Immunoinformatic based identification of cytotoxic T lymphocyte epitopes from the Indian isolate of SARS-CoV-2

Viswajit Mulpuru
Indian Institute of Information Technology Allahabad, Department of Applied Science, Prayagraj, 211012, India

Nidhi Mishra (nidhimishra@iiita.ac.in)
Indian Institute of Information Technology Allahabad, Department of Applied Science, Prayagraj, 211012, India

Research Article

Keywords: SARS-CoV-2, potential cytotoxic T lymphocyte (CTL) epitopes, Indian isolate, molecular docking

Posted Date: April 27th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-24566/v1

Version of Record: A version of this preprint was published on February 25th, 2021. See the published version at https://doi.org/10.1038/s41598-021-83949-9.

Abstract

The Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has turned into a pandemic with about a million confirmed cases worldwide. Being an airborne infection, it can be highly fatal to populous countries like India. This study sets to identify potential cytotoxic T lymphocyte (CTL) epitopes in the SARS-CoV-2 Indian isolate which can acts act as an effective vaccine candidate for the majority of the Indian population. The immunogenicity and the foreignness of the epitopes towards the human body have to studies to further confirm their candidacy. The top-scoring epitopes were subjected to molecular docking studies to study their interactions with the corresponding human leukocyte antigen (HLA) system. The CTL epitopes were observed to bind at the peptide-binding groove of the corresponding HLA system, indicating their potency as a vaccine candidate. The identified epitopes can be subjected to further studies for the development of SARS-CoV-2 vaccine.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a virus of genus Betacoronavirus and family Coronaviridae with ranging from a runny nose to severe acute respiratory syndrome and kidney failure [1]. Now that it has infected more than 100 countries worldwide, the World Health Organization has declared it as a pandemic. An airborne infection like SARS-CoV-2 in populous countries such as India can escalate very rapidly. Thus, finding a vaccine candidate for this deadly disease is of first priority to the researchers worldwide.

As it is known that cytotoxic T lymphocytes (CTL) with the help of specialised proteins known as T cell receptors (TCR) on their surface help fight virally-infected cells, using this natural mechanism of the immune system can be of great help in fighting this dreadful virus. Each TCR that is present on the surface of the TCL can specifically identify antigenic peptides bound to the major histocompatibility complex (MHC). whenever the TCR detects a viral antigen, the CTL releases cytotoxins to kill the virus-infected cell thus helping in an immune response.[2]

This study aimed to identify cytotoxic T cell (CTL) epitopes of SARS-CoV-2 Indian isolate for designing potential vaccine candidates which are effective on Indian population using an in-silico approach. We predicted the CTL epitopes for all those human leukocyte antigen supertypes (HLA) which have high allelic frequently in Indian population. We further studied the immunogenicity, foreignness, and interactions between the epitopes and the HLA molecules by the means of molecular docking studies.

Materials And Methods

Retrieving the complete genome and the coding sequence from an Indian isolate

The amino acid sequence of the complete genome of Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolate of India (MT050493.1) with 9950 AA was retrieved from NCBI. All the coding regions of the genome was retrieved from the features section containing the coding sequences of orf1ab

polyprotein, surface glycoprotein, orf3a protein, envelope protein, membrane glycoprotein, orf6 protein, orf7a protein, orf8 protein, nucleocapsid phosphoprotein andorf10 protein.

Prediction of cytotoxic T cell epitopes for the Indian population

NetCTLpan version 1.1 [3] was used to predict the CTL epitopes across the proteins coded by the SARS-CoV-2 Indian isolate. NetCTLpan uses a neural network to predict TAP-transporter binding and C terminal cleavage predictions in addition to HLA binding prediction. Considering the HLA supertype variation across populations, we predicted the epitopes only for those HLA supertypes which constitute for the majority of human leukocyte antigen (HLA) distribution in the Indian population. The study on the evolution of HLA-A and HLA-B polymorphisms reveals that HLA A3, B7 and B44 are the major HLA's present in Indian population [4].

Prediction of epitope immunogenicity

Although the binding affinities of the peptides towards HLA help in predicting the epitopes, the immunogenicity plays an important role in the immune response. All the predicted epitopes were subjected to the Immune Epitope Database (IEDB) immunogenicity tool [5,6] to predict their immunogenicity score. IEDB immunogenicity tool relays on physicochemical properties such as side chains composition, amino acid position to predict the immunogenicity of the peptide sequence.

Identification of unique epitopes

As the human body shows immune response only towards foreign antigens, it is of great importance to consider only those epitopes which are foreign to the human body as a potential vaccine candidate. To filter out the vaccine candidates which are foreign to the human body, all the epitopes that show positive immunogenicity are subjected to Multiple Peptide Match tool [7] against human reference proteome.

Docking studies

To further confirm the candidacy of the foreign epitopes as a vaccine, the top three foreign epitopes based on immunogenicity scores were subjected to molecular docking studies to confirm their interactions with the specified HLA at the peptide-binding groove. The molecular docking of the peptide epitope with the HLA structure was performed using HPEPDOCK Server [8]. The docking was performed without specifying the binding site residues to investigate if the studied epitopes would bind at the peptide-binding groove without any lead. The interaction diagrams are generated using LigPlot+ [9].

Results And Discussion

253 CTL epitopes are predicted by NetCTLpan across the proteins coded by SARS-CoV-2 Indian isolate towards HLA A3, B7 and B44. The epitopes as predicted by NetCTLpan are shown in Supplementary Table S1 online. The scores of the peptides constitute of combined scores of HLA binding, TAP-transporter binding and C terminal cleavage prediction scores.

The IEDB immunogenicity tool was used to calculate the immunogenicity of the epitopes as it plays an important role in examining the immune response. IEDB immunogenicity tool returned 139 epitopes with positive scores. The data showing the immunogenicity score of the epitopes are shown in Supplementary Table S2 online.

All the epitopes with positive immunogenicity scores are subjected to Multiple Peptide Match tool to filter out the epitopes that are foreign to the human body. It was observed that all the epitopes that showed positive score for immunogenicity calculation were foreign to the human proteome. The peptide matching step was performed on sequence dataset 'UniProtKB release 2020_01 plus isoforms | SwissProt | Isoform' with target organism set as 'Homo sapiens [9606]'. The output log had stated that 0 out of 139 unique peptides had matches in 0 protein(s) found in 0 organism(s) confirming their foreignness to the human body. The top five vaccine candidates as predicted by the above steps along with their epitope prediction score by NetCTLpan and immunogenicity score are given in Table 1.

Table 1. The epitope prediction score and immunogenicity score of top five vaccine candidates.

Epitope	Protein	HLA	NetCTLpan score	Immunogenicity score
LVAEWFLAY	orf1ab polyprotein	HLA-A03	0.70209	0.45285
WPWYIWLGF	surface glycoprotein	HLA-B07	0.68398	0.41673
REHEHEIAW	orf1ab polyprotein	HLA-B44	0.82399	0.37218
HVTFFIYNK	orf3a protein	HLA-A03	0.78979	0.36278
SPRWYFYYL	nucleocapsid phosphoprotein	HLA-B07	1.00321	0.34101

As stated earlier, the top three vaccine candidates based on the above investigation are subjected to molecular docking studies using HPEPDOCK Server. The interaction diagrams revealed that the peptide epitopes bind to the peptide-binding groove even though they were subjected to blind docking. Thus, confirming their vaccine candidacy. The interaction diagrams of the docked epitopes are shown in Fig 1-3. The common interactions between the reference peptide available in the PDB file and the studied epitope are circled to facilitate easier identification.

Conclusion

Designing a vaccine is of top priority in time of a pandemic. In this study, we attempted to identify potential CTL epitopes from the SARS-CoV-2 Indian isolate for Indian population using a bioinformatics approach.

The list of CTL epitopes was predicted using NetCTLPan server which considers HLA binding affinity, TAP transport efficiency and C-terminal cleavage to identify the epitopes. Further, Immunogenicity scores were

calculated using the IEDB immunogenicity tool to identify the potential vaccine candidates. As the human body shows immune response only towards a foreign antigen, all the epitopes with positive immunogenic score are subjected to peptide matching against the human proteome. The results of the peptide matching tool reveal that all the predicted immunogenic epitopes are foreign to the human body. These unique immunogenic epitopes were further docking with their respective HLA molecule to study their interactions with the HLA molecule. The docking studies revealed that all the studied epitopes bind at the peptide-binding site of the HLA confirming their vaccine candidacy. The epitopes identified in this study can be further subjected to in vitro and in vivo studies design a vaccine against the dreadful SARS-CoV-2.

Declarations

Data availability:

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Acknowledgements

The research work was carried out in the laboratory of the Bioinformatics, Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj-India. Viswajit Mulpuru is grateful to the Ministry of Human Resource Development, Government of India and Indian Institute of Information Technology Allahabad for the scholarship to pursue a doctoral degree.

Author contributions statement

N.M. and V.M. conceived the study, V.M. conducted the study and analysed the result(s). V.M wrote the main manuscript and prepared the figures. All authors reviewed the manuscript.

Additional information

Accession codes:

GenBank ID of the genome: MT050493.1 PBD ID for structure of HLA-A3: 6O9C PBD ID for structure of HLA-B7: 6AT5 PBD ID for structure of HLA-B44: 3KPS

Competing interests

The authors declare no competing interests.

References

1. Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., ... & Guan, L. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*.

2. Laing, K., & Fred Hutchinson Cancer Research Centre, Seattle, USA. (n.d.). Immune responses to viruses. Retrieved March 30, 2020, from https://www.immunology.org/public- information/bitesized-immunology/pathogens-and-disease/immune-responses-viruses
3. Stranzl, T., Larsen, M. V., Lundegaard, C., & Nielsen, M. (2010). NetCTLpan: pan-specific MHC class I pathway epitope predictions. *Immunogenetics*, *62*(6), 357-368.
4. dos Santos Francisco, R., Buhler, S., Nunes, J. M., Bitarello, B. D., França, G. S., Meyer, D., & Sanchez-Mazas, A. (2015). HLA supertype variation across populations: new insights into the role of natural selection in the evolution of HLA-A and HLA-B *Immunogenetics*, *67*(11-12), 651-663.
5. Calis, J. J., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., ... & Peters, (2013). Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS computational biology*, *9*(10).
6. Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., ... & Peters, B. (2019). The immune epitope database (IEDB): 2018 update. *Nucleic acids research*, *47*(D1), D339-D343.
7. Barker, C., Garavelli, J. S., Huang, H., McGarvey, P. B., Orcutt, B. C., Srinivasarao, G. Y., ... & Pfeiffer, F. (2000). The protein information resource (PIR). *Nucleic acids research*, *28*(1), 41- 44.
8. Zhou, P., Jin, B., Li, H., & Huang, S. Y. (2018). HPEPDOCK: a web server for blind peptide– protein docking based on a hierarchical algorithm. *Nucleic acids research*, *46*(W1), W443-W450.
9. Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand–protein interaction diagrams for drug discovery.

Figures

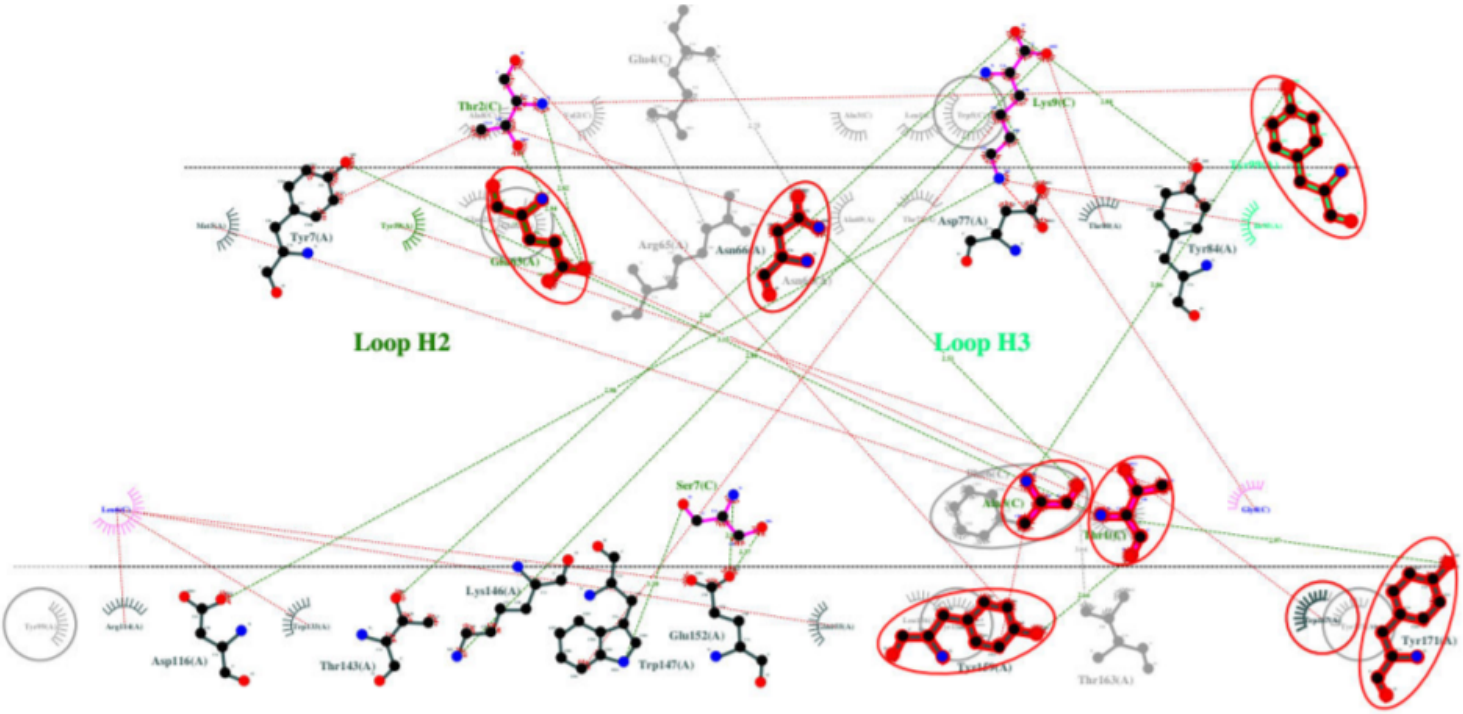

Figure 1

The interactions between HLA-A3 (PDB ID: 6O9C) and the epitope LVAEWFLAY from the SARS-CoV- 2 proteome.

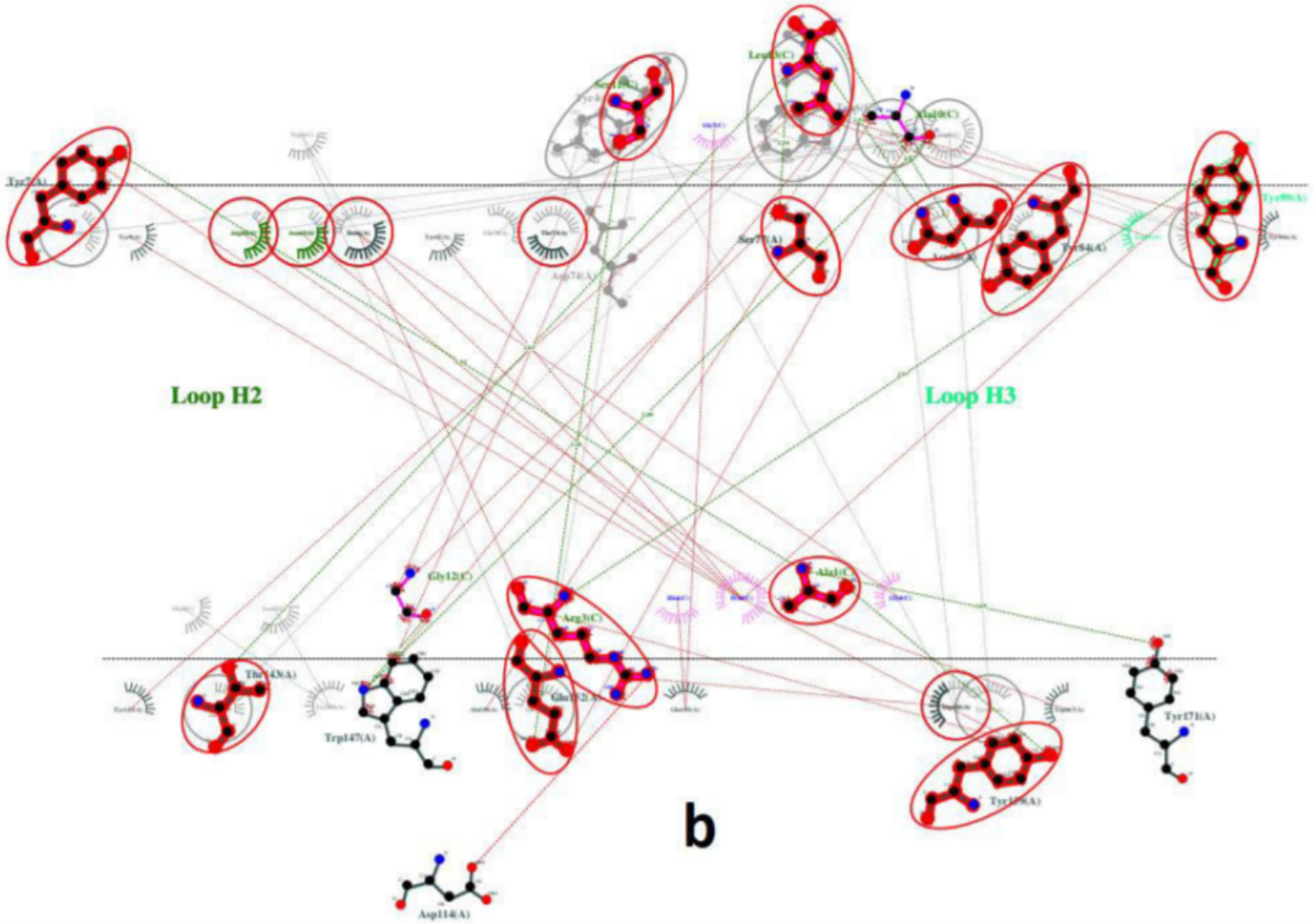

Figure 2

The interactions between HLA-B7 (PDB ID: 6AT5) and the epitope WPWYIWLGF from the SARS- CoV-2 proteome.

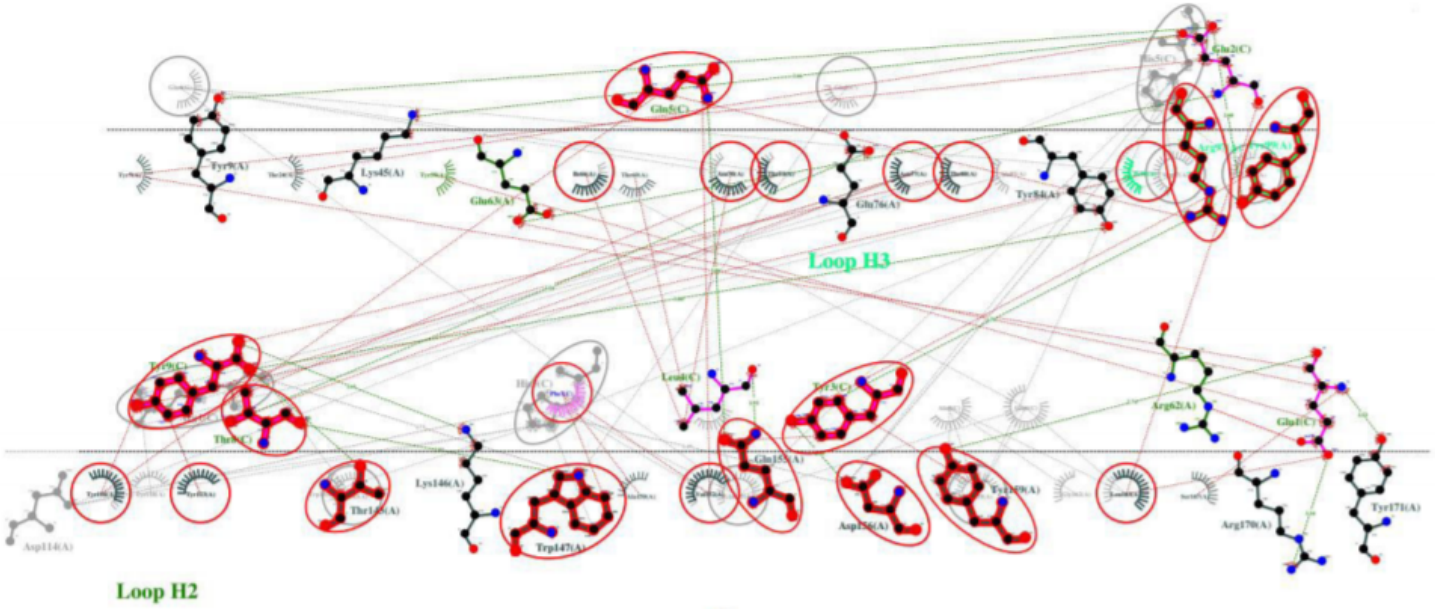

Figure 3

The interactions between HLA-B44 (PDB ID: 3KPS) and the epitope REHEHEIAW from the SARS-CoV- 2 proteome.

Supplementary Files

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

- Supplementaryinformation.pdf