered with strong binding affinity, according to NetMHCpan V.4.0 results (online supplementary table 2); of these, 73 were also considered probably antigenic based on the Vaxijen V.2.0 results. We carried out peptide docking simulations for all of them except for two (pep88 and pep111), which could not be docked by CABS-dock server, possibly due to some issues with the MHC-I structure. As result, 61 peptides had good docking predictions, as they showed average RMSD <3 Å (online supplementary table 3).

These complexes comprised 24 different MHC-I alleles: HLA-A*01:01 (four epitopes), HLA-A*02:01 (2), HLA-A*03:01 (3), HLA-A*11:01 (5), HLA-A*23:01 (4), HLA-A*24:02 (3), HLA-A*26:01 (3), HLA-A*30:02, HLA-A*31:01 (3), HLA-A*68:01 (5), HLA-B*07:02 (3), HLA-B*18:01 (1), HLA-B*35:01 (4), HLA-B*35:03 (2), HLA-B*38:01 (2), HLA-B*44:02 (2), HLA-B*44:03 (1), HLA-B*51:01 (1), HLA-C*05:01 (2), HLA-C*07:01 (2), HLA-C*07:02 (1), HLA-C*08:02 (1), HLA-C*15:02 (1) and HLA-C*17:01 (1). To evaluate the stability of these complexes, we completed MD simulations to make trajectory analyses.

Twenty-four MHC-I and protein S peptides (pep) complexes presented good results from MD simulations; these are summarised in table 2. Trajectory analysis of HLA-A alleles showed that the most compact and rigid complexes were: HLA-A*01:01 with the peptides pep2 and pep3 (online supplementary figure 1); HLA-A*02:01 with pep6 and pep7 (online supplementary figure 2); HLA-A*11:01 combined with pep15 and pep18 (online supplementary figure 4); HLA-A*24:02 with pep24 (online supplementary figure 6); HLA-A*68:01 with pep65 (online supplementary

figure 10); HLA-A*23:01 in complex with pep83 (online supplementary figure 5); HLA-A*26:01 with pep85 and pep87 (online supplementary figure 7); HLA-A*30:02 with pep92 (online supplementary figure 8); and HLA-A*31:01 in complex with pep96 (online supplementary figure 9). It is worth noting that even among the most stable complexes, the solvent accessibility varied.

This variation was due to the peptides binding in different positions to the HLA-A alleles. The other peptides (less compact and more flexible) showed more fluctuations in the loop conformations of the complex, including in the peptide structure by itself. In fact, a comparison among the structures confirmed that the major difference among the complex is in the peptides and not in the regions of HLA-A without secondary conformation (online supplementary figures 25 and 26). However, although these differences may affect the stability of the complexes, they do not induce structural modifications sufficient to unfold the protein.

Regarding HLA-B alleles, the complexes with minimal changes from the initial form during the simulations were: HLA-B*07:02 with pep59 (online supplementary figure 11); HLA-B*51:01 in complex with pep70 (online supplementary figure 18); HLA-B*35:01 with pep72, pep73 and pep75 (online supplementary figure 13); HLA-B*38:01 with pep79 (online supplementary figure 15); and HLA-B*35:03 in complex with the peptides pep97 and pep98 (online supplementary figure 14). Similar to the most compact and stable complexes of HLA-A, the major differences among the most rigid and flexible complexes were not in the peptides itself but in regions of HLA-B alleles without secondary conformations (online supplementary figures 27 and 28).

Analysing the trajectories for HLA-C, it was possible to observe that the most stable complexes were HLA-C*05:01, with pep103

(online supplementary figure 19); HLA-C*07:01, with pep105 (online supplementary figure 20) and HLA-C*15:02, with pep109 (online supplementary figure 23). The latter complex, though, appeared to be stable only by the end of the simulation.

There are a few works in the literature using a similar approach to evaluate possible immunogenic peptides in the SARS-CoV-2 protein S. One study identified five CD8⁺ T cells epitopes (YLQPRTEFL, GYVFASTK, EPVLKGVKL, VVNQNAQAL and WTAGAAAYY) and eight B cell epitopes that bind the MHC class I and II alleles more frequent in China.²⁵ A second study found 13 MHC-I (SQCVNLTR, GVYHKNNK, GKQGNFKNL, GIYQTSNFR, VSPTKLNLDL, KIADYNYKL, KVGGNYNLY, EGFNCYFPL, GPKKSTNLV, SPRRARSVA, LGAENSVAY, FKNHTSPDV and DEDDSEPV) and three MHC-II possible protein S antigenic peptides.²⁶ Joshi *et al*²⁷ reported the MHC-I IITLCFTLKR as a possible candidate for vaccine development. Comparing these previous results with our simulations, only GWTAGAAAYY complexed with HLA-A*01:01 and HLA-A*26:01 presented good NetMHC-Pan and Vaxijen scores, as well as good docking stability and RMSD during the MD.

In conclusion, being aware of the intrinsic limitations of *in silico* analysis (differences between the real and the PDB structure, accuracy of the methods for simulate proteasome cleavage, as well as molecular modelling, docking and dynamics' shortcomings), we described 24 epitopes present in the SARS-CoV-2 protein S that could interact with 17 different MHC-I alleles in the Brazilian population. These epitopes can elicit an effective CD8⁺ T cells immune response and could be useful to develop strategic methods for vaccines against COVID-19. Finally, our immunoinformatic approach could be a useful tool to determine a guided starting point to design and develop epitope-based vaccines.

Take home messages

- ▶ The understanding of which severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epitopes are immunogenic may help in the development of diagnostic kits and prophylactic vaccines.
- ▶ An immunoinformatic approach is suitable for Major Histocompatibility Complex Class I (MHC-I) screening, to predict the immunogenicity of a vast quantity of peptides and alleles in a cost-effective manner.
- ▶ SARS-CoV-2 protein S peptides have been analysed *in silico* aiming to prospect epitopes and their ability to elicit an immune response mediated by the most frequent MHC-I alleles in the Brazilian population.
- ▶ Twenty-four epitopes present in the SARS-CoV-2 protein S have been observed, which could interact with 17 different MHC-I alleles identified in the Brazilian population.

Handling editor Runjan Chetty.

Acknowledgements CAS-S and AMB-I thank CAPES (Coordination for the Improvement of Higher Education Personnel, Brazil) and CNPq (Brazilian National Council for Scientific and Technological Development) for financial support and scholarships.

Contributors RRdM, CAS-S, BRA and NS made substantial contributions to the conception or design of the work, or the acquisition, analysis or interpretation of data. RRdM, AA, SC, CAS-S, LB and AMB-I drafted the work or revising it critically for important intellectual content. SC and LB made the final approval of the version published. The authors do agree that they have the accountability for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This study has been supported by the following grants: Italian-Slovenian Ecosystem for Electronic and Mobile Health from European Community (grant number: 07/2019) and BioHub—A High-throughput Platform For OMICs Data Analysis And Integration from the Italian Ministry of Health (grant number: RC03/20). LB is recipient of a senior fellowship from the Brazilian National Council for Scientific and Technological Development (CNPq) (grant number: 308540/2017). RRdM is recipient of a senior fellowship from RC03/20 project from IRCCS Burlo Garofolo, Trieste, Italy.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. Raw data generated in this study may be requested by sending an email to lp.m.ccm@ufpe.br.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iD

Ronald Rodrigues de Moura <http://orcid.org/0000-0002-9112-7167>

REFERENCES

- 1 Li Q, Guan X, Wu P, *et al*. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020;382:1199–207.
- 2 Baud D, Qi X, Nielsen-Saines K, *et al*. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis* 2020;20:30195.
- 3 WHO. COVID 19 public health emergency of international concern (PHEIC), 2020. Available: [https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-\(pheic\)-global-research-and-innovation-forum](https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-(pheic)-global-research-and-innovation-forum)
- 4 WHO. Coronavirus disease (COVID-2019) situation reports, 2020. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200707-covid-19-sitrep-169.pdf?sfvrsn=c6c69c88_2
- 5 Data OWI. Coronavirus pandemic (COVID-19) – the data, 2020. Available: <https://ourworldindata.org/coronavirus-data?country=BRA>
- 6 Rabi FA, Al Zoubi MS, Kasasbeh GA, *et al*. Sars-cov-2 and coronavirus disease 2019: what we know so far. *Pathogens* 2020;9:231–15.
- 7 Petrosillo N, Viceconte G, Ergonul O, *et al*. COVID-19, SARS and MERS: are they closely related? *Clin Microbiol Infect* 2020;26:729–34.
- 8 Malik YS, Sircar S, Bhat S, *et al*. Emerging novel coronavirus (2019-nCoV)-current scenario, evolutionary perspective based on genome analysis and recent developments. *Vet Q* 2020;40:68–76.
- 9 Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 2020;12:254.
- 10 Yan R, Zhang Y, Li Y, *et al*. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;367:1444–8.
- 11 Ksiazek TG, Erdman D, Goldsmith CS, *et al*. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 2003;348:1953–66.
- 12 Coutard B, Valle C, de Lamballerie X, *et al*. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* 2020;176:104742.
- 13 Ren Y, Wu B, Pan Y, *et al*. Characterization of the binding profile of peptide to transporter associated with antigen processing (TAP) using Gaussian process regression. *Comput Biol Med* 2011;41:865–70.
- 14 Wilson E, Hirneise G, Singharoy A, *et al*. Total predicted MHC-I epitope load is inversely associated with mortality from SARS-CoV-2. *medRxiv* 2020.
- 15 Gonzalez-Galarza FF, McCabe A, Santos EJMdos, *et al*. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 2019;61:D783–8.
- 16 Waterhouse A, Bertoni M, Bienert S, *et al*. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018;46:W296–303.
- 17 Bhattacharya D, Nowotny J, Cao R, *et al*. 3Drefine: an interactive web server for efficient protein structure refinement. *Nucleic Acids Res* 2016;44:W406–9.
- 18 Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 1993;2:1511–9.
- 19 Lovell SC, Davis IW, Arendall WB, *et al*. Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* 2003;50:437–50.
- 20 Jurtz V, Paul S, Andreatta M, *et al*. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 2017;199:3360–8.
- 21 Doytchinova IA, Flower DR. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* 2007;8:4.

- 22 Kurcinski M, Jamroz M, Blaszczyk M, *et al.* CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res* 2015;43:W419–24.
- 23 Chew CH, Lim YAL, Chua KH. Heterologous expression of *Plasmodium vivax* apical membrane antigen 1 (PvAMA1) for binding peptide selection. *PeerJ* 2017;5:e3794.
- 24 Abraham MJ, Murtola T, Schulz R, *et al.* GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1-2:19–25.
- 25 Baruah V, Bose S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J Med Virol* 2020;92:495–500.
- 26 Bhattacharya M, Sharma AR, Patra P, *et al.* Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *J Med Virol* 2020;92:618–31.
- 27 Joshi A, Joshi BC, Mannan MA-U, *et al.* Epitope based vaccine prediction for SARS-COV-2 by deploying immuno-informatics approach. *Inform Med Unlocked* 2020;19:100338.