



Designing of cytotoxic T lymphocyte-based multi-epitope vaccine against SARS-CoV2: a reverse vaccinology approach

Reethu Vivekanandam, Kamarajan Rajagopalan, Madesh Jeevanandam, Harsha Ganesan, Vaishnavi Jagannathan, Jackson Durairaj Selvan Christyraj, Kalishwaralal Kalimuthu, Johnson Retnaraj Samuel Selvan Christyraj & Manikandan Mohan

To cite this article: Reethu Vivekanandam, Kamarajan Rajagopalan, Madesh Jeevanandam, Harsha Ganesan, Vaishnavi Jagannathan, Jackson Durairaj Selvan Christyraj, Kalishwaralal Kalimuthu, Johnson Retnaraj Samuel Selvan Christyraj & Manikandan Mohan (2021): Designing of cytotoxic T lymphocyte-based multi-epitope vaccine against SARS-CoV2: a reverse vaccinology approach, *Journal of Biomolecular Structure and Dynamics*, DOI: [10.1080/07391102.2021.1993338](https://doi.org/10.1080/07391102.2021.1993338)

To link to this article: <https://doi.org/10.1080/07391102.2021.1993338>



[View supplementary material](#)



Published online: 25 Oct 2021.



[Submit your article to this journal](#)



Article views: 188



[View related articles](#)



[View Crossmark data](#)



Designing of cytotoxic T lymphocyte-based multi-epitope vaccine against SARS-CoV2: a reverse vaccinology approach

Reethu Vivekanandam^{a*}, Kamarajan Rajagopalan^{b*}, Madesh Jeevanandam^c, Harsha Ganesan^d, Vaishnavi Jagannathan^e, Jackson Durairaj Selvan Christyraj^b, Kalishwaralal Kalimuthu^{f#}, Johnson Retnaraj Samuel Selvan Christyraj^b and Manikandan Mohan^{g,h}

^aDepartment of Biotechnology, Bharathiyar University, Coimbatore, Tamilnadu, India; ^bRegeneration and Stem Cell Biology Lab, Centre for Molecular and Nanomedical Sciences, International Research Centre, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India; ^cDepartment of Biochemistry, PSG college of Arts and Science, Coimbatore, Tamilnadu, India; ^dChettinad Academy of Research and Education (CARE), Chettinad Hospital and Research Institute (CHRI), Chennai, Tamilnadu, India; ^eInstitute of Forest Genetics and Tree Breeding (IFGTB), Coimbatore, Tamilnadu, India; ^fDepartment of Surgery, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA; ^gVaxigen International Research Center Private Limited, Coimbatore, Tamilnadu, India; ^hDepartment of Pharmaceutical and Biomedical Sciences, College of Pharmacy, University of Georgia, Athens, GA, USA

Communicated by Ramaswamy H. Sarma

ABSTRACT

SARS-CoV2 is a single-stranded RNA virus, gaining much attention after it out broke in China in December 2019. The virus rapidly spread to several countries around the world and caused severe respiratory illness to humans. Since the outbreak, researchers around the world have devoted maximum resources and effort to develop a potent vaccine that would offer protection to uninfected individuals against SARS-CoV2. Reverse vaccinology is a relatively new approach that thrives faster in vaccine research. In this study, we constructed Cytotoxic T Lymphocytes (CTL)-based multi-epitope vaccine using hybrid epitope prediction methods. A total of 121 immunogenic CTL epitopes were screened by various sequence-based prediction methods and docked with their respective HLA alleles using the AutoDock Vina v1.1.2. In all, 17 epitopes were selected based on their binding affinity, followed by the construction of multi-epitope vaccine by placing the appropriate linkers between the epitopes and tuberculosis heparin-binding hemagglutinin (HBHA) adjuvant. The final vaccine construct was modeled by the I-TASSER server and the best model was further validated by ERRAT, ProSA, and PROCHECK servers. Furthermore, the molecular interaction of the constructed vaccine with TLR4 was assessed by ClusPro 2.0 and PROtein binDIng enerGY prediction (PRODIGY) server. The immune simulation analysis confirms that the constructed vaccine was capable of inducing long-lasting memory T helper (Th) and CTL responses. Finally, the nucleotide sequence was codon-optimized by the JCAT tool and cloned into the pET21a (+) vector. The current results reveal that the candidate vaccine is capable of provoking robust CTL response against the SARS-CoV2.

ARTICLE HISTORY

Received 19 February 2021
Accepted 10 October 2021

KEYWORDS

Virus; Vaccine; alleles; CTL;
Reverse vaccinology

1. Introduction

SARS-CoV2 or 2019-nCoV, officially known as coronavirus disease of 2019 (COVID-19) was first observed in animals and individuals from the Wuhan seafood market in China has spread globally affecting millions of people (Lu et al., 2020; Wang et al., 2020). Coronaviruses (CoVs) belong to the *Coronoviridae* family and were initially reported in the early 1930s. SARS-CoV and SARS-CoV2 belong to the beta coronavirus class and primarily cause respiratory infections in humans and various animals (Cheng et al., 2007). Initially, there are four major SARS-CoV2 strains under circulation - HKU1, NL63, OC43, and E229 (Roussel et al., 2020). The origins of SARS-CoV2 is unclear at the moment. However, it

shares a high sequence similarity with the Bat-CoV-RaTG13 strain that infects bats in Southeast provinces of China (Zhang et al., 2020; Zhou et al., 2020; Zhu et al., 2020). Moreover, rats, pangolins, and raccoons are suspected intermediate hosts for human transmission (Wong et al., 2020; Zhang et al., 2020). SARS-CoV2 infected individuals exhibit symptoms like common cold, flu, headache, respiratory ailments, while pneumonia and ultimately fatal infections have been reported in severe cases (Guo et al., 2020). SARS-CoV2 has been found to rapidly spread among infants, children, pregnant women, elderly, immune-suppressed, and immune-compromised patients with ease, implying high rate of infections (Shereen et al., 2020). Majority of infected individuals show little or no symptoms, with

CONTACT Manikandan Mohan Manikandan.Mohan@uga.edu College of Pharmacy, University of Georgia, Athens, GA 30605, USA; Johnson Retnaraj Samuel Selvan Christyraj johnnbt@sathyabama.ac.in Regeneration and Stem Cell Biology Lab, Centre for Molecular and Nanomedical Sciences, International Research Centre, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India

*Authors contributed equally

#MK Bhan Young Researcher Fellow, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala, India.

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07391102.2021.1993338>

© 2021 Informa UK Limited, trading as Taylor & Francis Group

characteristic manifestations of the disease appearing after progression to a severe state. As a result, asymptomatic spreads play a major role in the COVID-19 pandemic (Dai et al., 2020).

SARS-CoV2 is a single-stranded, positive-sense RNA virus that encodes seven accessory proteins (ORF1ab, ORF3a, ORF6, ORF7a, ORF7b, ORF8 and ORF10) (Islam et al., 2020; Yang et al., 2020) and four structural proteins, namely Envelope (E), Membrane (M), Spike (S), and Nucleocapsid protein (Nguyen et al., 2021). The spike protein plays a central role in the viral cell entry and replication. Both humoral and cellular immune responses towards the viral spike protein play a major role in controlling SARS-CoV2 infection. Specific antibodies (IgG and IgM) are evident approximately two weeks post-infection (Long et al., 2020; Wu et al., 2020). Meanwhile, the increased CTL responses are also monitored during the first week of infection (Thevarajan et al., 2020). A key focus for majority of vaccines lies in inducing the humoral immune responses towards the viral glycoproteins. On the other hand, CTL immunity towards the SARS-CoV2 also plays an essential role viral clearance (Peng et al., 2020). However, the CTL mediated immunity is activated only after binding of short (9mer to 11mer) peptides to class I MHC and its subsequent interaction with T cell receptor (TCR) (Trolle et al., 2016). Hence, it is critical to reliably identify class I MHC binders that form the basis for CTL epitope vaccine design.

In the recent times, owing to the progress in computational power and techniques, vaccine design concepts can rapidly be translated into practical examples (He & Zhu, 2015). In contrast, traditional vaccine development is a time consuming and labor-intensive process. Modern immunoinformatics methods can be used to design the vaccine effectively and in a cost-efficient manner (Patronov & Doytchinova, 2013). The past three decades have witnessed the development of a wide range of algorithms for T cell epitope prediction. Early algorithms were primarily sequence-based, while modern ones powered by matrix-driven methods, motifs binding, QSAR approach, molecular docking, homology modeling, and machine-learning algorithms have a much-improved predictive power (Tomar & De, 2010). Modern immunoinformatics provides varied resources that aid in the design of potent peptide-based vaccine (Dhanda et al., 2017). In this present study, we utilize the sequence- and structure-based approach to predict potent CTL epitopes. Moreover, these epitopes were used to construct a multi-epitope vaccine targeting SARS-CoV2.

2. Methodology

2.1. Sequence-based methods

2.1.1. Protein sequence retrieval

SARS-CoV2 structural proteins; Envelope Protein (EP) (Accession No. QHD43418.1), Membrane Glycoprotein (MGP) (Accession No. QHD43419.1), Nucleocapsid Phosphoprotein (NP) (Accession No. QHD43423.2) and Surface Glycoprotein (SGP) (Accession No. QHD43416.1) and non-structural proteins: orf1ab polyprotein (Accession No. QHD43415.1), ORF3a protein (Accession No. QHD43417.1), ORF6 protein (Accession

No. QHD43420.1), ORF7a protein (Accession No. QHD43421.1), ORF8 protein (Accession No. QHD43422.1) and ORF10 protein (Accession No. QHI42199.1) were retrieved from NCBI database (<https://www.ncbi.nlm.nih.gov/>) for CTL epitope mapping. Furthermore, protein sequences from different SARS-CoV2 variants such as B.1.1.7 (Alpha), B.1.351 (Beta), B.1.427 (Epsilon), B.1.429 (Epsilon), B.1.526.1 (Epsilon), B.1.526 (Iota), B.1.617.1 (Delta), B.1.617.2 (Delta), P.1 (Gamma) and P.2 (Zeta) were retrieved from NCBI database which have been represented in Table 1.

2.1.2. Class I MHC binder's prediction

The high-frequency HLA supertype alleles (A1 (HLA-A*01:01), A2 (HLA-A*02:01), A3 (HLA-A*03:01), A24 (HLA-A*24:02), A26 (HLA-A*26:01), B7 (HLA-B*07:02), B8 (HLA-B*08:01), B27 (HLA-B*27:05), B39 (HLA-B*39:01), B40 (HLA-B*40:01), B58 (HLA-B*58:01), and B62 (HLA-B*15:01) were selected to predict all possible 8-mer, 9-mer, 10-mer, and 11-mer length epitopes using Immune Epitope Database (IEDB), respectively (Kim et al., 2012). The class I MHC binders were selected based on the percentile rank ≤ 1 with the default prediction methods.

2.1.3. Class I MHC peptide processing prediction

The CTL peptides undergo multiple processing steps prior to TCR presentation. The peptides cleaved by the proteasomes need to bypass multiple cytosolic proteases prior to their transport into the endoplasmic reticulum through TAP (Peters et al., 2003). Moreover, identification and mapping of a CTL epitope is critical step prior to its incorporation in a vaccine (Mohan et al., 2018). In this study, all class I MHC binders were first identified and further processed using the MHC-I processing prediction tool from IEDB to predict proteasomal processing peptides (<http://tools.iedb.org/processing/>) (Peters et al., 2003; Tenzer et al., 2005). The default prediction method was used for the processing, with positively predicted peptides being considered as plausible potent epitopes.

2.1.4. Class I MHC peptide immunogenicity prediction

The prediction of immunogenic epitopes is an essential step for CTL based epitope vaccine development. Hence, the class I Immunogenicity prediction tool provided by IEDB (<http://tools.iedb.org/immunogenicity/>) was used to differentiate between immunogenic and non-immunogenic peptides that were assigned positive values by the MHC-I processing prediction tool (Calis et al., 2013). Likewise, the prior step, default settings were utilized, while the peptides with positive predictions being assumed as potent epitopes.

2.1.5. Antigenicity, toxicity, non-allergic, non-homologous and conservancy analysis

It is critical to assess the antigenicity of a predicted immunogenic epitope prior to its *in vitro/in vivo* assessment. Here, the plausible antigenic nature of the positive peptides were assessed using the VaxiJen v2.0 tool (<http://www.ddg-pharm-fac.net/vaxijen/VaxiJen/VaxiJen.html>) (Doytchinova & Flower,

Table 1. SARS-CoV-2 variants proteome and their NCBI ID.

Covid-19 Variants/ Proteins	B.1.1.7_Southeast England (Alpha)	B.1.351_Nelson Mandela Bay	B.1.427_Southern California, USA (Epsilon)	B.1.429_Southern California, USA (Epsilon)	B.1.526_1_New York (Epsilon)	B.1.526_New York (Epsilon)	B.1.617.1_India (Delta)	B.1.617.2_UK (Delta)	P.1_Brazil (Gamma)	P.2_Brazil (Zeta)
EP	QQK53359.1	QRG27675.1	QRJ61563.1	QQQ16720.1	QRW77963.1	QTP07466.1	QUC98739.1	QQQ16720.1	QRX39427.1	
M	QQK53360.1	QRG27676.1	QRJ61564.1	QQQ16721.1	QRW77964.1	QTP07467.1	QUC98740.1	QQQ16721.1	QRX39428.1	
NP	QQK53365.1	QRG27681.1	QRJ61569.1	QQQ16726.1	QRW77969.1	QTP07472.1	QUC98745.1	QQQ16726.1	QRX39433.1	
ORF1AB	QQK53355.1	QRG27671.1	QRJ61559.1	QQQ16716.1	QRW77959.1	QTP07462.1	QUC98735.1	QQQ16716.1	QRX39423.1	
ORF 3	QQK53358.1	QRG27674.1	QRJ61562.1	QQQ16719.1	QRW77962.1	QTP07465.1	QUC98738.1	QQQ16719.1	QRX39426.1	
ORF 6	QQK53361.1	QRG27677.1	QRJ61565.1	QQQ16722.1	QRW77965.1	QTP07468.1	QUC98741.1	QQQ16722.1	QRX39429.1	
ORF 7a	QQK53362.1	QRG27678.1	QRJ61566.1	QQQ16723.1	QRW77966.1	QTP07469.1	QUC98742.1	QQQ16723.1	QRX39430.1	
ORF 7b	QQK53363.1	QRG27679.1	QRJ61567.1	QQQ16724.1	QRW77967.1	QTP07470.1	QUC98743.1	QQQ16724.1	QRX39431.1	
ORF 8	QQK53364.1	QRG27680.1	QRJ61568.1	QQQ16725.1	QRW77968.1	QTP07471.1	QUC98744.1	QQQ16725.1	QRX39432.1	
ORF 10	QQK53366.1	QRG27682.1	NII	QQQ16727.1	QRW77970.1	QTP07473.1	QUC98746.1	QQQ16727.1	QRX39434.1	
SGP	QQK53357.1	QRG27673.1	QRJ61561.1	QQQ16718.1	QRW77961.1	QTP07464.1	QUC98737.1	QQQ16718.1	QRX39425.1	

2007). This step was followed by a toxicity analysis using the ToxinPred tool (<http://crdd.osdd.net/raghava/toxinpred/>). The epitopes deemed non-toxic by ToxinPred were further assessed for their allergenicity by the AllerTop v.2.0 tool (Dimitrov et al., 2014) (<https://www.ddg-pharmfac.net/AllerTOP/>). The default prediction value was set by all three prediction tools to characterize the immunogenic epitopes. In order to avoid any autoimmune responses, the epitopes were subjected to a protein BLAST analysis with the human proteomes that yielded non-homologous peptides (e value < 0.05) (Mehla & Ramana, 2016). Finally, the epitope conservancy among other SARS-CoV2 variants was studied using the IEDB conservancy analysis tool (Bui et al., 2007) (<http://tools.iedb.org/conservancy/>). A sequence identity cutoff of 100% was used in the identification of highly conserved epitopes.

2.2. Structure-based methods

2.2.1. Retrieval of HLA molecules and homology modeling

The following class I MHC super alleles (HLA-A*01:01 (PDB id: 6AT9), HLA-A*02:01 (PDB id: 4NO3), HLA-A*03:01(PDB id: 2XPG), HLA-A*24:02 (PDB id: 4F7P), HLA-B*07:02 (PDB id: 5EO0), HLA-B*08:01 (PDB id: 4QRT), HLA-B*15:01 (PDB id: 5TXS), HLA-B*27:05 (PDB id: 4G9D), HLA-B*39:01 (PDB id: 4O2C), HLA-B*40:01 (PDB id: 6IEX), HLA-B*58:01 (PDB id: 5IND) crystallographic structures were retrieved from RCSB Protein Data Bank (PDB) (Berman et al., 2000) (<https://www.rcsb.org/>) and their interactions with corresponding peptides were evaluated. The crystallographic structures HLA-A*26:01 alleles is not available in the PDB database. Therefore, the HLA-A*26:01 alpha sub-unit amino acid sequence (Accession No. AAA03720.1) was retrieved from the NCBI database. Subsequently, the protein modeling was carried out with the help of SWISS-MODEL server (<https://swissmodel.expasy.org/interactive>) (Waterhouse et al., 2018). Later, the modeled protein was further refined by GalaxyRefine server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) (Ko et al., 2012). Furthermore, ProSA (Protein Structure Analysis) (<https://prosa.services.came.sbg.ac.at/prosa.php>) (Wiederstein & Sippl, 2007), ERRAT (<https://servicesn.mbi.ucla.edu/ERRAT>) (Colovos & Yeates, 1993) and PROCHECK (Programs to check the Stereochemical Quality of Protein Structures) (<https://servicesn.mbi.ucla.edu/PROCHECK/>) (Laskowski et al., 1993) were used to validate the quality of the modeled HLA-A*26:01 structure.

2.2.2. Molecular docking of MHC peptide

The Chimera tool (Pettersen et al., 2004) was utilized to build the peptide 3D structure and used as a ligand; whereas, the AutoDock Vina tool was used to carry out a peptide-MHC docking study (Trott & Olson, 2010). Prior to docking, the native peptide, β_2 microglobulin, co-crystallized solvents and heteroatoms were removed from all of the class HLA crystal structures using the Discovery studio 4.1 (Jain & Baranwal, 2019). The HLA allele was further processed by adding hydrogen atoms and Kollman partial charges using AutoDock tool. Meanwhile, the all-rotatable bonds except the backbone amide

bonds were set to be mobile for the peptides. Finally, the HLA alleles and peptides ligands were converted into PDBQT format prior to docking. The docking grid was set to an appropriate size that accommodated the entire receptor, while default settings were used for the docking (Mohammed et al., 2018). The default output (containing 10 models) depicting various interactions between the HLA allele and their respective peptides, were sorted based on their binding affinity. The output models were retrieved in the PDBQT file format and visualization in Discovery Studio viewer (<https://discover.3ds.com/discovery-studio-visualizer-download>).

2.2.3. Theoretical validation of T cell epitopes

The IEDB database provides the option to analyze experimentally validated epitopes from various organisms (Mohan et al., 2020). In this instance, putative MHC binders were searched in the IEDB database to identify the experimentally proven T cell epitopes. The query epitopes were subjected to a BLAST analysis with the exact match option to predict the experimentally validated SARS specific epitopes.

3. Vaccine constructs preparation

3.1. Construction of multi-epitope CTL vaccine

The CTL epitopes restricted by various HLA supertypes were selected for multi-epitope CTL vaccine construct. The potent immunogenic epitope was selected on the basis of higher binding affinity and higher number of hydrogen bonds involved in the interactions between the peptide and HLA allele which describes more stable epitopes (Chen et al., 2016). The N terminal region of the multi-epitope CTL vaccine construct was embedded by Heparin Binding Hemagglutinin Adjuvant (HBHA) and linked with CTL epitopes using EAAAK linker. Furthermore, AAY and GPGPG linkers were placed between the CTL epitopes and PADRE peptide (AKFVAAWTLKAAA) respectively, for better CTL activation.

3.2. Antigenicity, toxicity, allergenicity, and physicochemical properties assessment

The antigenicity, toxicity, and allergenicity of the vaccine construct were evaluated by VaxiJen v2.0, ToxinPred, and AllerTop v.2.0 tools, respectively. The ExPASy ProtParam tool (<https://web.expasy.org/protparam/>) was used to determine the physicochemical property of vaccine construct such as molecular weight, half-life, instability index, theoretical isoelectric point, charge and hydropathicity (Gasteiger et al., 2005). Finally, the SOLpro server (Magnan et al., 2009) was used to study the soluble nature of the final vaccine construct over the expression in *E.coli* (<http://scratch.proteomics.ics.uci.edu/>).

3.3. Secondary and tertiary structure prediction

SOPMA server was used to investigate the secondary structure of the vaccine construct (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). Based

on the amino acid sequence, the improved SOPMA server jointly predicts the secondary structure with neural network method and provides 82.2% accuracy (Geourjon & Deleage, 1995). The I-TASSER tool (Roy et al., 2010) was used to build the tertiary structure of vaccine construct. The final structure was selected based on the C score (range between -5 to 2), with a higher positive score implying a higher reliability of the vaccine construct's structure (<https://zhanglab.cmb.med.umich.edu/I-TASSER/>).

3.4. Tertiary structure refinement and validation

The tertiary structure refinement and validation of the vaccine construct was performed in an identical fashion to HLA-A*26:01 model and using the same tools (GalaxyRefine, ProSA, ERRAT and PROCHECK).

3.5. Molecular docking and simulation of vaccine construct with TLR4

It is crucial to assess the molecular interactions between vaccine construct and toll-like receptor (TLR) for determining the potent immune response. Briefly, TLR4 (PDB: 3FXI) crystallographic structure was retrieved from the PDB database and used as a receptor for docking studies. The final receptor was prepared by removing the other ligand and chemical molecules from the TLR4 crystal structure, while 3D vaccine model used as a ligand. The complete vaccine construct and the TLR immune receptor interactions were investigated with the help of ClusPro 2.0 server (<https://cluspro.bu.edu/queue.php>) (Destra et al., 2020). The best model was selected based on the cluster size, the PIPER energy of the cluster center and the lowest energy structure in the cluster. Moreover, the binding affinity (in kcal/mol) of the vaccine to the TLR4 receptor was identified using the PRODIGY server (<https://wenmr.science.uu.nl/prodigy/>) (Vangone et al., 2019). The interactions between the TLR and vaccine construct were analyzed using PDBsum server (Laskowski et al., 2018) (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>). As a final step, the dynamic interactions in between TLR4 and the vaccine construct were studied with the aid of the iMODS (<http://imods.chaconlab.org/>) server (López-Blanco et al., 2014). The dynamic parameters such as deformability values, Eigen values, B-factor, variance, and co-variance map were investigated for the candidate vaccine.

3.6. Codon optimization and in silico cloning of vaccine construct

The amino acid sequence of the vaccine construct was back-translated using EMBOSS Backtranseq (https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/) (Rice et al., 2000). To improve the translation efficiency, the Java Codon Adaptation Tool (JCAT) (Grote et al., 2005) was utilized to optimize the codon usage to the K12 *Escherichia coli* strain. Additionally, the NdeI and NheI restriction sites were added to the 5' and 3' end of the optimized sequence respectively. The SnapGene tool was

utilized to clone the optimized sequence into the pET28a (+) 6x-His-TEV vector.

3.7. Immune simulation efficiency of vaccine construct

The immunological response of the vaccine construct was determined with the help of *in silico* immune simulation C-ImmSim 10.1 server (Rapin et al., 2010). The server predicts the epitopes based on the Position Specific Scoring Matrix (PSSM). Also, machine-learning method was used to predict immune interaction to various cells (CTL, THL, B cell, Natural killer cells, macrophages and dendritic cells), immunoglobulins and cytokines (Shey et al., 2019). The following parameters were selected to run the immune simulation: Random Seed: 12345, Simulation Volume: 10, and Simulation Steps: 1050. Moreover, the combinations of HLA supertype alleles (Test set1: A0101, A0201, B0702 and B0801; Test set2: A0301, A2402, B1501 and B2705; Test set3: A020, A2601, B3901 and B4001; Test set4: A0101, A0201, B0702 and B5801) were entered in the respective columns to analyze the CTL response. The default simulation parameters were selected and the time of injection was set at 1, 84, and 126.

4. Results

4.1. Sequence-based methods

4.1.1. Class I MHC binder's prediction

A total of 2240 Class I MHC binders were predicted from all structural and non-structural proteomes of SARS-CoV2 that are restricted by all 12 Class I MHC super types (Supplementary Table 1). Among these, the highest numbers are attributed to A3 supertype 11.07% (Total: 248) epitopes are predicted. Followed by 10.36% (Total: 232), 10.27% (Total: 230), 10.13% (Total: 227), 10.13% (Total: 227), 9.42% (Total: 211), 7.86% (Total: 176), 7.68% (Total: 172), 7.41% (Total: 166), 5.80% (Total: 130), 5.22% (Total: 117), and 4.64% (Total: 104) epitopes are predicted in B62, A24, A1, A26, A2, B40, B58, B39, B8, B7, and B27 super types, respectively (Figure 1). In fixed length pattern (8–11mer), a major share (1142 epitopes (50.98%) belonged to 9-mer length (Figure 2). In addition, 802 (35.80%), 207 (9.24%), and 89 (3.97%) epitopes are predicted in 10-mer, 11-mer, and 8-mer length, respectively.

4.1.2. Class I MHC peptide processing prediction

In all, 876 proteasomal processing positive epitopes were predicted from 2240 class I binders (Supplementary Table 2). From amongst these, 22.49% (Total: 197) epitopes were predicted for B62 allele followed by 12.67% (Total: 111), 11.53% (Total: 101), 11.30% (Total: 99), 11.19% (Total: 98), 8.45% (Total: 74), 5.25% (Total: 46), 4.57% (Total: 40), 3.77% (Total: 33), 3.08% (Total: 27), 2.97% (Total: 26), and 2.74% (Total: 24) for A2, A1, A24, B58, A26, B40, B39, B8, B27, B7, and A3 supertypes, respectively. In total, 50.80% (Total: 445) positive proteasomal epitopes are predicted in 9-mer in length; whereas, 40.87% (Total: 358), 6.28% (Total: 55), and 2.17%

(Total: 19) epitopes are predicted in 10-mer, 11-mer, and 8-mer length, respectively (Figure 2).

4.1.3. Class I MHC peptide immunogenicity prediction

In total, 462 epitopes were deemed with positive immunogenicity out of 876 class I processing epitopes analyzed in this study (Supplementary Table 3). Highest numbers of immunogenic epitopes are identified in B62 (Total: 101 epitopes). However, 21.86% (Total: 77), 12.34% (Total: 57), 11.47% (Total: 53), 10.39% (Total: 48), 6.28% (Total: 29), 6.06% (Total: 28), 4.76% (Total: 22), 3.46% (Total: 16), 2.60% (Total: 12), 2.16% (Total: 10), and 1.95% (Total: 9) epitopes are found in A2, A1, A24, B58, B40, A26, B39, B27, B7, B8, and A3 supertypes, respectively. In total, 246 (53.25%) immunogenic epitopes are predicted 9-mer length; whereas, 40.04% (Total: 185) and 5.19 (Total: 24) epitopes are predicted in 10-mer and 11-mer length, respectively. Only 1.52% (Total: 7) immunogenic epitopes are found to occur in 8-mer length (Figure 2).

4.1.4. Antigenicity, toxicity, non-allergic, non-homologous and conservancy analysis

The immunogenic epitopes are further analyzed in VaxiJen v2.0, ToxinPred, and AllerTOP tools to predict the antigenicity, toxicity, and allergic nature of the epitopes, respectively. In total, 222 epitopes were predicted as antigenic from amongst 463 immunogenic epitopes. Any epitope with a threshold value of ≥ 0.4 is considered to be antigenic. Interestingly, 20/222 epitopes show > 1 as the probability of being antigenic. The ORF 8 epitope IQYIDIGNY restricted by B62 supertype exhibits the highest score (Score: 2.10). The antigenic epitopes were subjected to a toxicity analysis in ToxinPred. Three toxic epitopes (A24: TYKPNTWCI, B58: NTWCIRCLW, and B58: KPNTWCIRCLW) from the ORF1ab protein were excluded from the dataset, with the rest 219 being subjected to allergenicity analysis in AllerTOP server. In all, 121/219 epitopes were deemed non-allergenic, while simultaneously being predicted as potent immunogens and lacking any sequence homology to human genome (Supplementary Table 4). The IEDB conservancy analysis revealed that 67/121 exhibited 100% sequence similarity with various SARS-CoV2 variants such as B.1.1.7 (Alpha), B.1.351 (Beta), B.1.427 (Epsilon), B.1.429 (Epsilon), B.1.526.1 (Epsilon), B.1.526 (Iota), B.1.617.1 (Delta), B.1.617.2 (Delta), P.1 (Gamma) and P.2 (Zeta). In this instance, the ORF1ab protein was found to harbor high number ($N = 30$) of conserved epitopes followed by MGP (12), SGP (8), EP (6), NP (4), ORF3a (4), ORF6 (1), ORF7a (1) and ORF10 (1) protein respectively (Supplementary Table 5).

4.2. Structure-based methods

4.2.1. Homology modeling of HLA-A*26:01 (A26) alleles

The immunogenic epitopes predicted in sequence-based methods were further used to identify the interactions with their respective HLA alleles. The HLA-A*26:01 allele structure was not found to be available in the PDB that necessitated its sequence retrieval from the NCBI database and

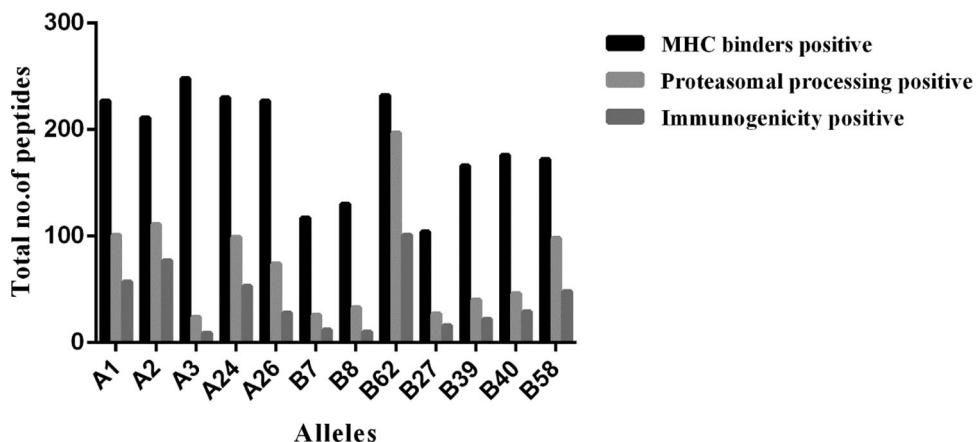


Figure 1. SARS-CoV2 epitopes predicted by the IEDB-AR tools.

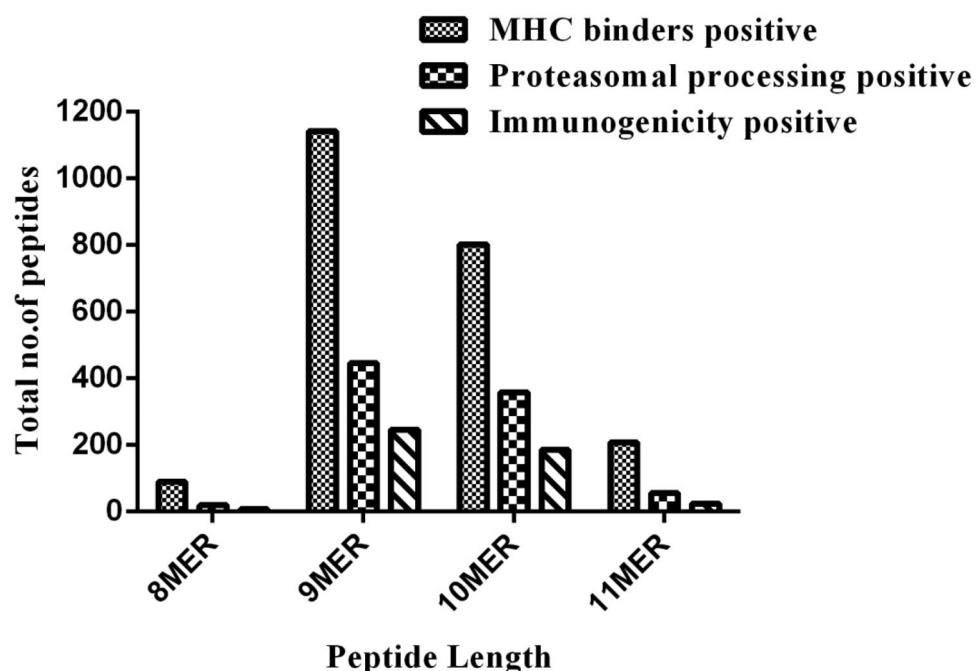


Figure 2. Length wise epitopes distribution of entire proteome of SARS-CoV2.

subsequent homology modeling using the SWISS-MODEL server. The modeled structures were then refined giving a final output of 5 models. The top-ranking model from amongst the 5 GDT-HA: 0.99; RMSD: 0.33; MolProbity: 1.58; Class score: 11.70; Poor rotamers: 0.4; Rama favored: 98.50) was selected for further validations. The overall quality of the model 1 was found to be 96.17 while the Z-score according to ProSA was -8.92. Finally, the PROCHECK analyses incorporating the Ramachandran plot revealed that 93% residues were situated in the most favored region; whereas, 6.60% and 0.40% residues were found in additionally allowed and generously allowed regions (Figure 3).

4.2.2. Molecular docking of MHC peptide

All 121 immunogenic epitopes identified by sequence-based methods were docked with their respective HLA alleles to identify potent binders. Post-docking analyses involving pose sorting on the basis of their binding affinity (kcal/mol) revealed that 81 epitopes had good binding affinity to their

respective HLA alleles (Supplementary Table 6). A total of 25 from 81 epitopes were localized in structural proteins (EP (Total: 4), NP (Total: 1), MGP (Total: 11), and SGP (Total: 9), leaving the rest 56 being localized in non-structural proteins (ORF1ab (Total: 45), ORF3a (Total: 2), ORF7a (Total: 4), ORF8 (Total: 4), and ORF10 (Total: 1). Intriguingly, ORF8 protein epitope FLGIITTVAAF exhibits the highest binding energy (-12.60) with the B62 supertype, while 13 epitopes with binding affinity (>11 kcal/mol) were also restricted with the B62 supertype. Moreover, 5 immunogenic epitopes were restricted by the A1 supertype, although none exhibited favorable binding with the respective supertype.

4.2.3. Theoretical validation of T cell epitopes

A total of 28 from 81 epitopes have been proven to be immunogenic from experiments have shown in Supplementary Table 7. From theses, 18 epitopes are potent MHC binders as confirmed from MHC ligand elution assays. Interestingly, the HLA-A*02:01 restricted epitope VMVELVAEL

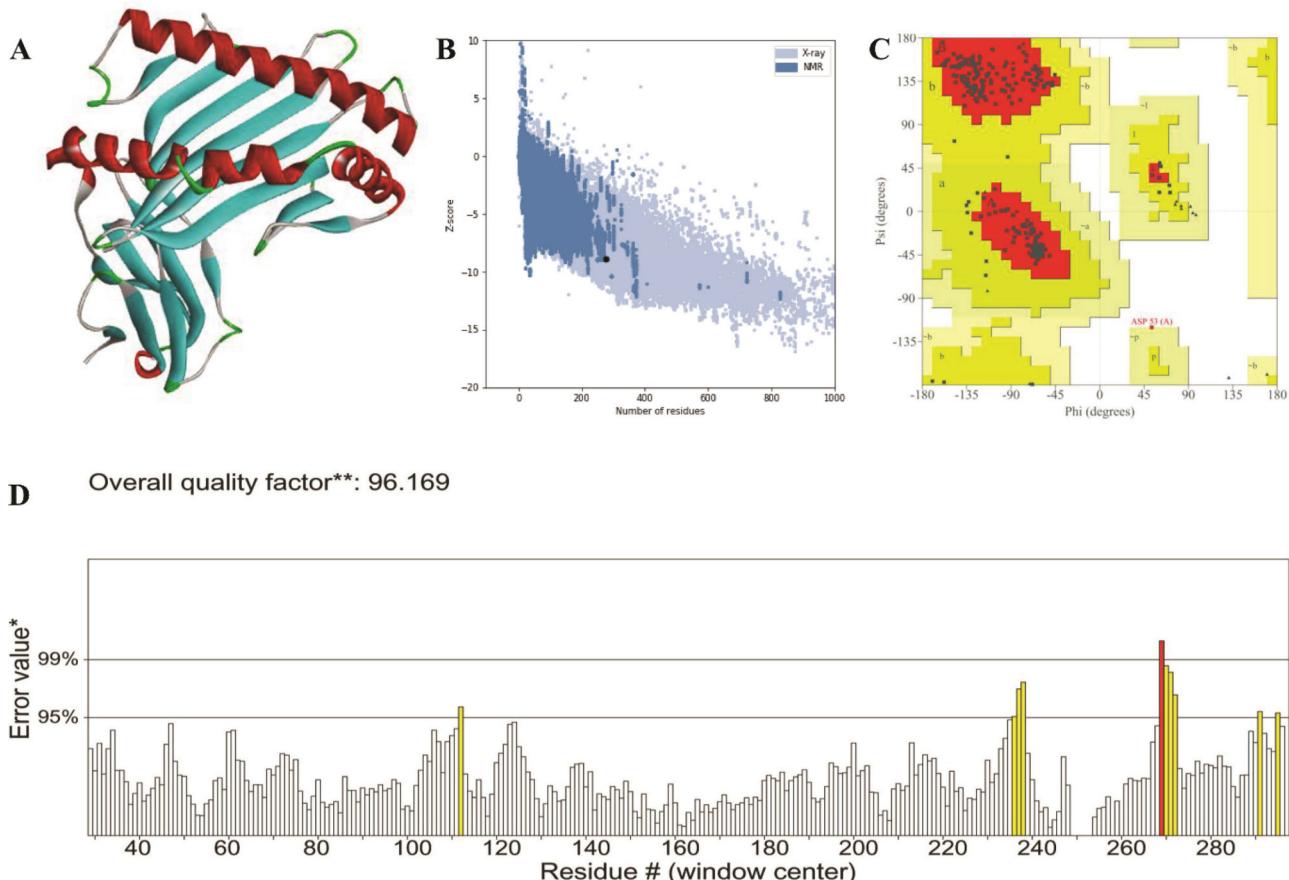


Figure 3. The HLA-A*26:01 molecules was modeled by SWISSPROT server and validated using different bioinformatics tools. A) The 3 D model of HLA-A*26:01 allele. B) Best refined model analyzed by ProSA server. The plot represents the Z-score of all experimentally validated similar size proteins structure from PDB database (dark blue dots: NMR spectroscopy and light blue dots: X-ray crystallography). The black spot in the plot indicates the validated model with Z-score of -8.92. C) Ramachandran plot generated by PROCHECK server and result showing the amino acid distribution in different regions. D) Overall quality factor was determined by ERRAT server. X-axis and Y-axis denotes the residues and error value, respectively. Two lines in the Y-axis represent the confidence and region above that error value considered to be lower quality.

(Protein: ORF1ab) and HLA-B*40:01 restricted epitope MEVTPSGTWL (Protein: NP) have shown positive result to more than 3 CTL assays.

4.3. Vaccine construct preparation

4.3.1. Construction of multi-epitope CTL vaccine

The epitope exhibiting the highest binding affinity and optimal hydrogen bonding may provoke a strong CTL response. Based on these criteria, 17 immunogenic CTL epitopes were selected that cover 11 HLA supertypes for multi-epitope vaccine construction. Among these, 7 epitopes (MGP: FLYIILKLFL (A2); MGP: LWPVTLACF (A24); SGP: RFPNITNLCPF (A24); SGP: WTAGAAAYY (A26); EP: NSVLLFLAF (B62); SGP: YQPYRVVVLSF (B62); NP: MEVTPSGTWL (B40) were selected from the structural proteins. Also, 10 epitopes (ORF1ab: WLLLTIITSL (A2); ORF1ab: HLYLQYIRK (A3); ORF1ab: VYSVIYLYLT (A24); ORF1ab: DVVAIDYKHY (A26); ORF3a: APFLFLYLYAL (B7); ORF1ab: AIWVTCLAY (B62); ORF1ab: LQAVGACVL (B62); ORF1ab: TRFFYVLGL (B27); ORF1ab: SEVGPEHSL (B40); ORF1ab: DVSFLAHIQW (B58) are selected from non-structural proteins. All the CTL epitopes were interlinked with the AAY linker while the HBHA adjuvant was attached via EAAAK linker in the N terminal region of the vaccine construct. Additionally, the universal T helper PADRE epitope (AKFVAAWTLKAAA) is added in

between structural and non-structural CTL epitope construct linked via GPGPG linker (Figure 4).

4.3.2. Antigenicity, toxicity, allergenicity, and physicochemical properties assessment

The vaccine construct was found to be immunogenic when subjected to immunogenicity analysis in the VaxiJen v2.0 server (Probability score: 0.86). Moreover, toxicity and allergenicity analyses using ToxinPred (Score: 0.50) and AllerTOP v.2.0 server deemed the construct non-toxic and non-allergic. The physicochemical properties of the vaccine construct like molecular weight, theoretical pI, Grand average of hydropathicity (GRAVY), instability index, and aliphatic index were found to be 47920.36, 8.88, 0.19, 39.04, and 98.79 respectively. In addition, the half-life of the vaccine constructs in different host systems was found to be 30 h (mammalian reticulocytes, *in vitro*), > 20 h (yeast, *in vivo*), and > 10 h (*E. coli*, *in vivo*). Lastly, the construct was found to have excellent solubility in water (0.96) according to the SOLPro server.

4.3.3. Secondary and tertiary structure prediction

Secondary structure analyses of the vaccine construct (Figure 5) using the SOPMA server predicted 72.44% for alpha helix,

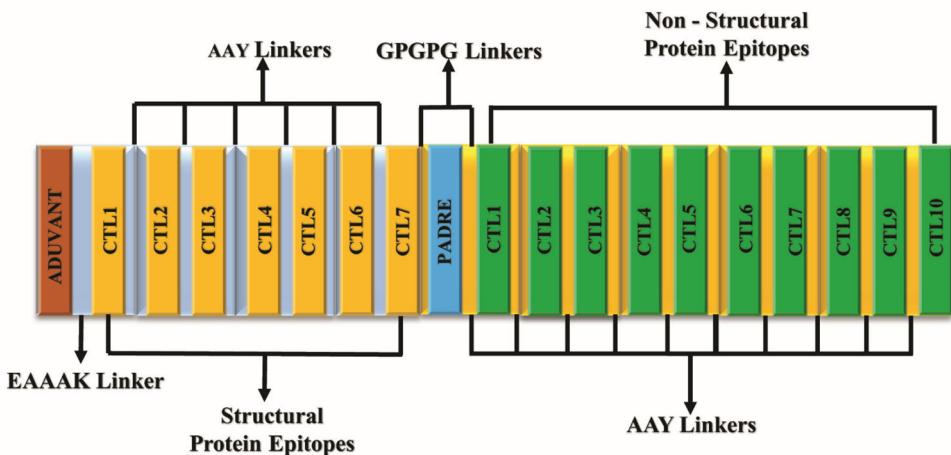


Figure 4. Schematic diagram of multi-epitope vaccine designed together with HBHA adjuvant (Red). The N terminal HBHA adjuvant linked with EAAK linker followed by structural proteins CTL epitopes (yellow), PADRE epitope (Sky blue) and non-structural proteins CTL epitopes (Green). All CTL epitopes linked with AAY linker and PADRE epitope linked with GPGPG linker.

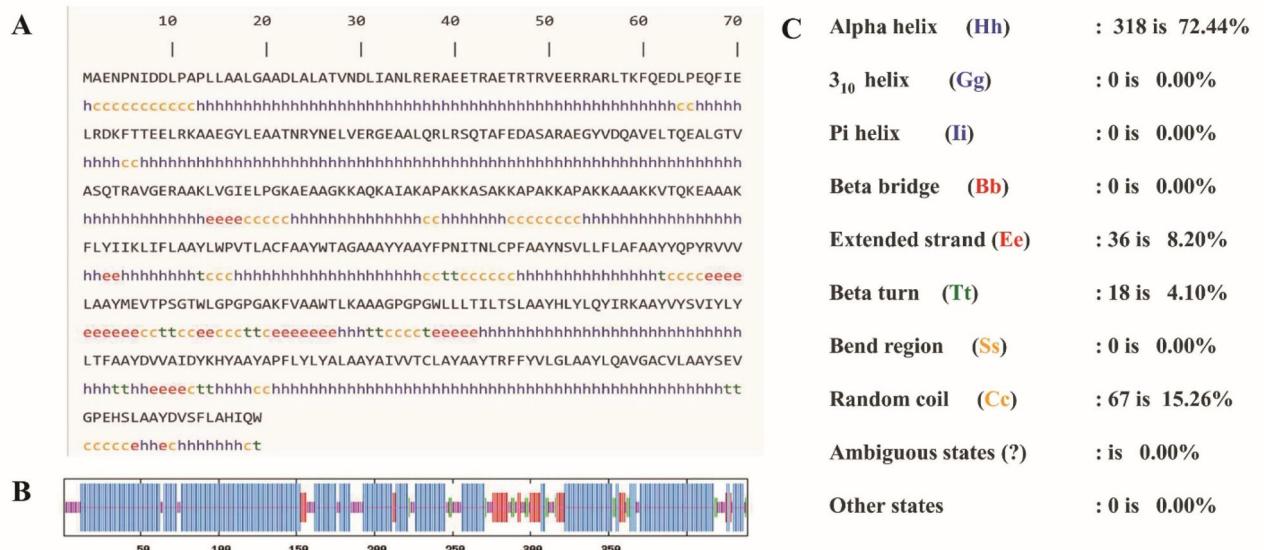


Figure 5. Predicted secondary structure of proposed multi-epitope vaccine construct. A) Visual representation of predicted secondary structure and their property analyzed by SOPMA server. B) Graphical representation of predicted secondary structure (alpha helix (blue), random coil (yellow) and extended strand (Red)).

15.20% for random coil, 8.20% for extended strand, and 4.10% for beta-turn. The tertiary structure validation using the I-TASSER server, revealed the model 3 (C score –1.48) to be the best and was used for further validation.

4.3.4. Tertiary structure refinement and validation

The selected 3D structure of the vaccine construct was further refined using Galaxy Refine tools as mentioned earlier. The mild perturbation model 1 was selected based on the scores generated by I-TASSER before and after refinement. Finally, the refined vaccine construct was analyzed by the ERRAT server and the overall quality factor score is observed as 95.94, while the ProSA Z-score was –4.89. Apart from this, the Ramachandran plot is generated by PROCHECK server and 87.40% residues are in the most favored region; whereas, 9%, 1.30%, and 2.30% residues are found in

additionally allowed, generously allowed, and disallowed regions, respectively (Figure 6).

4.3.5. Molecular docking and simulation of vaccine construct with TLR4

The interaction between refined vaccine construct with TLR4 was carried out with the aid of ClusPro 2.0 server. Based on the largest cluster size and the PIPER energy, cluster 0 (44 members) was selected out of 30 clusters ranked by the server. The centroid and lowest energy representative structures were found to have weighted scores of –1181 and –1289.1, respectively. Additionally, PRODIGY analysis revealed the binding affinity (ΔG) of the docked complex to be –14.10 kcal/mol. The intramolecular contacts were analyzed with the aid of PDBSum server. The vaccine construct formed 7 and 5 hydrogen bonds with B and D subunits of TLR4,

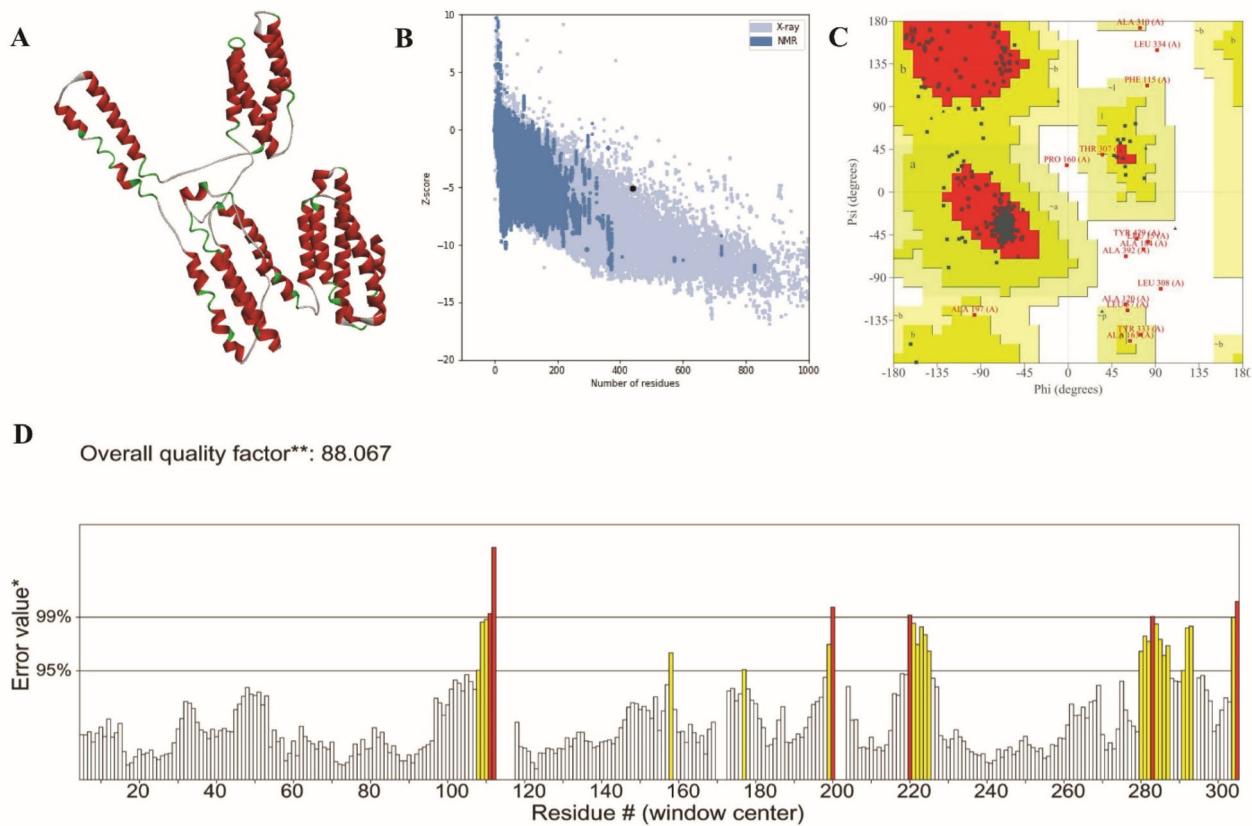


Figure 6. The tertiary structure of multi-epitope vaccine designed by I-TASSAR server and validated in different bioinformatics tools. A) The 3 D structure of final multi-epitope vaccine. B) Best refined model analyzed by ProSA server. The plot represents the Z-score of all experimentally validated similar size proteins structure from PDB database (dark blue dots: NMR spectroscopy and light blue dots: X-ray crystallography). The black spot in the plot indicates the validated model with Z-score of -4.89. C) Ramachandran plot generated by PROCHECK server and results showing the amino acid distribution in different regions. D) Overall quality factor was determined by ERRAT server. X-axis and Y-axis denotes the residues and error value, respectively. Two lines in the Y- axis represent the confidence and region above that error value considered to be lower quality.

respectively (Figure 7). Finally, interactions of the vaccine construct and TLRs are analyzed in PDBsum server. The molecular simulation study revealed that the Eigen values of the vaccine construct-TLRV4 complex were measured to be 1.82×10^{-6} (Figure 8(d)). However, the Eigen value of the protein is directly proportional to the deformability viz. The lower Eigen values are a representation of easy deformability or distortion of the candidate vaccine which ultimately depends on the distortion of each residue in the protein. The peaks corresponding to the regions of the proteins with increased deformability have been illustrated in Figure 8(d). The temperature factor or B-factor of the complex was displayed by their respective PDB field (Figure 8(c)). The normal mode analysis (NMA) and B-factors suggested that the vaccine construct is thermostable in nature (Figure 8(a, c)). The deformability value of approximately 1 was seen around the 1600th residue for the vaccine TLRV4 protein complex which is computed based on the data shown in 8 b. Individual variance is indicated by red bars, while cumulative variance is indicated by green bars in the variance graph as shown in 8e. The coupling of docked complex was analyzed by covariance matrix analysis represents uncorrelated, correlated and anti-correlated motions have presented in order of white, red and blue colors respectively have shown in Figure 8(f). The

stiffness and cumulative and individual variances of the complex has determined that have shown in Figure 8(g).

4.3.6. Codon optimization and *in silico* cloning of vaccine construct

The primary protein sequence was back-translated using the EMBOSS Backtranslseq server and optimized by the JCAT server. Initially, the codon usage of the vaccine construct was optimized based on the *E. coli* strain K12. The CAI value subsequently improved from 0.55 to 0.98; whereas, the GC content of the improved sequence was found to be 53.61. After optimization, the Ndel and Nhel restriction sites were added to the 5' and 3' end of the optimized sequence, respectively and cloned into the pET28a (+) 6x-His-TEV vector using the SnapGene tool (Figure 9).

4.3.7. Immune simulation efficiency of vaccine constructs

Activation of viral-specific CTL response is an important event that would aids in clearing intracellular infections (Woolard & Kumaraguru, 2010). Hence, the CTL effect of the multi-epitope vaccine was assessed by C-ImmSim 10.1 server. Different combinations of HLA alleles (Test set 1-Test set 4) were entered in the C-ImmSim server and details of MHC

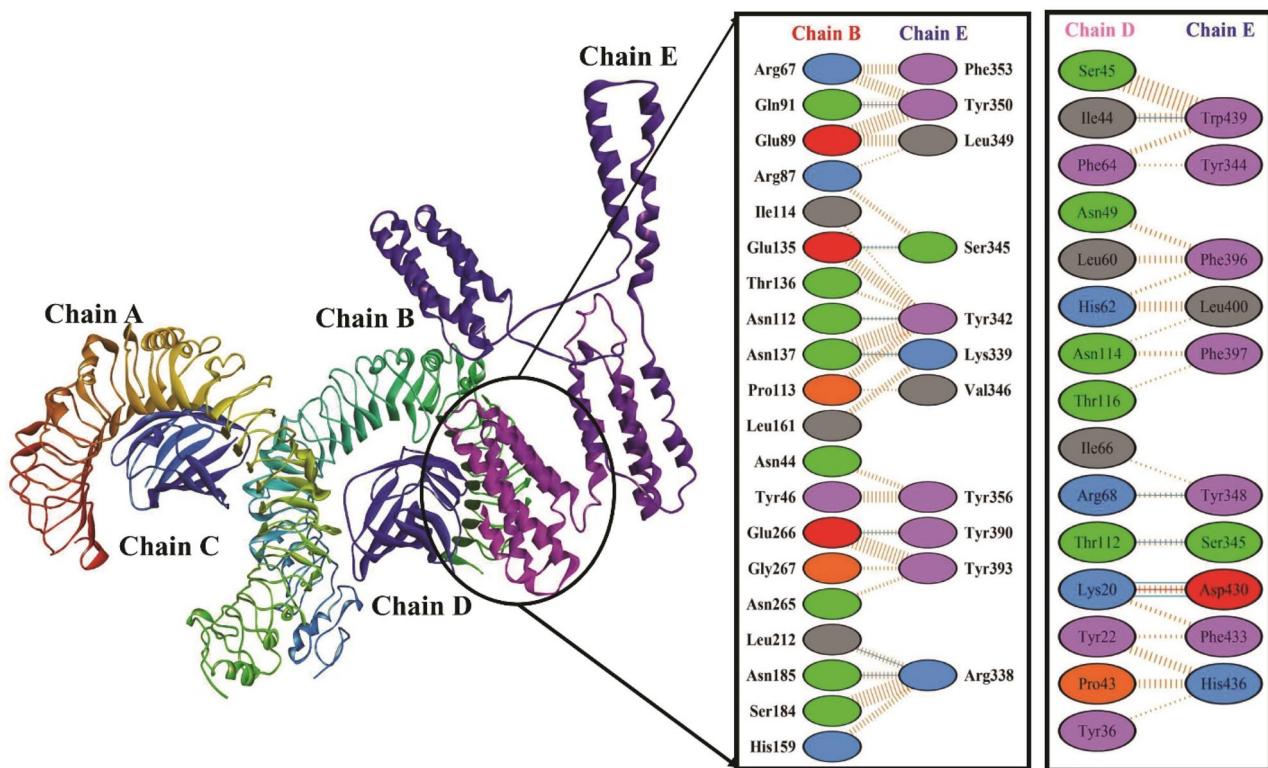


Figure 7. Multi-epitope vaccine docking with human TLR4 receptor. The colors of the interaction residues are reflecting properties of the amino acid (Positive: Blue, Negative: Red, Neutral: Green, Aliphatic: Grey, Aromatic: Purple and Pro & Gly: Orange).

test sets have been mentioned in the methodology section. The immune simulation revealed long-lasting period of active CTL and TH populations for all four test sets. Additionally, increased level of Dendritic cell (DC), Natural killer cells (NK), and Macrophage (MA) presentation was observed after exposure of vaccine [Supplementary File 8](#). Importantly, the robust level of IFN- γ and IL-2 elevation appears during secondary and tertiary vaccination periods ([Figure 10](#)), which confirms that the constructed multi-epitope vaccine is capable of eliciting a robust CTL response.

5. Discussion

Humoral and cell-mediated immune responses are considered as two arms of the immune system. Obviously, antibody response is mainly involved in neutralizing the viral infections. Several SARS-CoV2 vaccines are available in the market under emergency use authorizations and other vaccines are under the various stage of clinical trials. Most of the current vaccines were solely focused on the pre-fusion form of the SARS-CoV-2 Spike protein and their neutralizing antibody response ([Kyriakidis et al., 2021](#)). However, long-lived antibody response was observed in patients who have recovered from SARS-CoV2 infection. But, the emerging data suggest that the antibody response may be compromised due to severe viral infection. On the other hand, the Covid patients are displaying the robust CTL and Th response to both structural and non-structural viral proteins even after several weeks of disease onset ([Jarjour et al., 2021](#)). Hence, T cell response plays a vital role in viral clearances during the infection period ([Karlsson et al., 2020](#)). The Cytotoxic T

Lymphocytes (CTL) recognizes the Major Histocompatibility Complex (MHC) bounded peptides otherwise called epitopes and provokes the effector functions against the viral infected cells. Identification of such potent T cell epitopes through the traditional methods is time consuming and expensive process ([Parvizpour et al., 2020](#)). To overcome this problem, *in silico* prediction methods are widely employed to analyze the potent T cell epitopes. In this study, we first carried out different immunoinformatics analyses to predict the immunogenic CTL epitopes using a sequence-based method. Besides, class I MHC binders were predicted using the MHC-I binding predictions tool from Immune Epitope Database (IEDB). Based on the previous literature, 12 high prevalent HLA super-type alleles were selected for analyzing class I MHC binders. The selected HLA super-type almost covers 99% world population. A total of 2240 class I MHC binders were predicted that are restricted by high-frequency HLA super-types. From amongst these, 80.31% (Total: 1799) of epitopes were predicted in non-structural proteins of SARS-CoV2 virus; whereas, 19.69% (Total: 441) of epitopes were found in structural proteins. Interestingly, 245 epitopes (10.98%) are located in the SGP protein. Hence, the SGP-based epitopes predicted in this study could be the potent immunogenic target for vaccine development. Intracellular proteasomal cleavage is an important event for CTL activation ([Dönnies & Kohlbacher, 2005](#)). However, the predicted MHC binders should undergo processing by the proteasomes and transported through TAP for the formation of a peptide-MHC complex. Identification of proteasomal processing is thus a critical step prior to epitope incorporation in vaccine. The class I MHC binders using the MHC-I processing prediction

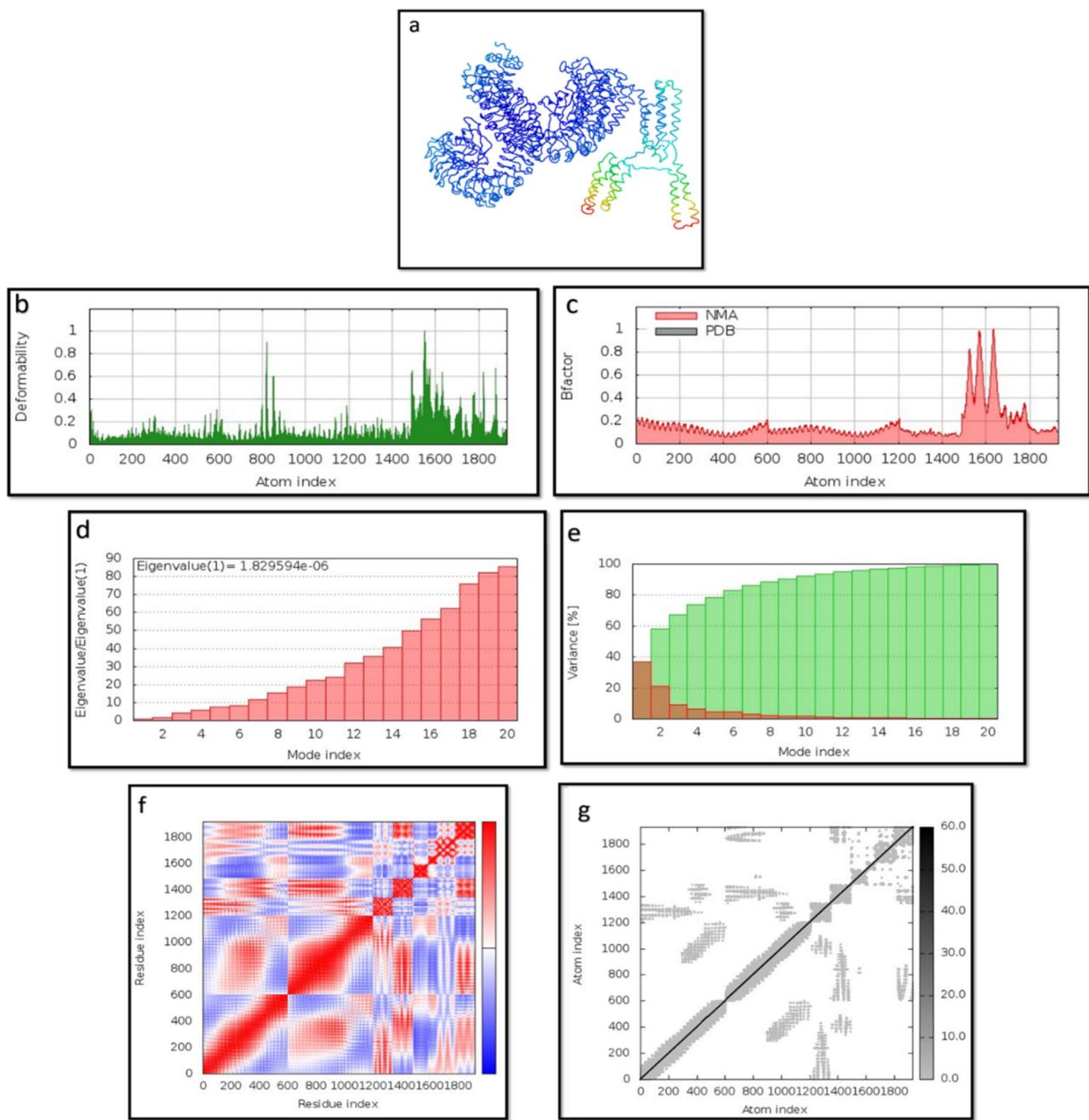


Figure 8. The results of molecular dynamics simulation of vaccine construct and TLR-V4 docked complex. a. NMA mobility and Protein domains, b. Deformability, c. B-factor, d. Eigen values, e. Variance f. Covariance matrix plot, g. Elastic network plot.

tool from IEDB 876/2240 (39.11%) epitopes received positive processing scores. From these 876 epitopes, 697 and 179 positive processing peptides are found in the non-structural and structural proteins of the SARS-CoV2 virus, respectively. However, the predicted MHC binders should be processed. Moreover, association of the peptide-MHC complex with the T Cell Receptor (TCR) is another key step in activation of effector CTL functions (Calis et al., 2013). Predicting the association capabilities of immunogenic epitopes would be very useful for multi-epitope vaccine development. A total of 462/876 epitopes with positive processing were found to have favorable association propensities. From a structural perspective, 22.51% (104 out of 462) epitopes were predicted as immunogenic in structural proteins; whereas, 77.49% (358

out of 462) epitopes were predicted in non-structural proteins.

A potent CTL epitope must exhibit antigenicity and the ability to trigger the cell-mediated immunity. The antigenicity prediction using the VaxiJen server revealed 222/462 positive immunogenic epitopes exhibiting optimal antigenicity, with the degree of antigenicity ranging from 0.40 to 2.10. Moreover, the desired vaccine candidate must be non-toxic and non-allergenic (Mohan et al., 2020). The toxicity and allergenicity analyses using ToxinPred followed by AllerTop predicted 219/222 epitopes to be nontoxic, while 121 epitopes were predicted to be non-allergic. A BLASTp query against human proteomes revealed all epitopes to have e values below 0.05, implying their inability to provoke an

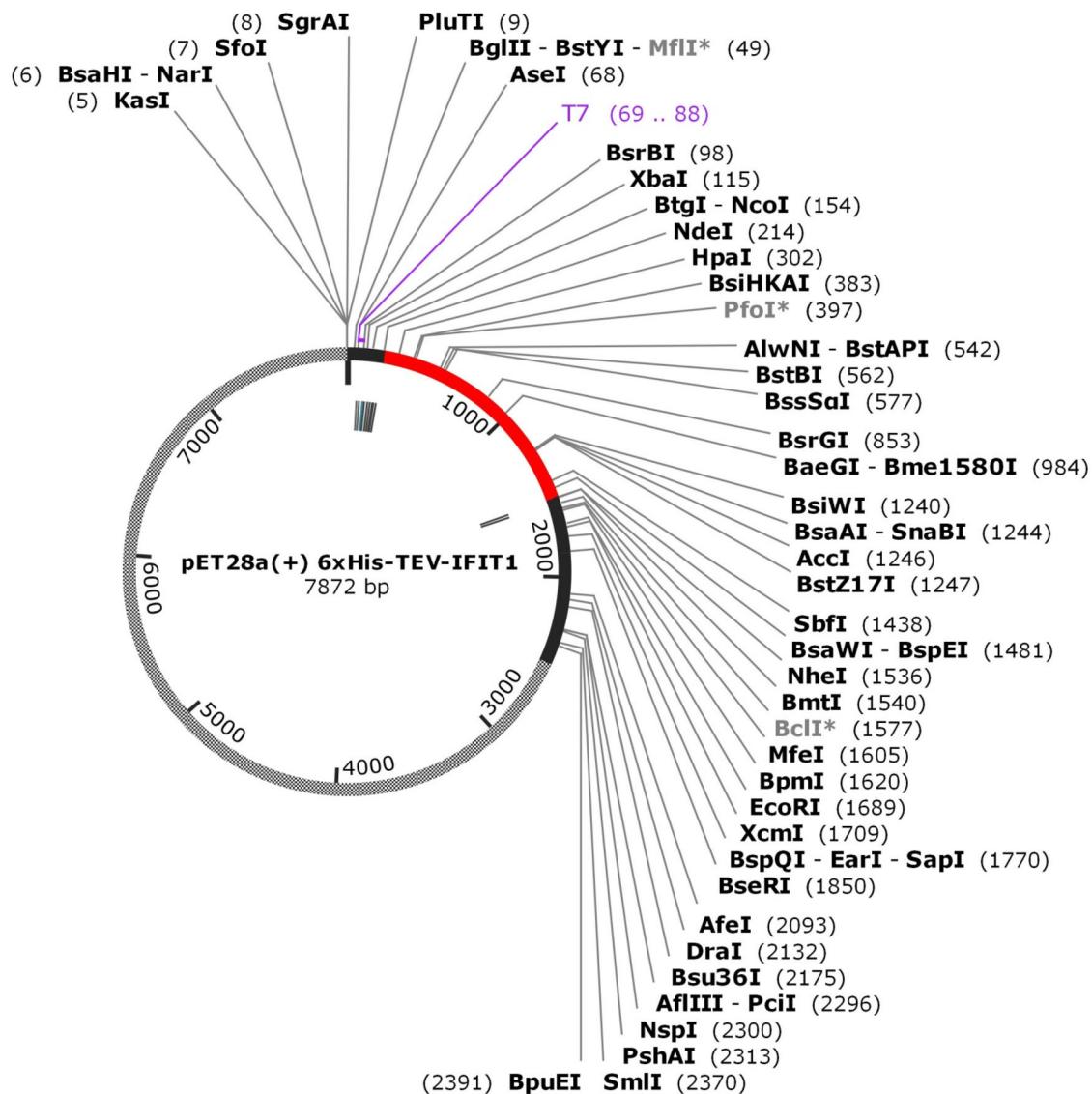


Figure 9. *In silico* cloning of final vaccine construct into the pET28a (+) 6x-His-TEV vector, where the red color represents the vaccine insert and black color represents vector backbone.

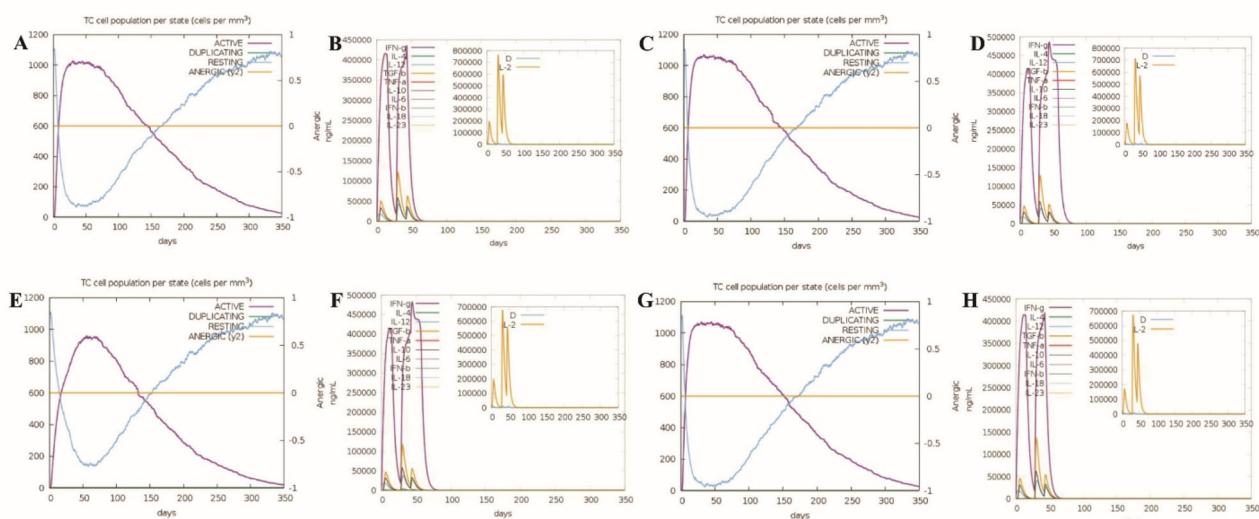


Figure 10. *In silico* immune simulation of the designed multi-CTL epitope vaccine. A, B) CTL and cytokine response of HLA-A*01:01, HLA-A*02:01, HLA-B*07:02 and HLA-B*08:01 restricted populations; C, D) HLA-A*03:01, HLA-A*24:02, HLA-B*15:01 and HLA-B*27:05 restricted populations; E, F) HLA-A*02:01, HLA-A*26:01, HLA-B*39:01 and HLA-B*40:01 restricted populations; G, H) HLA-A*01:01, HLA-A*02:01, HLA-B*07:02 and HLA-B*58:01 restricted populations.

autoimmune response. Moreover, a conservancy analyses showed that 67 CTL cell epitopes with 100% conservancy with other SARS-CoV2 variants. However, the conserved epitopes predicted in this present study could provoke the CTL immune response against multiple SARS-CoV2 variants. Interestingly, 8 epitopes (YIWLGFIAGL, KLNDLCFTNV, PYRVVVLSF, RFPNITNLCPF, FPNITNLCPF, YQPYRVVVL, YQPYRVVVL, AEIRASANL) predicted in the SGP protein exhibited 100% conservancy with other SARS-CoV2 variants. Among this, KLNDLCFTNV epitopes was experimentally confirmed to bind with HLA-A*02:01 (Poran et al., 2020), whereas, HLA-A*24:02 allele restricted epitopes RFPNITNLCPF provoke specific CD8 T cell activity (Tarde et al., 2021). Moreover, experimental evidence showed that PYRVVVLSF epitope binds with HLA-A*01:01, HLA-A*23:01, HLA-A*24:02 and HLA-A*26:01 alleles and FPNITNLCPF epitope binds with HLA-B*07:02, HLA-B*35:01, HLA-B*51:01, HLA-B*53:01 and HLA-B*54:01alleles (IEDB submission). Generally, four methods - sequence, structural, hybrid, and consensus approaches have been extensively used to predict the immunogenic epitopes. Comparatively, the combination of sequence- and structure-based (hybrid) methods were found to exhibit higher predictive accuracy than other methods (X. Yang & Yu, 2009). Hence, the epitopes were predicted by sequence-based methods and their interaction with either experimentally validated MHC or high-quality MHC homology models were extensively analyzed using molecular docking. From amongst 81 epitopes binding to their respective alleles, 13 epitopes (ORF1ab: ILHCANFNV (A02); ORF8: YIDIGNYTV (A02); MGP: SYFIASFRL (A24); EP: QWNLVIGFL (A24); MGP: LLWPVTLACF (B62); ORF1ab: FMGRIRSVY (B62); ORF1ab: KQLIKVTLVF (B62); ORF1ab: LAYILFTRF (B62); ORF1ab: GQSTQLGIEF (B62); ORF1ab: LQAVGACVL (B62); ORF8: FLGIITTVAAF (B62); SGP: YQPYRVVVL (B62); and NP: MEVTPSGTWL (B40) exhibiting the highest binding affinity (\geq 11 kcal/mol). A majority of these were restricted with the B62 supertype. In this study, we considered the binding affinity and hydrogen bonding while constructing a multi-CTL epitope vaccine. A high number of hydrogen bond interaction between the peptide and MHC were indicators of stable and high affinity binding (Hossain et al., 2021). Thus, these metrics were used to evaluate the docked peptides. In comparison to previous studies using limited HLA alleles for epitope prediction, we have for the first time utilized all 12 high prevalent HLA supertype to not only predict potent CTL epitopes but also their interaction with all high prevalence alleles. To the best of our knowledge, this is the first study to utilize the structure-based epitopes prediction using all 12 high prevalent HLA supertypes.

A comparison of the predicted epitopes against the experimentally proven ones in the IEDB database revealed a 34.57% (Total: 28/81) match. Moreover, the HLA-A*02:01 restricted ORF1ab epitope VMVELVAEL has been shown to elicit potent T cell response to HLA-A*0201 Tg mice (Takagi & Matsui, 2020). Intriguingly, 4/28 epitopes (MGP Protein-FVLAAYVRI, ORF1ab Protein- KLNVDYFV, ORF1ab Protein-ILHCANFNV, and EP Protein- LLFLAFVVFL) have multi-allele binding potential (Ishizuka et al., 2009). A total of five

epitopes (ORF1ab Protein- HLYLQYIRK, EP Protein-LWPVTLACF, EP Protein- QWNLVIGFLF, SGP Protein-FPNITNLCPF, and MGP Protein- SELVIGAVI) exhibited high binding affinity to various class I MHC molecules (IEDB submission). Moreover, 18/81 epitopes matched those peptides which have been experimentally proven to bind MHC1 molecules as deduced from ligand elution assays (Supplementary Table 7). These results strongly suggest that hybrid-based epitope prediction methods provide higher prediction efficiency over sequence-based prediction methods. The aforementioned top-ranking epitopes ($N=17$) were selected for the vaccine construct restricted by 11 HLA supertypes. The IEDB conservancy analyses revealed that 11 epitopes of vaccine construct have 100% conservancy, 3 epitopes have 90% and 3 epitopes have 80% conservancy with other SARS-CoV2 variants. Importantly, in the SGP protein, epitopes RFPNITNLCPF and YQPYRVVVL were identified with 100% conservancy. Similar values were observed for FLYIIKLIFL and LWPVTLACF (MGP), NSVLLFLAF (EP), MEVTPSGTWL (NP), WLLLTLTSL, AIVVTCCLAY, TRFFYVLGL, DVSFLAHIQW (ORF 1AB), and APFLYLYAL (ORF 3 A) epitopes. Lower conservancy values of 90% and 80% were observed in the case of VYSVIYLYLTF, DVVAIDYKHY, SEVGPEHSL (ORF 3 A) and WTAGAAAYY (SGP), and HLYLQYIRK, LQAVGACVL (ORF 1AB) peptides. In addition, incorporation of the universal PADRE epitope in the vaccine ensured activation of the CD4 T helper cells during the immune response, while ensuring robust binding to 16 high-frequency HLA-DR alleles (Alexander et al., 2000). The addition of PADRE epitope and the HBHA adjuvant at the N terminal of the vaccine construct might lead to a robust immunogenic response. The antigenicity, toxicity and allergenicity analyses of the vaccine construct showed potent antigenic properties whilst exhibiting low toxicity and hypo-allergenicity. It is known that the molecular weight of the vaccine is inversely related to the lymph node exposure. Moreover, proteins with molecular weight $>$ 16 kDa play a major role in lymph node absorption (Majee et al., 2020; Wu et al., 2020). Thus, the vaccine construct with a weight of 47.92 kDa is expected to be absorbed in the lymph nodes and exhibit good antigenicity. Based on the instability index score of 39.04 and an aliphatic index of 98.79, the vaccine construct was found to be thermostable (Instability index $<$ 40 = stable protein) (Jaydari et al., 2020). The vaccine was found to linger around for extended periods in mammalian cells (30 h), yeast ($>$ 20 h), and *E. coli* cells ($>$ 20 h). Moreover, the predicted half-life of our vaccine construct was also found to be higher (Akhand et al., 2020; Dong et al., 2020; Rahman et al., 2020) and equal in comparison to the previously reported *in silico* designed vaccines candidate (Abraham Peele et al., 2021; Ahmad et al., 2020; Ismail et al., 2020; Kar et al., 2020; Ojha et al., 2020; Saha & Prasad, 2020; Tahir Ul Qamar et al., 2020). The secondary and tertiary structure analyses of the vaccine construct revealed its high quality. The modeling of the interactions between TLR4 and the vaccine construct using the ClusPro server and the iMODS server, revealed that the construct bound with a high affinity (-14.1 kcal/mol) and established stable hydrogen bonds with TLR4. Moreover, the MD simulation of the TLR4-

vaccine complex was found to have low deformity (Figure 8(a)) and Eigen values (Figure 8(a)). The TLR and vaccine complex was also stable as evident from the B-factor analysis (Figure 8(b)). All these evidences suggest that the designed vaccine is very stables *in vivo* condition. Our vaccine construction procedure also considered the CAI value (codon optimization) in order to improve its expression in numerous expression systems. A higher CAI value is associated with a corresponding shift in the gene expression in the host system (Tosta et al., 2021). Hence, the vaccine construct exhibited the ability to be expressed in *E. coli* cells (CAI: 0.98). Moreover, cloning of the construct was performed in pET28a (+) 6x-His-TEV vector (Figure 9), that contains 6X His tag sequence and a TEV protease cleavage site conferring ease of recombinant vaccine product purification in the downstream process.

In contrast to other studies focusing on the spike glycoprotein, we utilized the entire SARS-CoV2 proteome to designed the CTL based vaccine using robust epitope prediction methods. The immune system simulation profile from C-ImmSim online server reveals that the constructed vaccine is capable of stimulating many immune cells especially CD8+ and CD4+ T cells population after vaccination. Also, increased DC and macrophage levels are observed at primary and booster dose of vaccine administration. Increased IFN- γ and IL-2 elevation is observed after immunization. To the best of our knowledge, this is the first study to determine the immune simulation profile using all 12 MHC supertypes. The present *in silico* results suggested that the vaccine construct is capable of provoking a potent CTL response that would help to combat SARS-CoV2 infection.

6. Conclusion

SARS-CoV2 is highly contagious pathogen that has resulted in a global pandemic unseen since the influenza pandemic of 1918. In this present study, potent immunogenic epitopes are identified in entire proteome of SARS-CoV2 using combination of sequence- and structure-based methods, wherein many peptides have been already experimentally validated as potent CTL epitopes. A broad gamut of prediction methods has revealed the predicted epitopes to be safe, effective and hypoallergic. With several potent CTL epitopes being used to construct the vaccine, a broad immune response is expected. This was seen in an *in silico* immune simulation study that revealed that our designed vaccine could provoke adequate level of CTL immunity to different sets of allele population covering wide swathes of global population. Further, *in vitro* and *in vivo* studies are needed to determine whether the designed vaccine could offer the protective T cell immunity to SARS-CoV2.

Acknowledgements

Authors thanks the researchers worldwide for sharing the sequential information of SARS-CoV2. This helps us for effective construction of potent vaccine candidate against the virus using various bioinformatics tools.

Disclosure statement

Authors declare no conflict of interest.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Johnson Retnaraj Samuel Selvan Christyraj  <http://orcid.org/0000-0001-7662-8150>

Manikandan Mohan  <http://orcid.org/0000-0002-0132-7510>

References

- Abraham Peele, K., Srihansa, T., Krupanidhi, S., Ayyagari, V. S., & Venkateswarulu, T. C. (2021). Design of multi-epitope vaccine candidate against SARS-CoV-2: A *in-silico* study. *Journal of Biomolecular Structure & Dynamics*, 39(10), 3793–3801.
- Ahmad, S., Navid, A., Farid, R., Abbas, G., Ahmad, F., Zaman, N., Parvaiz, N., & Azam, S. S. (2020). Design of a novel multi epitope-based vaccine for pandemic coronavirus disease (COVID-19) by vaccinomics and probable prevention strategy against avenging zoonotics. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*, 151, 105387.
- Akhand, M. R. N., Azim, K. F., Hoque, S. F., Moli, M. A., Joy, B., Das, Akter, H., Afif, I. K., Ahmed, N., & Hasan, M. (2020). Genome based evolutionary study of SARS-CoV-2 towards the prediction of epitope based chimeric vaccine. *BioRxiv*.
- Alexander, J., del Guercio, M.-F., Maewal, A., Qiao, L., Fikes, J., Chesnut, R. W., Paulson, J., Bundle, D. R., DeFrees, S., & Sette, A. (2000). Linear PADRE T helper epitope and carbohydrate B cell epitope conjugates induce specific high titer IgG antibody responses. *Journal of Immunology (Baltimore, MD: 1950)*, 164(3), 1625–1633. <https://doi.org/10.4049/jimmunol.164.3.1625>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235–242. <https://doi.org/10.1093/nar/28.1.235>
- Bui, H.-H., Sidney, J., Li, W., Fuseder, N., & Sette, A. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*, 8(1), 1–6. <https://doi.org/10.1186/1471-2105-8-361>
- Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., Keşmir, C., & Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Computational Biology*, 9(10), e1003266.
- Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S. M., & Savidge, T. C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science Advances*, 2(3), e1501240.
- Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y., & Yuen, K. Y. (2007). Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews*, 20(4), 660–694.
- Colovos, C., & Yeates, T. (1993). ERRAT: An empirical atom-based method for validating protein structures. *Protein Science*, 2(9), 1511–1519. <https://doi.org/10.1002/pro.5560020916>
- Dai, Y., Chen, H., Zhuang, S., Feng, X., Fang, Y., Tang, H., Dai, R., Tang, L., Liu, J., Ma, T., & Zhong, G. (2020). Immunodominant regions prediction of nucleocapsid protein for SARS-CoV-2 early diagnosis: A bioinformatics and immunoinformatics study. *Pathogens and Global Health*, 114(8), 463–470.
- Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., & Vajda, S. (2020). Performance and its limits in rigid body protein-protein docking. *Structure (London, England: 1993)*, 28(9), 1071–1081.
- Dhanda, S. K., Usmani, S. S., Agrawal, P., Nagpal, G., Gautam, A., & Raghava, G. P. S. (2017). Novel *in silico* tools for designing peptide-

- based subunit vaccines and immunotherapeutics. *Briefings in Bioinformatics*, 18(3), 467–478.
- Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v. 2—A server for in silico prediction of allergens. *Journal of Molecular Modeling*, 20(6), 1–6. <https://doi.org/10.1007/s00894-014-2278-5>
- Dong, R., Chu, Z., Yu, F., & Zha, Y. (2020). Contriving multi-epitope sub-unit of vaccine for COVID-19: Immunoinformatics approaches. *Frontiers in Immunology*, 11, 1784.
- Dönnes, P., & Kohlbacher, O. (2005). Integrated modeling of the major events in the MHC class I antigen processing pathway. *Protein Science: A Publication of the Protein Society*, 14(8), 2132–2140.
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8(1), 1–7. <https://doi.org/10.1186/1471-2105-8-4>
- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Eds.), *The proteomics protocols handbook*. Springer protocols handbooks (pp. 571–607). Humana Press.
- Geourjon, C., & Deleage, G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Computer Applications in the Biosciences: CABIOS*, 11(6), 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681>
- Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, 33(Web Server issue), W526–W531.
- Guo, Y.-R., Cao, Q.-D., Hong, Z.-S., Tan, Y.-Y., Chen, S.-D., Jin, H.-J., Tan, K.-S., Wang, D.-Y., & Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—an update on the status. *Military Medical Research*, 7(1), 1–10. <https://doi.org/10.1186/s40779-020-00240-0>
- He, L., & Zhu, J. (2015). Computational tools for epitope vaccine design and evaluation. *Current Opinion in Virology*, 11, 103–112.
- Hossain, M. S., Hossan, M. I., Mizan, S., Moin, A. T., Yasmin, F., Akash, A.-S., Powshri, S. N., Hasan, A. K. R., & Chowdhury, A. S. (2021). Immunoinformatics approach to designing a multi-epitope vaccine against Saint Louis Encephalitis Virus. *Informatics in Medicine Unlocked*, 22, 100500. <https://doi.org/10.1016/j.imu.2020.100500>
- Ishizuka, J., Grebe, K., Shenderov, E., Peters, B., Chen, Q., Peng, Y., Wang, L., Dong, T., Pasquetto, V., Oseroff, C., Sidney, J., Hickman, H., Cerundolo, V., Sette, A., Bennink, J. R., McMichael, A., & Yewdell, J. W. (2009). Quantitating T cell cross-reactivity for unrelated peptide antigens. *Journal of Immunology (Baltimore, MD: 1950)*, 183(7), 4337–4345.
- Islam, M. R., Hoque, M. N., Rahman, M. S., Alam, A. S. M. R. U., Akther, M., Puspo, J. A., Akter, S., Sultana, M., Crandall, K. A., & Hossain, M. A. (2020). Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Scientific Reports*, 10(1), 1–9. <https://doi.org/10.1038/s41598-020-70812-6>
- Ismail, S., Ahmad, S., & Azam, S. S. (2020). Immunoinformatics characterization of SARS-CoV-2 spike glycoprotein for prioritization of epitope based multivalent peptide vaccine. *Journal of Molecular Liquids*, 314, 113612.
- Jain, S., & Baranwal, M. (2019). Computational analysis in designing T cell epitopes enriched peptides of Ebola glycoprotein exhibiting strong binding interaction with HLA molecules. *Journal of Theoretical Biology*, 465, 34–44.
- Jarjour, N. N., Masopust, D., & Jameson, S. C. (2021). T cell memory: Understanding COVID-19. *Immunity*, 54(1), 14–18. <https://doi.org/10.1016/j.jimmuni.2020.12.009>
- Jaydari, A., Nazifi, N., & Forouharmehr, A. (2020). Computational design of a novel multi-epitope vaccine against *Coxiella burnetii*. *Human Immunology*, 81(10–11), 596–605.
- Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*, 10(1), 1–24. <https://doi.org/10.1038/s41598-020-67749-1>
- Karlsson, A. C., Humbert, M., & Buggert, M. (2020). The known unknowns of T cell immunity to COVID-19. *Science Immunology*, 5(53), eabe8063. <https://doi.org/10.1126/sciimmunol.abe8063>
- Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundsgaard, C., Sette, A., Lund, O., Bourne, P. E., Nielsen, M., & Peters, B. (2012). Immune epitope database analysis resource. *Nucleic Acids Research*, 40(Web Server issue), W525–W530.
- Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research*, 40(Web Server issue), W294–W297.
- Kyriakidis, N. C., López-Cortés, A., González, E. V., Grimaldos, A. B., & Prado, E. O. (2021). SARS-CoV-2 vaccines strategies: A comprehensive review of phase 3 candidates. *NPJ Vaccines*, 6(1), 1–17. <https://doi.org/10.1038/s41541-021-00292-w>
- Laskowski, R. A., Jabłońska, J., Pravda, L., Váreková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Science: A Publication of the Protein Society*, 27(1), 129–134.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291. <https://doi.org/10.1107/S0021889892009944>
- Long, Q.-X., Liu, B.-Z., Deng, H.-J., Wu, G.-C., Deng, K., Chen, Y.-K., Liao, P., Qiu, J.-F., Lin, Y., Cai, X.-F., Wang, D.-Q., Hu, Y., Ren, J.-H., Tang, N., Xu, Y.-Y., Yu, L.-H., Mo, Z., Gong, F., Zhang, X.-L., ... Huang, A.-L. (2020). Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nature Medicine*, 26(6), 845–848. <https://doi.org/10.1038/s41591-020-0897-1>
- López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S., & Chacón, P. (2014). iMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Research*, 42(Web Server issue), W271–W276.
- Lu, H., Stratton, C. W., & Tang, Y. (2020). Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *Journal of Medical Virology*, 92(4), 401–402.
- Magnan, C. N., Randall, A., & Baldi, P. (2009). SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics (Oxford, England)*, 25(17), 2200–2207.
- Majee, P., Jain, N., & Kumar, A. (2020). Designing of a multi-epitope vaccine candidate against Nipah virus by in silico approach: A putative prophylactic solution for the deadly virus. *Journal of Biomolecular Structure and Dynamics*, 39 (4), 1461–1480.
- Mehla, K., & Ramana, J. (2016). Identification of epitope-based peptide vaccine candidates against enterotoxigenic Escherichia coli: A comparative genomics and immunoinformatics approach. *Molecular bioSystems*, 12(3), 890–901. <https://doi.org/10.1039/c5mb00745c>
- Mohammed, A. A., ALnaby, A. M., Sabeel, S. M., AbdElmarouf, F. M., Dirar, A. I., Ali, M. M., Khandgawi, M. A., Yousif, A. M., Abdulgadir, E. M., Sabahalkhair, M. A., Abbas, A. E., & Hassan, M. A. (2018). Epitope-based peptide vaccine against fructose-bisphosphate aldolase of *Madurella mycetomatis* Using immunoinformatics approaches. *Bioinformatics and Biology Insights*, 12, 1177932218809703.
- Mohan, M., Haribalaganesh, R., Coico, R., & Sundar, K. (2018). HLA-directed bioinformatics approach for genome-wide mapping of dengue CTL epitopes. *Future Virology*, 13(5), 331–342. <https://doi.org/10.2217/fvl-2017-0157>
- Mohan, M., Shanmugaraja, P., Krishnan, R., Rajagopalan, K., & Sundar, K. (2020). In silico prediction of b-cell epitopes of dengue virus—A reverse vaccinology approach. *Journal of Applied Pharmaceutical Science*, 10(10), 77–85.
- Nguyen, T. T., Pathirana, P. N., Nguyen, T., Nguyen, Q. V. H., Bhatti, A., Nguyen, D. C., Nguyen, D. T., Nguyen, N. D., Creighton, D., & Abdelrazeq, M. (2021). Genomic mutations and changes in protein secondary structure and solvent accessibility of SARS-CoV-2 (COVID-19 virus). *Scientific Reports*, 11(1), 1–16.
- Ojha, R., Gupta, N., Naik, B., Singh, S., Verma, V. K., Prusty, D., & Prajapati, V. K. (2020). High throughput and comprehensive approach to develop multiepitope vaccine against minacious COVID-19. *European Journal of Pharmaceutical Sciences: Official Journal of the European Federation for Pharmaceutical Sciences*, 151, 105375.
- Parvizpour, S., Pourseif, M. M., Razmara, J., Rafi, M. A., & Omidi, Y. (2020). Epitope-based vaccine design: A comprehensive overview of bioinformatics approaches. *Drug Discovery Today*, 25(6), 1034–1042.
- Patronov, A., & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biology*, 3(1), 120139. <https://doi.org/10.1098/rsob.120139>
- Peng, Y., Mentzer, A. J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattaisai, W., Rostron, T., Supasa, P., Liu, C., López-Camacho, C., Slon-Campos, J., Zhao, Y., Stuart, D. I., Paesen, G. C., Grimes, J. M., Antson, A. A.,

- Bayfield, O. W., Hawkins, D. E. D. P., ... Dong, T. (2020). Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nature Immunology*, 21(11), 1336–1345.
- Peters, B., Bulik, S., Tampe, R., Van Endert, P. M., & Holzhütter, H.-G. (2003). Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *Journal of Immunology (Baltimore, MD: 1950)*, 171(4), 1741–1749.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612.
- Poran, A., Harjanto, D., Malloy, M., Arieta, C. M., Rothenberg, D. A., Lenkala, D., van Buuren, M. M., Addona, T. A., Rooney, M. S., Srinivasan, L., & Gaynor, R. B. (2020). Sequence-based prediction of SARS-CoV-2 vaccine targets using a mass spectrometry-based bioinformatics predictor identifies immunogenic T cell epitopes. *Genome Medicine*, 12(1), 1–15. <https://doi.org/10.1186/s13073-020-00767-w>
- Rahman, N., Ali, F., Basharat, Z., Shehroz, M., Khan, M. K., Jeandet, P., Nepovimova, E., Kuca, K., & Khan, H. (2020). Vaccine design from the ensemble of surface glycoprotein epitopes of SARS-CoV-2: An immunoinformatics approach. *Vaccines*, 8(3), 423. <https://doi.org/10.3390/vaccines8030423>
- Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS One*, 5(4), e9862. <https://doi.org/10.1371/journal.pone.0009862>
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in Genetics: TIG*, 16(6), 276–277.
- Roussel, Y., Giraud-Gatineau, A., Jimeno, M.-T., Rolain, J.-M., Zandotti, C., Colson, P., & Raoult, D. (2020). SARS-CoV-2: Fear versus data. *International Journal of Antimicrobial Agents*, 55(5), 105947.
- Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738.
- Saha, R., & Prasad, B. V. L. S. (2020). In silico approach for designing of a multi-epitope based vaccine against novel Coronavirus (SARS-CoV-2). *BioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.03.31.017459v1>
- Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, 24, 91–98.
- Shey, R. A., Ghogomu, S. M., Esoh, K. K., Nebangwa, N. D., Shintouo, C. M., Nongley, N. F., Asa, B. F., Ngale, F. N., Vanhamme, L., & Souopgui, J. (2019). In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports*, 9(1), 1–18. <https://doi.org/10.1038/s41598-019-40833-x>
- Tahir Ul Qamar, M., Rehman, A., Tusleem, K., Ashfaq, U. A., Qasim, M., Zhu, X., Fatima, I., Shahid, F., & Chen, L.-L. (2020). Designing of a next generation multiepitope based vaccine (MEV) against SARS-CoV-2: Immunoinformatics and in silico approaches. *PLoS One*, 15(12), e0244176.
- Takagi, A., & Matsui, M. (2020). Identification of HLA-A* 02: 01-restricted candidate epitopes derived from the nonstructural polyprotein 1a of SARS-CoV-2 that may be natural targets of CD8+ T cell recognition *in vivo*. *Journal of Virology*, 95(5), e01837-20.
- Tarke, A., Sidney, J., Kidd, C. K., Dan, J. M., Ramirez, S. I., Yu, E. D., Mateus, J., da Silva Antunes, R., Moore, E., Rubiro, P., Methot, N., Phillips, E., Mallal, S., Frazier, A., Rawlings, S. A., Greenbaum, J. A., Peters, B., Smith, D. M., Crotty, S., ... Sette, A. (2021). Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Reports Medicine*, 2(2), 100204. <https://doi.org/10.1016/j.xcrm.2021.100204>
- Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., Kloetzel, P.-M., Rammensee, H.-G., Schild, H., & Holzhütter, H.-G. (2005). Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cellular and Molecular Life Sciences: CMSL*, 62(9), 1025–1037.
- Thevarajan, I., Nguyen, T. H. O., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C. E., Jia, X., Nicholson, S., Catton, M., Cowie, B., Tong, S. Y. C., Lewin, S. R., & Kedzierska, K. (2020). Breadth of concomitant immune responses prior to patient recovery: A case report of non-severe COVID-19. *Nature Medicine*, 26(4), 453–455.
- Tomar, N., & De, R. K. (2010). Immunoinformatics: An integrated scenario. *Immunology*, 131(2), 153–168. <https://doi.org/10.1111/j.1365-2567.2010.03330.x>
- Tosta, S. F. O., Passos, M. S., Kato, R., Salgado, Á., Xavier, J., Jaiswal, A. K., Soares, S. C., Azevedo, V., Giovanetti, M., & Tiwari, S. (2021). Multi-epitope based vaccine against yellow fever virus applying immunoinformatics approaches. *Journal of Biomolecular Structure and Dynamics*, 39(1), 219–235. <https://doi.org/10.1080/07391102.2019.1707120>
- Trolle, T., McMurtrey, C. P., Sidney, J., Bardet, W., Osborn, S. C., Kaever, T., Sette, A., Hildebrand, W. H., Nielsen, M., & Peters, B. (2016). The length distribution of class I-restricted T cell epitopes is determined by both peptide supply and MHC allele-specific binding preference. *Journal of Immunology*, 196(4), 1480–1487. <https://doi.org/10.4049/jimmunol.1501721>
- Trott, O., & Olson, A. J. (2010). 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.
- Vangone, A., Schaarschmidt, J., Koukos, P., Geng, C., Citro, N., Trellet, M. E., Xue, L. C., & Bonvin, A. M. J. J. (2019). Large-scale prediction of binding affinity in protein-small ligand complexes: The PRODIGY-LIG web server. *Bioinformatics (Oxford, England)*, 35(9), 1585–1587. <https://doi.org/10.1093/bioinformatics/bty816>
- Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223), 470–473. [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303.
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(Web Server issue), W407–W410.
- Wong, G., Bi, Y.-H., Wang, Q.-H., Chen, X.-W., Zhang, Z.-G., & Yao, Y.-G. (2020). Zoonotic origins of human coronavirus 2019 (HCoV-19/SARS-CoV-2): Why is this work important? *Zoological Research*, 41(3), 213–219.
- Woolard, S. N., & Kumaraguru, U. (2010). Viral vaccines and CTL response. *Journal of Biomedicine and Biotechnology*, 2010, 1–6. <https://doi.org/10.1155/2010/141657>
- Wu, D., Wu, T., Liu, Q., & Yang, Z. (2020). The SARS-CoV-2 outbreak: What we know. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, 94, 44–48. <https://doi.org/10.1016/j.ijid.2020.03.004>
- Yang, X., & Yu, X. (2009). An introduction to epitope prediction methods and software. *Reviews in Medical Virology*, 19(2), 77–96. <https://doi.org/10.1002/rmv.602>
- Yang, H.-C., Chen, C., Wang, J.-H., Liao, H.-C., Yang, C.-T., Chen, C.-W., Lin, Y.-C., Kao, C.-H., & Liao, J. C. (2020). Genomic, geographic and temporal distributions of SARS-CoV-2 mutations. *BioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.04.22.055863v1>
- Zhang, T., Wu, Q., & Zhang, Z. (2020). Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. *Current Biology : CB*, 30(7), 1346–1351. <https://doi.org/10.1016/j.cub.2020.03.022>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 588(7836), E6. <https://doi.org/10.1038/s41586-020-2951-z>
- 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, G. F., & Tan, W. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 382(8), 727–733. <https://doi.org/10.1056/NEJMoa2001017>