

Computationally validated SARS-CoV-2 CTL and HTL Multi-Patch vaccines, designed by reverse epitomics approach, show potential to cover large ethnically distributed human population worldwide

Sukrit Srivastava^{a,b} , Sonia Verma^c, Mohit Kamthania^b , Deepa Agarwal^b, Ajay Kumar Saxena^a , Michael Kolbe^{d,e} , Sarman Singh^f , Ashwin Kotnis^g, Brijesh Rathi^h, Seema A. Nayar^{i,j}, Ho-Joon Shin^k, Kapil Vashisht^c and Kailash C. Pandey^c

^aMolecular Medicine Lab., School of Life Science, Jawaharlal Nehru University, New Delhi, India; ^bInfection Biology Group, Indian Foundation for Fundamental Research, RaeBareli, India; ^cParasite-Host Biology Group, Protein Biochemistry & Engineering Lab, ICMR-National Institute of Malaria Research, New Delhi, India; ^dDepartment for Structural Infection Biology, Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research, Hamburg, Germany; ^eFaculty of Mathematics, Informatics and Natural Sciences, University of Hamburg, Hamburg, Germany; ^fDepartment of Microbiology, All India Institute of Medical Sciences (AIIMS), Bhopal, India; ^gDepartment of Biochemistry, All India Institute of Medical Sciences (AIIMS), Bhopal, India; ^hLaboratory For Translational Chemistry and Drug Discovery, Hansraj College, University of Delhi, New Delhi, India; ⁱDepartment of Microbiology, Government Medical College, Trivandrum, India; ^jDepartment of Microbiology, Sree Gokulam Medical College, Trivandrum, India; ^kDepartment of Microbiology, School of Medicine, Ajou University, Suwon, South Korea

Communicated by Ramaswamy H. Sarma

ABSTRACT

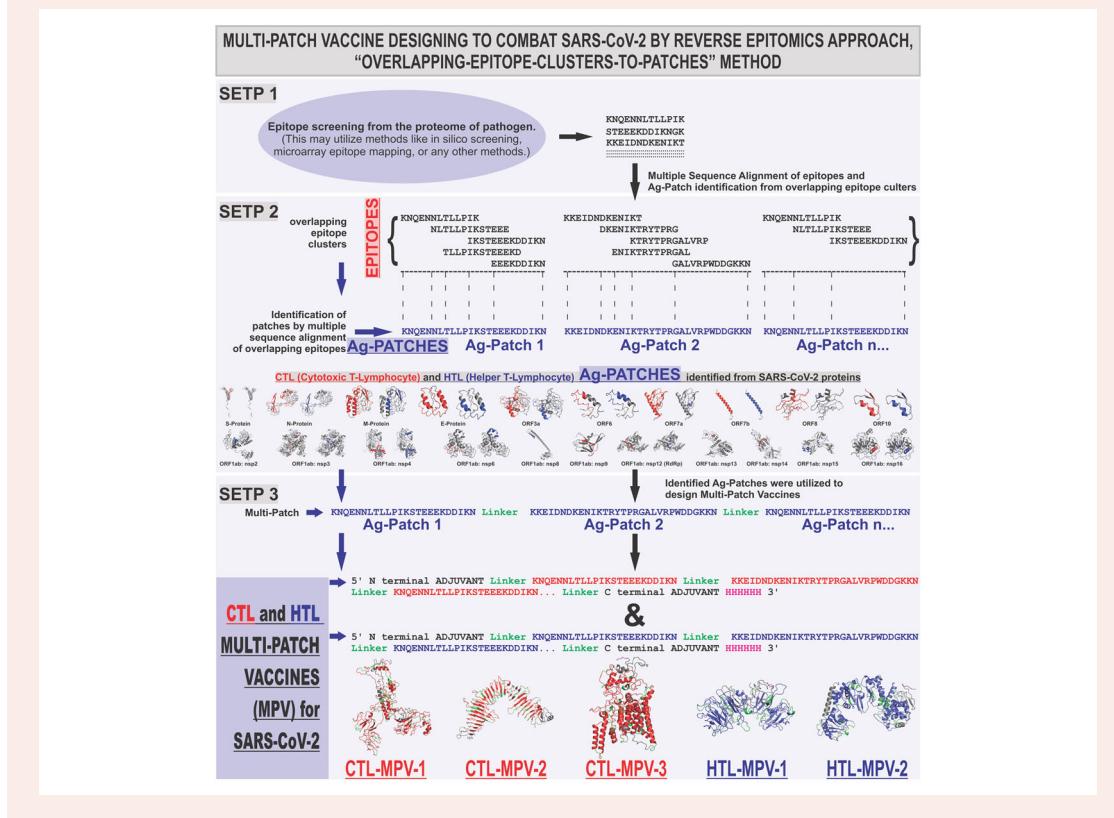
The SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) is responsible for the COVID-19 outbreak. The highly contagious COVID-19 disease has spread to 216 countries in less than six months. Though several vaccine candidates are being claimed, an effective vaccine is yet to come. A novel reverse epitomics approach, 'overlapping-epitope-clusters-to-patches' method is utilized to identify the antigenic regions from the SARS-CoV-2 proteome. These antigenic regions are named as 'Ag-Patch or Ag-Patches', for Antigenic Patch or Patches. The identification of Ag-Patches is based on the clusters of overlapping epitopes rising from SARS-CoV-2 proteins. Further, we have utilized the identified Ag-Patches to design Multi-Patch Vaccines (MPVs), proposing a novel method for the vaccine design. The designed MPVs were analyzed for immunologically crucial parameters, physiochemical properties and cDNA constructs. We identified 73 CTL (Cytotoxic T-Lymphocyte) and 49 HTL (Helper T-Lymphocyte) novel Ag-Patches from the proteome of SARS-CoV-2. The identified Ag-Patches utilized to design MPVs cover 768 overlapping epitopes targeting 55 different HLA alleles leading to 99.98% of world human population coverage. The MPVs and Toll-Like Receptor ectodomain complex shows stable complex formation tendency. Further, the cDNA analysis favors high expression of the MPVs constructs in a human cell line. We identified highly immunogenic novel Ag-Patches from the entire proteome of SARS CoV-2 by a novel reverse epitomics approach and utilized them to design MPVs. We conclude that the novel MPVs could be a highly potential novel approach to combat SARS-CoV-2, with greater effectiveness, high specificity and large human population coverage worldwide.

ARTICLE HISTORY

Received 1 October 2020
Accepted 13 October 2020

KEYWORDS

COVID-19; Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2); Coronavirus; epitope; Ag-Patch (antigenic patch); reverse epitomics; overlapping-epitope-clusters-to-patches; Multi-Patch Vaccine; Multi-Epitope Vaccine; Toll-Like Receptor (TLR)



Introduction

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), a novel coronavirus, causes the highly communicable disease COVID-19, the ongoing 2019–2020 disease outbreak, which has turned into a pandemic. The disease has already spread to more than 216 countries and territories worldwide. Above 26,415,380 people have already been infected and 870,286 died worldwide, with a mortality rate of ~3.295% (cases/deaths*100) (WHO Coronavirus Disease (COVID-19) Dashboard; as on 5th September, 2020; WHO Weekly Epidemiological Record (WER); as on 4th September, 2020). The outbreak demands an urgent requirement of a specific and an efficient vaccine against the SARS-CoV-2 virus.

The long-term adaptive immunity crucially involves the presentation of antigen as epitope on the surface of antigen-presenting cells (APC). The proteolytic chop down processing of antigen by proteasome and lysosome into small size peptide epitopes of different lengths like (7,9,13,15 ... amino acid) paves the way for epitope formation and eventual presentation. The 'Transporter associated with antigen processing' (TAP) and further the HLA allele (human leukocyte antigen) molecules facilitate the epitope presentation (Figure 1(A)). The crucial step in the process of antigen presentation is the cleavage of the antigenic protein molecules to provide small length peptides, which act as epitopes at later stage after their presentation on the surface of APC (Antoniou et al., 2003; Kahan, 2003; Melief, 2005; Tenzer et al., 2005).

In this study, we have performed the reverse epitomics analysis to identify antigenic patches from SARS-CoV-2

proteins by novel 'overlapping-epitope-clusters-to-patches' method. The term 'reverse epitomics' is coined here for first time as an approach which would involve identification of antigenic patches from pathogen's protein on the basis of overlapping multiple epitope clusters. This method of identifying antigenic patches (Ag-Patches) from a protein is termed here as 'Overlapping-epitope-clusters-to-patches' method. These antigenic patches are termed here as Ag-Patch or Ag-Patches. These Ag-Patches, once administered as vaccine candidates, are expected to be chopped down to produce cluster of overlapping multiple epitope. We have identified 73 CTL (Cytotoxic T-Lymphocyte) and 49 HTL (Helper T-Lymphocyte) such Ag-Patches from the structural and non-structural proteins of SARS-CoV-2 virus proteome. Next, we utilized these identified Ag-Patches from SARS-CoV-2 proteome to design Multi-Patch Vaccine (MPV) candidates against SARS-CoV-2.

The designed MPVs are supposed to be more effective than both the subunits or single antigenic protein-based vaccines and the Multi-Epitope Vaccines (MEVs). MPVs would involve Ag-Patches from entire proteome of SARS-CoV-2 and hence would be more effective than subunit or single protein vaccine. In comparison with the epitope-based MEVs, the Ag-Patches-based MPVs will provide a more natural antigenic vaccine candidate. The proteasome and lysosomal chop down processing of MPVs by the APC would release larger number of epitopes in their intact form for which the MEVs would have a challenge. Moreover, the Ag-Patch-based MPVs will cover a larger number of overlapping epitopes in comparison with the

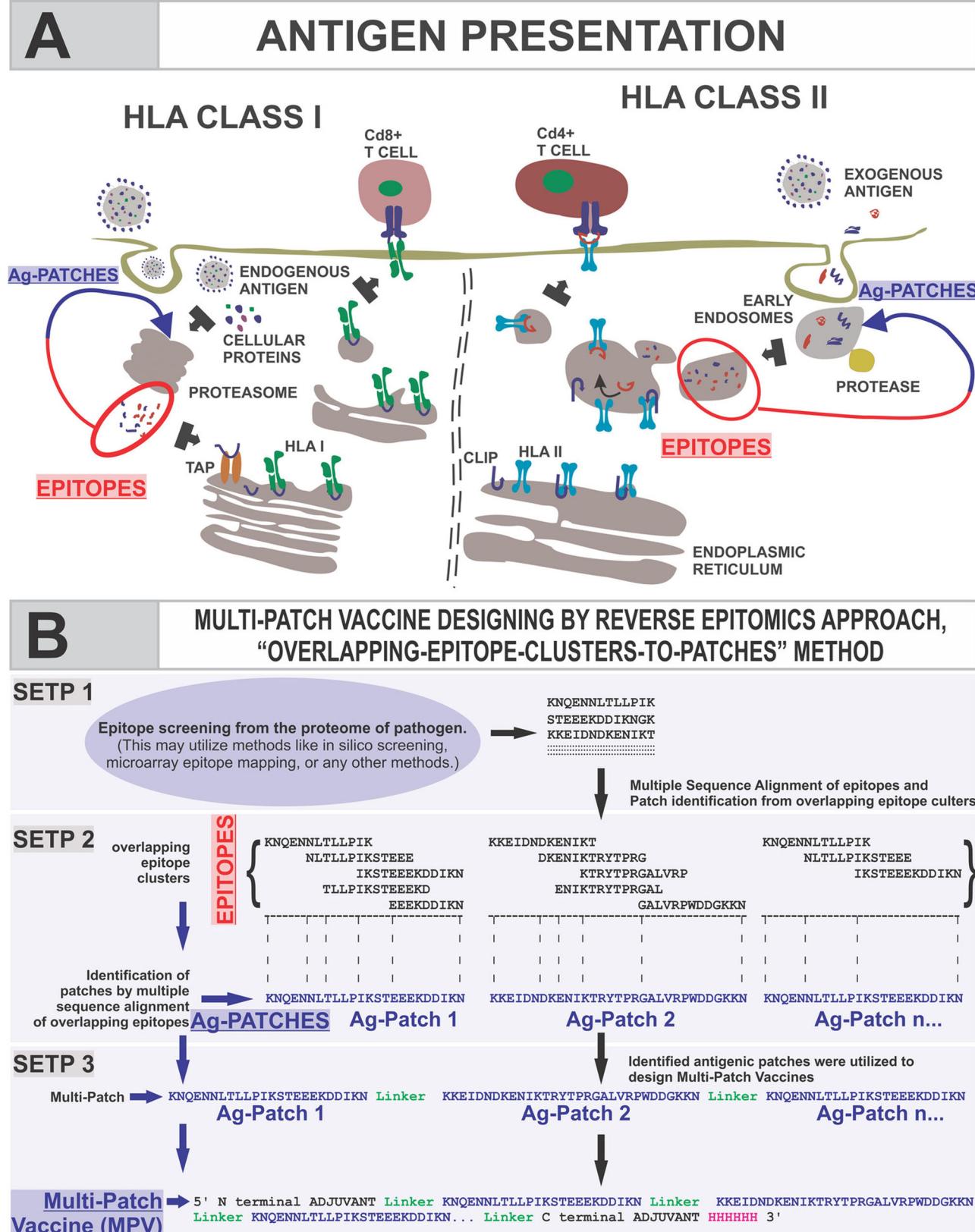


Figure 1. Schematic presentation for reverse epitomics approach, 'overlapping-eptope-clusters-to-patches' method. (A) Schematic presentation of the process of epitope presentation by the antigen-presenting cells facilitated by TAP transporter and HLA classes I & II alleles; (B) presentation of 'overlapping-eptope-clusters-to-patches' method (Steps 1 & 2) to identify antigenic patches (Ag-Patches) from SARS-CoV-2/pathogen. The cluster of overlapping epitopes leads us to identify antigenic patches from the source protein, a reverse epitomics approach. The identified antigenic patches could be used to design Multi-Patch Vaccines as shown in Step 3.

limited number of epitopes covered by the epitope-based MEVs (Srivastava et al., 2020a, 2020b). Larger the number of epitopes covered, larger would be the number of HLA

alleles targeted and thus, larger human population with ethnic distribution of HLA alleles will be covered. Therefore, the MPVs would provide more effective and specific

solution with larger human population coverage for vaccine development against SARS-CoV-2. The schematic of antigen presentation, the identification of Ag-Patches by the reverse epitomics ‘Overlapping-epitope-clusters-to-patches’ method and MPV designing is shown in Figure 1.

Methodology

Herein, we have identified antigenic patches from all the eleven ORF proteins of SARS-CoV-2 proteome (Supplementary Figure S1) (NCBI: SARS-CoV-2 isolate Wuhan-Hu-1, complete genome; gi|1798174254|ref|NC_045512.2|). These Ag-Patches are identified by the overlapping epitope clusters. These identified Ag-Patches are further utilized to design a MPV candidate against the SARS-CoV-2 infection. To identify the Ag-Patches from the SARS-CoV-2 proteins, we have first screened high scoring CTL and HTL epitopes from all the eleven ORF proteins of the virus. We further analyzed the immunogenic properties of the screened CTL and HTL epitopes on the basis of their amino acid sequences and epitope databases. Multiple sequence alignment (MSA) was performed for all the screened epitopes leading us to obtain clusters of overlapping epitopes from the proteins of SARS-CoV-2. These CTL and HTL epitope clusters further led us to identify the Ag-Patches of the SARS-CoV-2 proteins from where the overlapping epitopes clusters were arising. The identified Ag-Patches, in other words, act as the source of multiple CTL and HTL epitopes and hence are the potential stretch of SARS-CoV-2 proteins, which can be utilized to elicit T-cell immune response. The pairs of overlapping epitope forming clusters and their HLA allele binders were also utilized to analyze the world population coverage by the screened epitopes on the basis of HLA allele distributed worldwide among different ethnic human population.

Next, the identified Ag-Patches were implemented to design chimeric fusion MPVs against SARS-CoV-2. The human β defensin 2 and 3 (hBD-2 and hBD-3) were utilized as adjuvants at N and C terminals of MPVs, respectively. Flexible and rigid linkers GGGGS and EAAAK, respectively, were utilized to design the chimeric MPVs (Biragyn et al., 2001; Chen et al., 2013; Duits et al., 2003; Hajighahramani et al., 2017; Hoover et al., 2000; Hu et al., 2004; Kohlgraf et al., 2010; Srivastava et al., 2018; 2019; Wilson et al., 2013; Yang et al., 2002). Further, discontinuous B-cell epitopes, linear B-cell epitopes, and IFN- γ -inducing epitopes were also screened from the generated tertiary models of all the MPVs, designed by the identified Ag-Patches of SARS-CoV-2.

Next, we tested all the CTL and HTL MPV models for their molecular interaction with the ectodomain of the Toll-Like Receptor 3 (TLR3) by molecular docking and molecular dynamics (MD) simulation studies (Delneste et al., 2007; Farina et al., 2005; Totura et al., 2015). Further, the codon-optimized cDNA of all the CTL and HTL MPVs was also analyzed and found to favor high expression in human cell line. The Schematic work flowchart and the methodology utilized in this study are shown in Supplementary Figure S2.

Screening of potential epitopes from SARS-CoV-2 proteome

Screening of CTL epitope. The screening of CTL epitopes was performed by the IEDB (Immune Epitope Database) tools ‘MHC-I Binding Predictions’ (<http://tools.iedb.org/mhci/>) and ‘MHC-I Processing Predictions’ (Hoof et al., 2009; Peters et al., 2003; Tenzer et al., 2005). The tools use six different methods (viz. Consensus, NN-align, SMM-align, Combinatorial library, Sturniolo and NetMHCIpan) and generate ‘Percentile rank’ and a ‘total score’, respectively, indicating the immunogenic potential of the screened epitopes. The screening methods utilized are based on the total amount of cleavage sites in the submitted protein sequence. The TAP score of a peptide estimates an effective $-\log$ of (IC₅₀) values (half-maximal inhibitory concentration) for binding to TAP (Transporter associate with Antigen Processing) or its N-terminal prolonged precursors. The MHC (major histocompatibility complex) binding prediction score of a peptide is the $-\log$ of (IC₅₀) values for binding to MHC molecule (Calis et al., 2013). The IC₅₀ (nM) value for each epitope and MHC allele binding pairs were also obtained by the IEDB tool. Epitopes having high, intermediate and least affinity of binding to their respective HLA allele binders have IC₅₀ values of < 50 nM, < 500 nM and < 5000 nM, respectively.

Immunogenicity of all the screened CTL epitopes was also obtained by using ‘MHC I Immunogenicity’ tool of IEDB with all the parameters set to default analyzing 1st, 2nd and C-terminus amino acids of the given epitope (Calis et al., 2013). The tool predicts the immunogenicity of a given peptide and MHC (pMHC) complex based on the physicochemical properties of constituting amino acid and their position within the peptide sequence.

Screening of HTL epitopes. To screen out the HTL epitopes from SARS-CoV-2 proteins, the IEDB tool ‘MHC-II Binding Predictions’ was used. The tool generates ‘Percentile Rank’ for each potential screened peptide. The lower the value of percentile, the higher would be the immunogenic potential of the peptide. This percentile score is generated by the combination of three different methods, which includes combinatorial library, SMM_align and Sturniolo method. The generated percentile score is further compared with the other random five million 15-mer peptides of SWISSPROT database (Nielsen et al., 2007; Sidney et al., 2008; Sturniolo et al., 1999; Wang et al., 2010). The percentile rank of the screened peptides is generated from the consensus of all above-mentioned methods by the median percentile rank.

CTL and HTL epitope toxicity prediction. The tool ToxinPred was used to analyze the toxicity of screened CTL and HTL epitopes. The tool facilitates to identify the highly toxic or non-toxic short peptides. The toxicity check analysis was done by the ‘SVM (Swiss-Prot) based’ (support vector machine) method utilizing a dataset of 1805 sequences as positive (toxic) and 3593 sequences as negative (nontoxic) from Swissprot along with a comparison to an alternative dataset comprising of the same 1805 positive sequences and 12,541 other negative sequences from TrEMBLE (Gupta et al., 2013).

Overlapping CTL and HTL epitope clusters to Ag-Patches

CTL and HTL multi-epitope cluster-based Ag-Patch identification. To identify the potentially immunogenic Ag-Patches from SARS-CoV-2 proteome, all the screened high-scoring epitopes were aligned against their respective source protein sequence by MSA tool Clustal Omega available on the EBI server (Sievers et al., 2011). The Patches of the SARS-CoV-2 protein sequences showing consensus with the clusters of overlapping epitopes were chosen and shortlisted as Ag-Patches. This approach of search and identification of antigenic patches from source protein in a reverse epitomics manner, i.e. from epitopes to antigenic patches of source protein, is here defined as ‘Overlapping-epitope-clusters-to-patches’ method (Figure 1). The here provided reverse epitomics approach, ‘Overlapping-epitope-clusters-to-patches’ method to identify Ag-Patches from pathogen’s (here SARS-CoV-2) protein and further method to design MPV is included in filed Patent No: 202011037585 (Srivastava, S., 2020c).

Population Coverage by the overlapping CTL and HTL epitopes. The ‘Population Coverage’ tool of IEDB was used to analyze the world human population coverage for both the CTL and HTL overlapping epitopes and their respective HLA allele-binding pairs (Bui et al., 2006). The T-cells recognize the complex of a specific MHC molecule with a particular pathogen-derived epitope. The given epitope will elicit an immune response only in an individual that expresses the epitope binding the MHC molecule. This denomination of the MHC restricted T-cell responses, and the MHC polymorphism among human population provides the basis for population coverage analysis. The MHC types are expressed at dramatically different frequencies in the different ethnicities of human population worldwide. In this way, the population coverage by an epitope–MHC pair could be determined (Sturniolo et al., 1999).

Conservation analysis of antigenic patches. The shortlisted CTL and HTL epitope cluster-based Ag-Patches identified from 11 SARS-CoV-2 ORF proteins were analyzed for their amino acid sequence conservancy by ‘Epitope Conservancy Analysis’ tool of IEDB. The epitope conservancy is the percentage of SARS-CoV-2 ORF protein sequences (retrieved from NCBI) containing the particular epitope cluster-based Ag-Patch sequence in 100% amino acid sequence match. The analysis was done for all the identified Ag-Patches against all the source ORF protein sequences of SARS-CoV-2 proteins retrieved from the NCBI protein database (Bui et al., 2007).

Multi-Patch Vaccines

The identified Ag-Patches from the SARS-CoV-2 proteome were further utilized to design three CTL and two HTL MPVs as explained in the result section.

Physicochemical property analysis of designed MPVs. The physicochemical properties of the amino acid sequences of the designed three CTL and two HTL MPVs were analyzed by the ProtParam tool (Gasteiger et al., 2005). The ProtParam

tool performs an empirical investigation for the given query amino acid sequence of a protein. ProtParam computes various physicochemical parameters including amino acid length, molecular weight, theoretical pI, expected half-life (in *E. coli*, Yeast and Mammalian cell), Aliphatic index, Grand average of hydropathicity (GRAVY) and instability index score. The aliphatic index and grand average of hydropathicity (GRAVY) indicate the globular and hydrophilic nature of the protein. The instability index score indicates the stable nature of the protein molecule.

Interferon-gamma-inducing epitope prediction from the MPVs. From the designed amino acid sequence of all the three CTL and two HTL MPVs, the potential interferon-gamma (IFN- γ) epitopes were screened by utilizing the ‘IFN epitope’ tool with ‘Motif and SVM hybrid’ method (MERCI: Motif-EmeRging and with Classes-Identification, and SVM: support vector machine). The tool predicts peptides from the given protein sequences having the potential to induce release of IFN-gamma from CD4+ T cells. This module generates the overlapping, IFN-gamma-inducing peptides sequences from the given query sequence. The prediction is based on the IEDB database with 3705 IFN-gamma-inducing and 6728 non-inducing MHC class II binding epitopes (Dhanda et al., 2013; Nagpal et al., 2015).

MPVs allergenicity and antigenicity prediction. All the designed CTL and HTL MPVs were further analyzed for allergenicity and antigenicity prediction by utilizing the AlgPred and the Vaxigen tools, respectively (Doytchinova & Flower, 2007; Saha & Raghava, 2006). The AlgPred prediction is based on the similarity with already known allergens and non-allergens database, with any region of the submitted protein. For the screening of allergenicity, the Swiss-prot dataset consisting of 10,1725 non-allergens and 323 allergens is utilized. The VaxiJen tool utilizes an alignment-free approach, solely based on the physicochemical properties of the given query protein sequence, to predict antigenicity of a give protein sequence. For the prediction of antigenicity, the bacterial, viral and the tumor protein datasets are utilized by VaxiJen to predict the whole protein antigenicity.

Tertiary structure modelling and refinement of MPVs. The tertiary structure of all the designed three CTL and two HTL MPVs was generated by homology modelling utilizing the I-TASSER modelling tool. The I-TASSER is a tool that utilizes the sequence-to-structure-to-function paradigm for protein structure prediction (Roy et al., 2010). The tool generates three-dimensional (3D) atomic models from multiple threading alignments and iterative structural assembly simulations for the submitted amino acid sequence of the protein. It works based on the structure templates identified by LOMETS, a meta-server, from the PDB library. The I-TASSER tool only uses the templates of the highest Z-score, which is the difference between the raw and average scores in the unit of standard deviation with reference to the templet protein structure. For each target model, the I-TASSER simulations generate a large ensemble of the structural conformations, known as decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pairwise structure similarity and reports up to

five top-scoring models. The normalized Z-score >1 indicates a good alignment. The Cov score represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the amino acid length of the query protein sequence. Ranking of templet proteins is based on TM-score of the structural alignment between the query structure model and the known templet protein structures. The root-mean-square deviation (RMSD) is the deviation between templet residues and query residues that are structurally aligned by TM-align.

The refinement of all the generated three CTL and two HTL MPV models were performed by ModRefiner and GalaxyRefine tools (Dong et al., 2011; Ko et al., 2012). The TM-score generated by ModRefiner indicates the structural similarity of the refined model with the original input protein structure. Closer the TM-score to 1, higher would be the similarity of original and the refined model. The RMSD value of the refined model shows the conformational deviation from the initial input protein model.

The GalaxyRefine tool refines the input tertiary structure by repeated structure perturbation as well as by subsequent structural relaxation and MD simulation. The tool GalaxyRefine generates reliable core structures from multiple templates and then re-builds the loops or termini by using an optimization-based refinement method (Ko et al., 2012; Wang & Xu, 2013). To avoid any breaks or gaps in the generated 3D model, the GalaxyRefine uses the triaxial loop closure method. The MolProbity score generated for a given refined model indicates the log-weighted combination score of the clash score, the percentage of Ramachandran not favored residues and the percentage of bad side-chain rotamers.

Validation of CTL and HTL MPVs refined models. Both the refined three CTL and two HTL MPV tertiary models were further validated by RAMPAGE analysis tool (Lovell et al., 2003; Ramakrishnan & Ramachandran, 1965). The generated Ramachandran plots for all the MPV models show the sterically allowed and disallowed residues along with their dihedral psi (ψ) and phi (ϕ) angles.

Linear and discontinuous B-cell epitope prediction from the MPVs. The IEDB tool, ElliPro (ElliPro: Antibody Epitope Prediction tool) available at IEDB server was used to screen the linear and the discontinuous B-cell epitopes from all the CTL and HTL MPVs tertiary structure models. The ElliPro method screened B-cell epitope from the given tertiary structure based on the location of residues in the protein 3D structure. The farthest residue to be considered was limited to 6 Å. The residues lying outside of an ellipsoid cover 90% of the inner core residues of the protein score highest Protrusion Index (PI) of 0.9, and so on. The discontinuous epitopes predicted by the ElliPro tool are clustered based on the distance 'R' in Å between two residues centers of mass lying outside of the largest possible ellipsoid of the protein tertiary structure. The larger the value of R, the greater would be the distance between the residues (residue discontinuity) of the screened discontinuous epitopes (Kringelum et al., 2012; Ponomarenko et al., 2008).

Molecular interaction analysis of MPVs and immune receptor complexes. Molecular interaction analysis of all the three CTL and two HTL MPVs with TLR3 was performed by molecular docking followed by a MD simulation study. Protein–protein molecular docking was performed by the tool PatchDock (Schneidman-Duhovny et al., 2005). The PatchDock tool utilizes an algorithm for unbound (mimicking real-world environment) docking of molecules for protein–protein complex formation (Duhovny et al., 2002). The algorithm carries out the rigid protein–protein docking with the surface variability/flexibility implicitly addressed through liberal intermolecular penetrations. The algorithm focuses on the (a) initial molecular surface fitting on localized and curvature-based patches of the protein surface, (b) use of Geometric Hashing and Pose Clustering for initial transformation detection, (c) computation of shape complementarity utilizing the Distance Transformation, (d) efficient steric clash detection and geometric fit score based on a multi-resolution shape representation and (e) utilization of the biological information by focusing on the hot-spot rich surface patches (Duhovny et al., 2002; Schneidman-Duhovny et al., 2005). For molecular docking, the X-ray crystal structure of human TLR3 ectodomain (ECD) was retrieved from PDB databank (PDB ID: 2A0Z) (Bell et al., 2005). The PatchDock molecular docking study provides dynamical properties of the designed system with the complex and also it gives 'exact' predictions of bulk properties including the hydrogen bond formation and conformation of the molecules forming the complex.

MD simulations study of MPVs and immune receptor complexes. The MPVs-TLR3 complex molecular interactions were further evaluated using MD simulation analysis. MD simulation studies were performed for 20 nanoseconds (ns) by using the YASARA tool (Yet Another Scientific Artificial Reality Application) (Krieger & Vriend, 2015). The MD simulations studies were carried out in an explicit water environment in a dodecahedron simulation box at a stabilized temperature of 298 K, the pressure of 1 atm and pH 7.4, with periodic cell boundary condition. The solvated systems were neutralized with counter ions (NaCl) (concentration 0.9 M). The AMBER14 force field was used on the systems during MD simulation (Case et al., 2014; Maier et al., 2015). The long-range electrostatic energy and forces were calculated using particle mesh-based Ewald method (Toukmaji et al., 2000). The solvated structures were energy minimized by the steepest descent method at a temperature of 298 K and a stable pressure of 1 atm. Further, the complexes were equilibrated for period of 1 ns. After equilibration, a production MD simulation was run for 20 ns at a stable temperature and pressure and time frames were saved at every 10 ps, for each MD simulations. The RMSD and RMSF values for $\text{C}\alpha$, Backbone and all the atoms of all the CTL and HTL MPVs in complex with TLR3 were analyzed for each MD simulation study.

Analysis of MPVs cDNA for expression in human host cell line. Codon-optimized complementary DNA (cDNA) of all the three CTL and two HTL MPVs was generated for favored expression in Mammalian cell line (Human) by Java Codon Adaptation Tool. The generated cDNA of all the MPVs was

further analyzed by GenScript Rare Codon Analysis Tool for its large-scale expression potential. The tool analyses the GC content, Codon Adaptation Index (CAI) and the Tandem rare codon frequency in a given query cDNA sequence (Morla et al., 2016; Wu et al., 2010). The CAI indicates the possibility of cDNA expression in the chosen expression system. The tandem rare codon frequency indicates the presence of low-frequency codons in the given query cDNA sequence.

Results

Screening of potential epitopes from SARS-CoV-2 proteome

Screening of CTL epitope. CTL epitopes were screened by 'MHC-I Binding Predictions' and 'MHC-I Processing Predictions' tools of IEDB. A total of 1013 CTL epitopes and HLA I alleles pairs were screened (Supplementary Tables S1, S2). The immunogenicity analysis of all the screened CTL epitopes was also performed by the tool 'MHC I Immunogenicity' of IEDB. The higher immunogenicity score of all the screened CTL epitopes suggests the high immunogenic potential of the CTL epitopes (Supplementary Table S1, S2; Srivastava et al., 2020a, 2020b; Grifoni et al., 2020a).

Screening of HTL epitopes. The screening of HTL epitopes from 11 different ORF proteins of SARS-CoV-2 was performed by the 'MHC-II Binding Predictions' tool of IEDB. The tool generates percentile rank for all the screened epitopes. The smaller the value of percentile indicates higher 'Percentile Rank' of the epitope. Furthermore, the higher Percentile rank suggests a higher affinity of the peptide to bind its respective HLA allele binders. A total of 314 HTL epitopes-HLA II allele pairs with high percentile rank were screened from the entire proteome of the SARS-CoV-2 virus (Supplementary Table S3; Srivastava et al., 2020a, 2020b; Grifoni et al., 2020a).

CTL and HTL epitope toxicity prediction. Toxicity analysis of all the screened CTL and HTL epitopes was also performed by the tool 'ToxinPred'. The ToxinPred study of all the screened CTL and HTL epitopes revealed that all the screened epitopes were nontoxic (Supplementary Tables S1, S2 and S3).

Overlapping CTL and HTL epitope clusters to Ag-Patches

CTL and HTL multi-epitope cluster-based Ag-Patch identification. A total of 73 Ag-Patches from CTL and a total of 49 Ag-Patches from HTL high-scoring 768 overlapping epitopes (518 CTL and 250 HTL epitopes) were identified (Tables 1 and 2; Figures 2–4) by the 'Overlapping-epitope-clusters-to-patches' method (Figure 1(A,B), Supplementary Tables S1, S2, S3). The Ag-Patches hence obtained from the CTL and HTL epitope clusters are expected to produce back clusters of upto 768 overlapping epitopes, targeting different HLA allele, while their proteolytic chop down processing is by proteasome and lysosome, in the process of antigen presentation by the APC. This would be crucial step for the MPVs and

showed strong possibility for the epitopes to be raised in their intact form and presented by APC. On the other hand, such large numbers of epitopes are not possible to be accommodated by the MPVs. The MPVs would also face challenges to give raise to the epitopes in their intact form upon proteolytic chop down processing by proteasome or lysosome, for presentation by APC (Figures 1–4). The identified SARS-CoV-2 Ag-Patches are included in filed Patent No: 202011037939 (Srivastava, 2020d).

Population coverage by the overlapping CTL and HTL epitopes. The population coverage by all the overlapping CTL and HTL epitopes based on their HLA allele binders was analyzed by the 'Population Coverage' tool of IEDB. The Ag-Patches identified (73 CTL and 49 HTL Ag-Patches) cover a large number of overlapping epitopes (518 CTL and 250 HTL epitopes) targeting a large number of HLA alleles (27 HLA class I and 28 HLA class II alleles, a total of 55 HLA alleles) are predicted to cover large human population ethnically distributed worldwide (Figures 2–4; Supplementary Table S4). The population coverage analysis has shown very convincing results with 99.98% of world population to be covered with an average of 91.11 and standard deviation of 16.97. In our analysis, all the human population ethnicity distribution worldwide was included. The countries most affected by COVID-19 have shown significant coverage like France, 100%; Italy, 99.82%; United States, 100%; England, 100%; Germany 100%; Russia, 100%; Brazil, 99.99%; India, 99.94%; China, 99.83%, etc. (Supplementary Table S5). Epitopes-HLA allele pairs are given in Supplementary-2.

Conservation analysis of antigenic patches. The identified Ag-Patches were further analyzed for their amino acid sequence conservancy across the source protein sequence library retrieved from NCBI protein database, by the 'Epitope Conservancy Analysis' tool of IEDB. We found that all the identified immunogenic Ag-Patches are highly conserved and both the CTL and HTL Ag-Patches are significantly conserved with their 100% amino acid residues largely conserved among the protein sequences of SARS-CoV-2 retrieved from NCBI Protein database (CTL epitopes cluster Ag-Patches were 47.50% to 100% conserved [largely above 91.23%] and HTL epitopes cluster Ag-Patches were 91.83% to 100% conserved [Tables 1 and 2]).

Multi-Patch Vaccines

Design of MPVs. The identified overlapping epitope cluster-based Ag-Patches from the proteome of the SARS-CoV-2 were utilized to design three CTL and two HTL MPVs (Figure 4). Short-peptide linkers EAAAK and GGGGS were utilized as rigid and flexible linkers, respectively. The short-peptide linker EAAAK facilitates the domain formation and provides a rigid link between two domains facilitating the protein to fold in a stable tertiary conformation. The short and flexible peptide linker GGGGS provides conformational flexibility and hence facilitates stable conformation to the final folded protein structure. The GGGGS was utilized to fuse the Ag-Patches together, and to ease folding of the protein into its tertiary conformation. The rigid linker EAAAK was utilized to fuse the human β defensin 2 and 3 (hBD-2 and hBD-3) at N

Table 1. Identified ‘overlapping-epitope-clusters-to-patches’ method-based CTL Ag-Patches from the entire proteome of the SARS-CoV-2 virus. The identified highly immunogenic Ag-Patches were utilized to design three CTLs (CTL-MPV-1, CTL-MPV-2 and CTL-MPV-3) Multi-Patch Vaccines. The identified Ag-Patches were highly conserved in nature and were identified on the basis of significant number of overlapping epitope-forming clusters (Figures 2, 3 and 4).

S.No.	CTL-MPVs	CTL epitope cluster-based Ag-Patches	Number of epitopes in cluster	Conservancy	Source SARS-CoV-2 protein
1	CTL-MPV-1	GTITVEELKKLLEQWNLVIGFLFLTICLLQFAYANRNRFYIIL FLWLLWPVTLACFVLA	36	96.64%(461/477)	M-Protein/6-70
2	CTL-MPV-1	RINWITGIAIAMACLVGLMWLSYFIASFRLFARTRSMSFN	28	96.01%(458/477)	M-Protein/72-113
3	CTL-MPV-1	KEITVATSRSLTYYKLGASQRVAGDSGFAAYSRYRIGNYKL	21	97.69%(466/477)	M-Protein/166-206
4	CTL-MPV-1	TTRTQLPPAYTNFSFRGVYYPDKVFRSSV	6	93.02%(467/502)	S-Protein/19-47
5	CTL-MPV-1	STQDLFLPFFSNVTWF	4	91.43%(459/502)	S-Protein/50-65
6	CTL-MPV-1	VLPFNDGVYVFASTEKSNIIRGWIF	6	93.62%(470/502)	S-Protein/83-106
7	CTL-MPV-1	GVYYHKNNSKWSMESEFRVYSSANNCTFEYVSQPFL	7	91.23%(458/502)	S-Protein/142-176
8	CTL-MPV-1	KQGNFKNLREFVFKNIDGYFKIYSKHTPI	4	91.83%(461/502)	S-Protein/182-210
9	CTL-MPV-1	EPLVDLPIGINTRFQTLLA	3	93.22%(468/502)	S-Protein/224-243
10	CTL-MPV-1	TPGDSSSGWTAGAAAAYYVGVLQPRFTLLK	6	93.42%(469/502)	S-Protein/250-278
11	CTL-MPV-1	SETKCTLKSFTEVKGIVQTSNFR	4	93.22%(468/502)	S-Protein/297-319
12	CTL-MPV-1	FPNITNLCPFGEVFNAATRFASV	3	93.42%(469/502)	S-Protein/329-350
13	CTL-MPV-1	RISNCVADYSVLYNSASFSTFKCY	7	93.82%(471/502)	S-Protein/357-380
14	CTL-MPV-1	NVYADSFVIRGDEVR	2	93.62%(470/502)	S-Protein/394-408
15	CTL-MPV-1	SKVGGNNYNYLRLFRKSNLKPFER	10	92.62%(465/502)	S-Protein/443-466
16	CTL-MPV-1	NTSNQAVAVLYQDVNCTEV	3	67.33%(338/502)	S-Protein/603-620
17	CTL-MPV-1	RVYSTGSNVF	2	93.82%(471/502)	S-Protein/634-643
18	CTL-MPV-1	SPRRARSVASQSIIAY	5	93.82%(471/502)	S-Protein/680-695
19	CTL-MPV-1	FTISVTTEILPVSMSTKTSVDCMTY	4	93.82%(471/502)	S-Protein/718-741
20	CTL-MPV-1	NTQEVAQVKQIYKTPPIK	3	91.83%(461/502)	S-Protein/777-795
21	CTL-MPV-1	KRSFIEDLLFNKVTLA	3	93.02%(467/502)	S-Protein/814-829
22	CTL-MPV-1	PLLTDEMIAQY	3	93.42%(469/502)	S-Protein/863-873
23	CTL-MPV-1	LQIPFAMQMAYRFNGI	5	93.82%(471/502)	S-Protein/894-909
24	CTL-MPV-1	FPQSAPHGVVF	2	93.42%(469/502)	S-Protein/1052-1062
25	CTL-MPV-1	FVSNGTHWVFTQR	2	93.22%(468/502)	S-Protein/1095-1107
26	CTL-MPV-1	YEQYIKWPWYIWLGFIAGLIAIV	8	93.42%(469/502)	S-Protein/1206-1228
27	CTL-MPV-2	NQRNAPRITFGGPS	2	96.78%(482/498)	N-Protein/8-21
28	CTL-MPV-2	RSKORRPQGLPNNTASWFTALTQHGK	6	99.19%(494/498)	N-Protein/36-61
29	CTL-MPV-2	NSSPDDQIGYYRRAIRRIRGGDGKMKDLSPRWYFYYL	14	97.38%(485/498)	N-Protein/77-113
30	CTL-MPV-2	LPYGANKDGIW	2	98.99%(493/498)	N-Protein/121-132
31	CTL-MPV-2	KSAAEASKPKRQTKTAKAYNVTQAF	5	97.38%(485/498)	N-Protein/249-274
32	CTL-MPV-2	QELIROGTDY	2	97.18%(484/498)	N-Protein/289-298
33	CTL-MPV-2	IAQFAPSASAFFGMSRIGMEVTPSGTWTLY	12	97.18%(484/498)	N-Protein/304-333
34	CTL-MPV-2	LLNKHIDAYKTFPPTEPKK	4	96.98%(483/498)	N-Protein/352-370
35	CTL-MPV-2	DEWSMATYYLF	2	99.55%(451/453)	ORF1ab/898-908/nsp3
36	CTL-MPV-2	ETISLAGSYK	2	100%(453/453)	ORF1ab/1503-1512/nsp3
37	CTL-MPV-2	TQVVDMISMTY	2	99.77%(452/453)	ORF1ab/1581-1590/nsp3
38	CTL-MPV-2	AVMYMGTLSYEQF	3	97.79%(443/453)	ORF1ab/1767-1779/nsp3
39	CTL-MPV-2	LVAEWFIALFTRFYV	3	100%(453/453)	ORF1ab/2323-2340/nsp3
40	CTL-MPV-2	RMYIFFASFYVVWKSY	3	99.11%(449/453)	ORF1ab/2382-2397/nsp3
41	CTL-MPV-2	AYVNNTFSSTFNVPMEK	3	98.89%(448/453)	ORF1ab/2593-2608/nsp3
42	CTL-MPV-2	CLAYYFMRFRRAF	3	96.02%(435/453)	ORF1ab/3059-3071/nsp4
43	CTL-MPV-2	MVMFTPLVPFWITIAY	4	100%(453/453)	ORF1ab/3129-3144/nsp4
44	CTL-MPV-2	FYWFFSNYLKRRVVF	2	98.67%(447/453)	ORF1ab/3153-3167/nsp4
45	CTL-MPV-2	LPSLATVAYFNMVY	2	98.01%(444/453)	ORF1ab/3641-3654/nsp6
46	CTL-MPV-2	LILMTARTVY	2	100%(453/453)	ORF1ab/3692-3701/nsp6
47	CTL-MPV-2	TACTDDNALAYY	5	99.77%(452/453)	ORF1ab/4161-4172/nsp9
48	CTL-MPV-2	YTMADLVYAL	2	100%(453/453)	ORF1ab/4514-4523/nsp12
49	CTL-MPV-2	KLFDRYFKYWDQTY	2	100%(453/453)	ORF1ab/4673-4686/nsp12
50	CTL-MPV-2	ARLYYDSMSY	2	99.11%(449/453)	ORF1ab/4904-4913/nsp12
51	CTL-MPV-2	VDTDFVNEFYAYLRKHFSM	3	100%(453/453)	ORF1ab/5129-5147/nsp12
52	CTL-MPV-2	FPLCANGQVFGLY	2	99.55%(451/453)	ORF1ab/5405-5417/nsp13
53	CTL-MPV-2	IPLMYKGLPWNVVR	2	100%(453/453)	ORF1ab/6075-6088/nsp14
54	CTL-MPV-2	YVMHANYIFW	2	98.89%(448/453)	ORF1ab/7020-7029/nsp16
55	CTL-MPV-3	SEETGTLIVNSVLLFLAFVVFLVTAILTALRLCAY	20	99.17%(478/482)	E-Protein/6-42
56	CTL-MPV-3	VSLVKPFSYVSYRKVNLNSSR	5	99.58%(480/482)	E-Protein/49-69
57	CTL-MPV-3	FMRIFTGTVTLK	5	99.37%(478/481)	ORF3a/4-16
58	CTL-MPV-3	TIPIQASLPFGWLIVGVALLAVFQSASKIITLKRW	14	77.33%(372/481)	ORF3a/34-69
59	CTL-MPV-3	HFVCNLLLFLFTV	5	97.50%(469/481)	ORF3a/78-90
60	CTL-MPV-3	LLVAAGLEAPFLYLYALVYFLQSQINFRIIMRLWLCWKCR	27	95.42%(459/481)	ORF3a/95-134
61	CTL-MPV-3	YFLCWHTNCYDY	3	97.08%(467/481)	ORF3a/145-156
62	CTL-MPV-3	TTSPISEHDYQIGGY	2	97.92%(471/481)	ORF3a/175-189
63	CTL-MPV-3	YYQLYSTQLSTDTGVEHVTFFIYNKI	6	99.16%(477/481)	ORF3a/211-236
64	CTL-MPV-3	FHLVDFQVTIAEILLIIMRTFKVSIWNLDYII	12	98.33%(473/481)	ORF6/2-33
65	CTL-MPV-3	MKIILFLALITLATCELYHY	7	99.58%(478/480)	ORF7a/1-20
66	CTL-MPV-3	LLKEPCSSGTYEGRSPFHPLADNKFALTFCSTQFAFACPDG VKHVYQLRARSVSPKLFIR	18	98.75%(474/480)	ORF7a/30-89
67	CTL-MPV-3	EVQELYSPIFIIVAAIVFILCFTLKRK	11	97.70%(469/480)	ORF7a/92-119

(continued)

Table 1. Continued.

S.No.	CTL-MPVs	CTL epitope cluster-based Ag-Patches	Number of epitopes in cluster	Conservancy	Source SARS-CoV-2 protein
68	CTL-MPV-3	MIELSLIDFYLCFLAFLFLVLVIMLIIFWFSLEL	24	97.88%(231/236)	ORF7b/1-34
69	CTL-MPV-3	MKFLVFLGIITVAAF	9	98.95%(475/480)	ORF8/1-16
70	CTL-MPV-3	YVVDDPCPIHFYSKWYIRVGARKSAPLIELC	12	98.95%(475/480)	ORF8/31-61
71	CTL-MPV-3	IQYIDIGNYTVCSCPFTINCQEPKLGSLVVRCSFYEDFLEYHDVRVVL	19	51.25%(246/480)	ORF8/71-118
72	CTL-MPV-3	IQYIDIGNYTVCSCPFTINCQEPKLGSLVVRCSFYEDFLEYHDVRVVL	19	47.50%(228/480)	ORF8/71-118
73	CTL-MPV-3	MGYINVFAFPFTIYSLLLCR	15	99.37%(476/479)	ORF10/1-20

and C terminal of the MPVs, respectively. The human β defensin 2 and 3 have been utilized here as an adjuvant to enhance immunogenic response. COVID-19 patients usually have lung infection and lung infection is usually coupled with increased expression of hBD-2 and hBD-3. Also, the β -defensins are involved in the chemotactic activity for memory T-cells, monocytes, immature dendritic cells as well as in degranulation of mast cells. The hBDs also enhance the innate and adaptive immunity and therefore are chosen here as adjuvants in the design of the MPVs (Biragyn et al., 2001; Chen et al., 2013; Duits et al., 2003; Hajighahramani et al., 2017; Hoover et al., 2000; Hu et al., 2004; Kohlgraf et al., 2010; Srivastava et al., 2018, 2019, 2020a, 2020b; Wilson et al., 2013; Yang et al., 2002) (Figures 1–5). MPV design and the method thereof are included in the filed Patent No.: 202011037939 (Srivastava, 2020d).

Physicochemical property analysis of designed MPVs. ProtParam analysis for all the designed three CTL and two HTL MPVs was performed to analyze their physicochemical properties. The empirical physicochemical properties of the CTL and HTL MPVs are given in Table 3. The molecular weight of all the MPVs ranges from 66.36 to 89.96 kDa. The expected half-life of all the MPVs to \sim 30 h in mammalian cells is very favorable for the vaccine protein expression and purification *in vitro*. The aliphatic index (53.51 to 100.86) and grand average of hydropathicity (GRAVY) (-0.274 to 0.445) of all the MPVs indicate their globular and hydrophilic nature. The instability index score of all the MPVs (39.68 to 53.37) indicates the stable nature of the protein molecules on expression *in vitro*. Overall, all the physicochemical parameters of all MPVs suggest a favorable expression of MPVs *in vitro* (Table 3).

Interferon-gamma-inducing epitope prediction from the MPVs. The interferon-gamma (IFN- γ)-inducing epitopes are involved in both innate and adaptive immune response. The IFN- γ -inducing 15 mer peptide epitopes were screened from the amino acid sequence of all the three CTL and two HTL MPVs by utilizing the IFN epitope tool. A total of 257 IFN- γ -inducing epitopes were screened from three CTL MPVs (108 from CTL-MPV-1; 66 from CTL-MPV-2 and 83 from CTL-MPV-3); likewise, a total of 181 IFN- γ -inducing epitopes were screened from two HTL MPVs (83 from HTL-MPV-1 and 98 from HTL-MPV-2). In the screening, only INF- γ -inducing POSITIVE epitopes with a score of 1 or more than 1 were included (Supplementary Tables S7 and S8).

MPVs allergenicity and antigenicity prediction. For all the MPVs, AlgPred and Vaxijen tools were implemented to

analyze the allergic and antigenic nature, respectively. All the three CTL and two HTL MPVs were suggested as NON-ALLERGEN (scoring: -0.46602545 for CTL-MPV-1; -0.71579187 for CTL-MPV-2; -0.90796056 for CTL-MPV-3; -0.56197384 for HTL-MPV-1 and -0.72617142 for HTL-MPV-2; and the threshold cutoff, -0.4). Likewise, all the three CTL and two HTL MPVs were found to be potential ANTIGENS as suggested by Vaxijen analysis (scoring: 0.5241 for CTL-MPV-1, 0.4811 for CTL-MPV-2, 0.6534 for CTL-MPV-3, 0.5016 for HTL-MPV-1 and 0.5096 for HTL-MPV-2; default threshold being 0.4). Largely, AlgPred and Vaxijen analysis of all the three CTL MPVs and two HTL MPVs suggests that all the MPVs are non-allergic in nature but they are antigenic. These both properties are essential and favorable for a potential vaccine candidate.

Tertiary structure modelling and refinement of MPVs. The tertiary structure models of all the designed three CTL and two HTL MPVs were generated by utilizing the I-TASSER modeling tool (Figure 6). The parameters (C-score, TM-Score and RMSD) of homology modeling have shown acceptable values for all the MPV models (Supplementary Table S9). The C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It indicates the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score typically ranges from -5 to 2, where higher value signifies a model with high confidence and vice versa. The TM-score indicates the structural alignment between the query structure and the templet structures. The RMSD is the deviation between the residues that are structurally aligned (by TM-align) to the templet structure (Xu & Zhang, 2011).

All the generated three CTL and two HTL MPV tertiary models were further refined using ModRefiner and GalaxyRefine refinement tools to fix any gaps and improve stoichiometric conformational parameters of the models. The refinement by ModRefiner shows the TM-score of 0.9729, 0.9324, 0.9849, 0.9847 and 0.9854 for CTL-MPV-1, CTL-MPV-2, CTL-MPV-3, HTL-MPV-1 and HTL-MPV-1 models, respectively. The TM-score being close to 1 suggests that the initial and the refined models were structurally similar. After refinement, the RMSD for CTL-MPV-1, CTL-MPV-2, CTL-MPV-3, HTL-MPV-1 and HTL-MPV-1 models with respect to the initial model was 2.183 Å, 2.634 Å, 1.185 Å, 1.120 Å and 1.162 Å, respectively, indicating not much deviation from the initial structure.

Next, all the three CTL and two HTL MPV models were refined by GalaxyRefine and the top model was chosen based on the best scoring parameters. All the three CTL and

Table 2. Identified ‘overlapping-epitope-clusters-to-patches’ method-based HTL Ag-Patches from the entire proteome of the SARS-CoV-2 virus. The identified highly immunogenic Ag-Patches were utilized to design two HTL (HTL-MPV-1 and HTL-MPV-2) Multi-Patch Vaccines. The identified Ag-Patches were highly conserved in nature and were identified on the basis of significant number of overlapping epitope-forming clusters (Figures 2, 3 and 4).

S.No.	HTL-MPVs	CTL epitope cluster based Ag-Patches	Number of epitopes in cluster	Conservancy	Source SARS-CoV-2 protein
1	HTL-MPV-1	VNSVLLFLAFVVFLVLTLAILT	8	99.58%(480/482)	E-Protein/14-35
2	HTL-MPV-1	IKLIFLWLLWPVTLAC	2	98.11%(468/477)	m-protein/49-64
3	HTL-MPV-1	VGLMWLSYFIASFRLFAR	4	97.27%(464/477)	m-protein/88-105
4	HTL-MPV-1	SRTLSYYKLGASQRVAGDS	5	98.53%(470/477)	m-protein/173-191
5	HTL-MPV-1	SKTQSLLIVNNATNVVIKVE	7	93.22%(468/502)	S-protein/112-132
6	HTL-MPV-1	REFVFKNIDGYFKIYSK	3	91.83%(461/502)	S-protein/190-206
7	HTL-MPV-1	GINITRFQTLALHRSYLTGPDSSSGWTAG AAAYYVGY	9	92.82%(466/502)	S-protein/232-269
8	HTL-MPV-1	QPTESIVRFPNITNLCPFG	5	93.42%(469/502)	S-protein/321-339
9	HTL-MPV-1	ADYSVLYNSASFSTFKC	3	93.82%(471/502)	S-protein/363-379
10	HTL-MPV-1	YADSFVIRGDEVRQIAPGQ	5	93.62%(470/502)	S-protein/396-414
11	HTL-MPV-1	VVLSFELLHAPATVCGPK	5	92.23%(463/502)	S-protein/510-528
12	HTL-MPV-1	IAIPTNFTISVTTEILPVS	5	93.82%(471/502)	S-protein/712-730
13	HTL-MPV-1	ECSNLLLQYGSFCTQLNR	4	93.82%(471/502)	S-protein/748-765
14	HTL-MPV-1	QQLIRAAEIRASANLA	2	93.22%(468/502)	S-protein/1010-1025
15	HTL-MPV-1	REGVFVSGNTHWFVTQ	2	93.22%(468/502)	S-protein/1091-1106
16	HTL-MPV-1	PWYIWLGFIAGLIAIVMVTIMLCC	9	93.42%(469/502)	S-protein/1213-1236
17	HTL-MPV-1	DDQIGYYRRATRIRRGGDG	5	98.79%(492/498)	N-Protein/81-99
18	HTL-MPV-1	GTRNPANNAAIVLQLPQGTTL	7	97.38%(485/498)	N-Protein/147-167
19	HTL-MPV-1	WPQIAQFAPSASAFFGMSR	5	97.38%(485/498)	N-Protein/301-319
20	HTL-MPV-1	TPSGTWLTYTGAIKLDKD	5	97.18%(484/498)	N-Protein/325-343
21	HTL-MPV-1	SDFVRATATIPIQASLP	3	99.37%(478/481)	3a/26-42
22	HTL-MPV-1	APFLYLYALVYFLQSLNFV	5	96.46%(464/481)	3a/103-121
23	HTL-MPV-1	DTGVEHVTFFIYNKIVDEPEEHV	7	99.37%(478/481)	3a/222-244
S.No.	HTL-MPVs	CTL epitope cluster-based Ag-Patches	Number of epitopes in cluster	Conservancy	Source SARS-CoV-2 protein
24	HTL-MPV-2	AIILASFSASTSASFVET	3	100%(453/453)	1ab/474-490/nsp2
25	HTL-MPV-2	YAFASEAARVVRISFS	2	99.55%(451/453)	1ab/534-549/nsp2
26	HTL-MPV-2	TLEETKFLTENLLLYIDIN	5	99.77%(452/453)	1ab/1242-1260/nsp3
27	HTL-MPV-2	VQQESPFWMMSSAPPQAQEL	5	99.77%(452/453)	1ab/1798-1816/nsp3
28	HTL-MPV-2	LLQLCTFTRSTNSRIKASM	5	98.23%(445/453)	1ab/2176-2194/nsp3
29	HTL-MPV-2	SKLINIIIIWFLLLSVCL	3	97.79%(443/453)	1ab/2224-2240/nsp3
30	HTL-MPV-2	EWFLAYILFTRFFYVGLAAlMQLFFSYFAVHFI SNSWLMWLIINLVQMA	17	100%(453/453)	1ab/2326-2375/nsp3
31	HTL-MPV-2	VRMYIFFASFYVVWKSYV	4	98.45%(446/453)	1ab/2381-2398/nsp3
32	HTL-MPV-2	WLKQLIKVTLVFLVAAIFYLITPVH	9	99.77%(452/453)	1ab/2769-2794/nsp4
33	HTL-MPV-2	FSAVGNICYTPSKLIEYT	4	99.55%(451/453)	1ab/2889-2906/nsp4
34	HTL-MPV-2	DISASIVAGGIVAIIVTCL	5	100%(453/453)	1ab/3042-3060/nsp4
35	HTL-MPV-2	FTPLVPFWITIAYIICIST	5	100%(453/453)	1ab/3132-3150/nsp4
36	HTL-MPV-2	VQSTQWSLFFFYENAFLP	5	92.27%(418/453)	1ab/3595-3613/nsp6
37	HTL-MPV-2	QAIASEFSSLPSYAAFAT	4	99.33%(450/453)	1ab/3942-3959/nsp8
38	HTL-MPV-2	TQMNLKYAISAKNRART	3	99.77%(452/453)	1ab/4932-4948/nsp12
39	HTL-MPV-2	RAMPNMLRIMASLVLARKH	5	100%(453/453)	1ab/5016-5034/nsp12
40	HTL-MPV-2	FVNEFYAYLRKHFSMMIL	4	100%(453/453)	1ab/5133-5150/nsp12
41	HTL-MPV-2	PEVKILNNLGVDIAAN	2	99.77%(452/453)	1ab/6519-6534/nsp15
42	HTL-MPV-2	SQMEIDFLEAMDEFIERY	5	99.11%(449/453)	1ab/6659-6677/nsp15
43	HTL-MPV-2	FAWWTAFTVNNASSS	2	98.89%(448/453)	1ab/6985-7000/nsp16
44	HTL-MPV-2	HANYIFWRNTNPIQLSSY	4	98.89%(448/453)	1ab/7023-7040/nsp16
45	HTL-MPV-2	EILLIMRTFKVSIWNLDYIINLIIK	6	99.16%(477/481)	orf6/13-38
46	HTL-MPV-2	MKIILFLALITLATCELY	4	99.37%(478/481)	orf7a/1-18
47	HTL-MPV-2	LSLIDFYLCFLAFLFLVLIIMLIWFWSLEL	14	97.88%(231/236)	7b/4-34
48	HTL-MPV-2	CTQHQPYVVDPCPIHFYS	5	99.16%(476/480)	orf8/25-43
49	HTL-MPV-2	YINVFAFPFTIYSLLCRM	5	99.37%(476/479)	orf10/3-21

two HTL MPV refinement output models have acceptable Rama favored, GDT-HA, RMSD value and MolProbity scores (Supplementary Tables S9 and S10). The Rama favored is the percentage of residues, which come in the favored region of the Ramachandran plot, the GDT-HA (global distance test-High Accuracy) indicates the accuracy of the backbone structure, the RMSD value indicates deviation from the initial model and the MolProbity score indicates the log-weighted combination score of the clash score, the percentage of

Ramachandran not favored residues and the percentage of bad side chain rotamers.

Overall, after refinement, all the CTL and HTL MPV models had all their stoichiometric parameters in acceptable range and hence, all these models were carried forward for further analysis (Supplementary Tables S9 and S10; Figure 6).

Validation of CTL and HTL MPVs refined models. All the three CTL and two HTL MPV models were further analyzed again by the RAMPAGE analysis tool after refinement to their

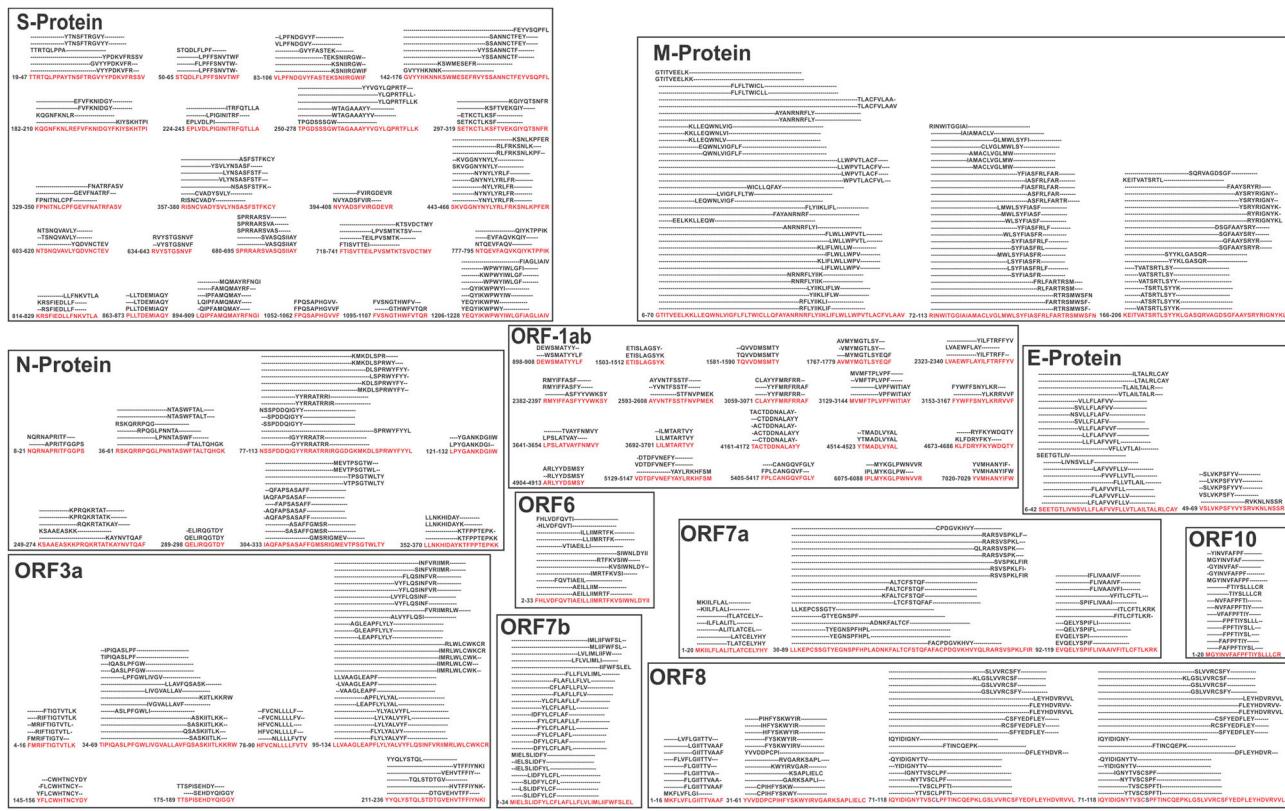


Figure 2. Immunogenic Ag-Patches (antigenic patches) identified from overlapping CTL epitope clusters of SARS-CoV-2 proteins. Immunogenic Ag-Patches are identified (red amino acid sequences) on the basis of the overlapping epitope clusters by the reverse epitomics approach, 'Overlapping-epitope-clusters-to-patches' method.

stoichiometric acceptable conformation. All the three refined CTL MPV models (CTL-MPV-1, CTL-MPV-2 and CTL-MPV-3) were found to have 87.0%, 89.20% and 88.8% favored region residues; 10.2%, 8% and 7.6% allowed region residues and 2.7%, 2.8% and 3.6% outlier region residues, respectively. Likewise, all the two refined HTL MPV models (HTL-MPV-1 and HTL-MPV-2) were found to have 86.9% and 93.0% favored region residues, 9.3% and 5.9% allowed region residues, and only 3.7% and 1.1% outlier region residues, respectively (Supplementary Figure S3). Hence, all the three CTL and two HTL MPV models showed acceptable stoichiometric conformational parameters.

Linear and discontinuous B-cell epitope prediction from the MPVs. The generated tertiary models of CTL and HTL MPVs were further screened for linear and discontinuous B-cell epitopes. The B-cell epitope screening was performed by the ElliPro tool available of IEDB. The screening revealed that the CTL MPV carries a total of 41 linear [CTL-MPV-1 (12 epitopes), CTL-MPV-2 (14 epitopes), and CTL-MPV-3 (15 epitopes)] and 19 discontinuous B-cell epitopes [CTL-MPV-1 (7 epitopes), CTL-MPV-2 (5 epitopes), and CTL-MPV-3 (7 epitopes)]. Likewise, the HTL MPVs carries a total of 30 linear [HTL-MPV-1 (16 epitopes) and HTL-MPV-2 (14 epitopes)] and 13 potential discontinuous epitopes [HTL-MPV-1 (8 epitopes) and HTL-MPV-2 (5 epitopes)]. The ElliPro associates each predicted epitope with a score, defined as a PI (Protrusion Index) value. The high PI score of the linear and discontinuous epitopes originating from the CTL and HTL

MPVs suggests all the MPVs to have high potential to cause humoral immune response (Supplementary Tables S11–S14).

Molecular interaction analysis of MPVs and immune receptor complexes. The immune receptor TLR3 acts as sentinel for our immune system. The molecular signaling by multiple TLRs is an essential component of the innate immune response against viral infection. Upon lung infection, the peripheral blood mononuclear cells from infected patients (PBEC) were found to express TLR3, which is involved in recognition of RNA (Delneste et al., 2007; Farina et al., 2005; Totura et al., 2015). Hence, we tested all the CTL and HTL MPV models for their molecular interaction with the ectodomain of the TLR3 receptor by molecular docking and MD simulation studies. Molecular docking study for the CTL and HTL MPVs models with the TLR3-ECD crystal structure model (PDB ID: 2A0Z) was performed by utilizing the PatchDock tool. The generated docking complex conformations with the highest docking score were chosen for further study [CTL-MPV-1:TLR3 (docking score: 17696), CTL-MPV-2:TLR3 (docking score: 17118), CTL-MPV-3:TLR3 (docking score: 16562), HTL-MPV-1:TLR3 (docking score: 21432) and HTL-MPV-2:TLR3 (docking score: 17620)]. The highest docking score indicates the MPV and TLR3-ECD complexes to have best geometric shape complementarity fitting conformation as predicted by the PatchDock tool. All the three CTL and two HTL MPVs were fitting into the ectodomain region of TLR3 receptor (Figure 7).

Figure 3. Immunogenic Ag-Patches (antigenic patches) identified from overlapping HTL epitope clusters of SARS-CoV-2 proteins. Immunogenic Ag-Patches identified (blue amino acid sequences) on the basis of the overlapping epitope clusters by the reverse epitomics approach, 'Overlapping-epitope-clusters-to-patches' method.

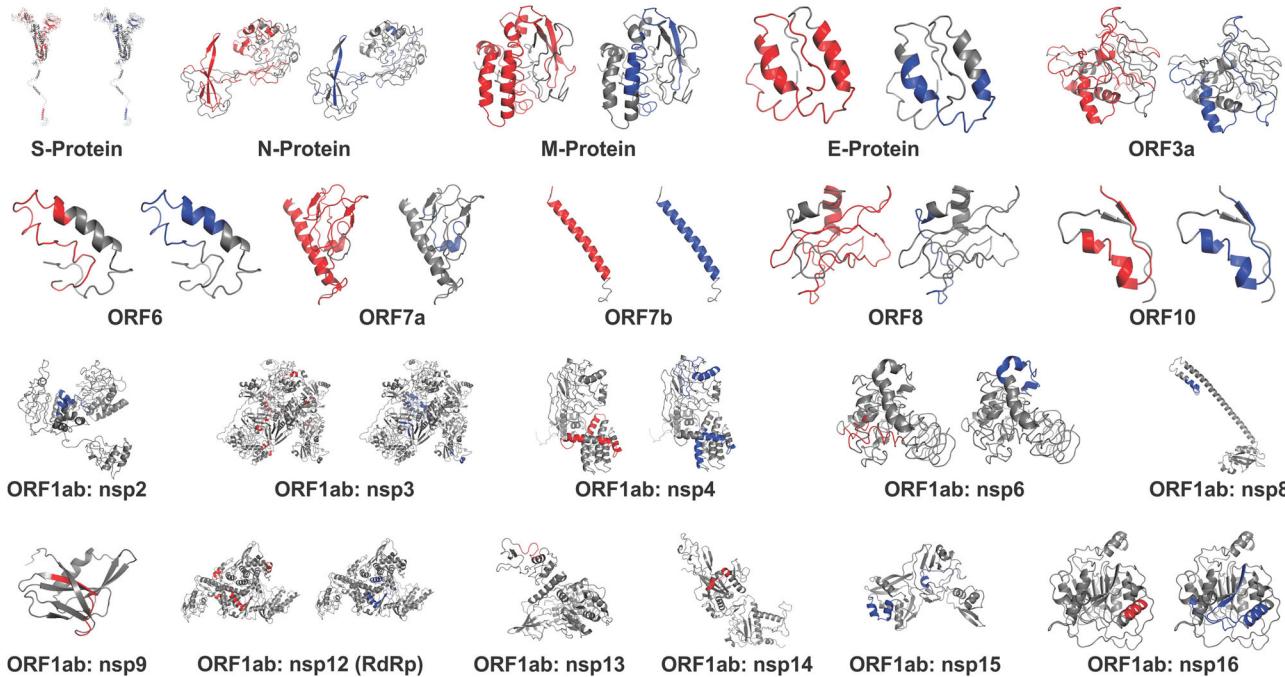


Figure 4. The identified immunogenic Ag-Patches (antigenic patches) shown in the tertiary structure models of the SARS-CoV-2 proteins. The tertiary structure models of all the SARS-CoV-2 proteins have been retrieved from I-TASSER homology modeling server (<https://zhanglab.ccmb.med.umich.edu/COVID-19/>). The CTL Ag-Patches are shown in red, and the HTL Ag-Patches are shown in blue color. The identification of novel Ag-Patches has been done by the reverse epitomics approach, ‘overlapping-epitope-clusters-to-patches’ method. Most of the Ag-Patches identified in the SARS-CoV-2 proteins are observed to be on the exposed surface of the proteins.

The B-factor analysis of all the MPV and TLR3-ECD complexes was also performed. The B-factor of a tertiary structure indicates the displacement of the atomic positions from an average (mean) value of the structure; that is, the more

flexible an atom is the larger the displacement from the mean position will be (mean-square displacement) (Figure 7: A-I, B-IV, C-VII, D-X, E-XIII). The B-factor analysis of the CTL and HTL MPVs in complex with TLR3-ECD shows that most

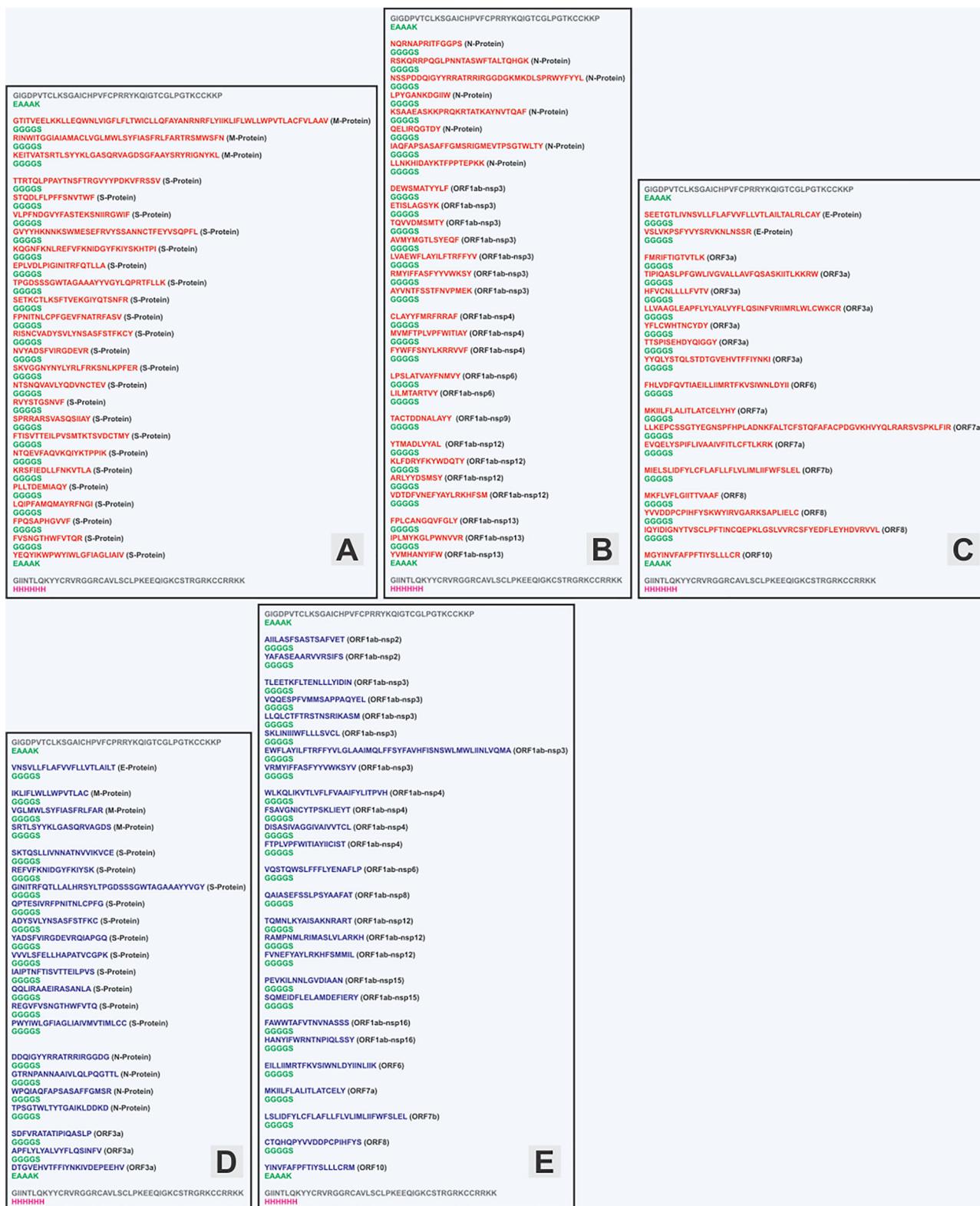


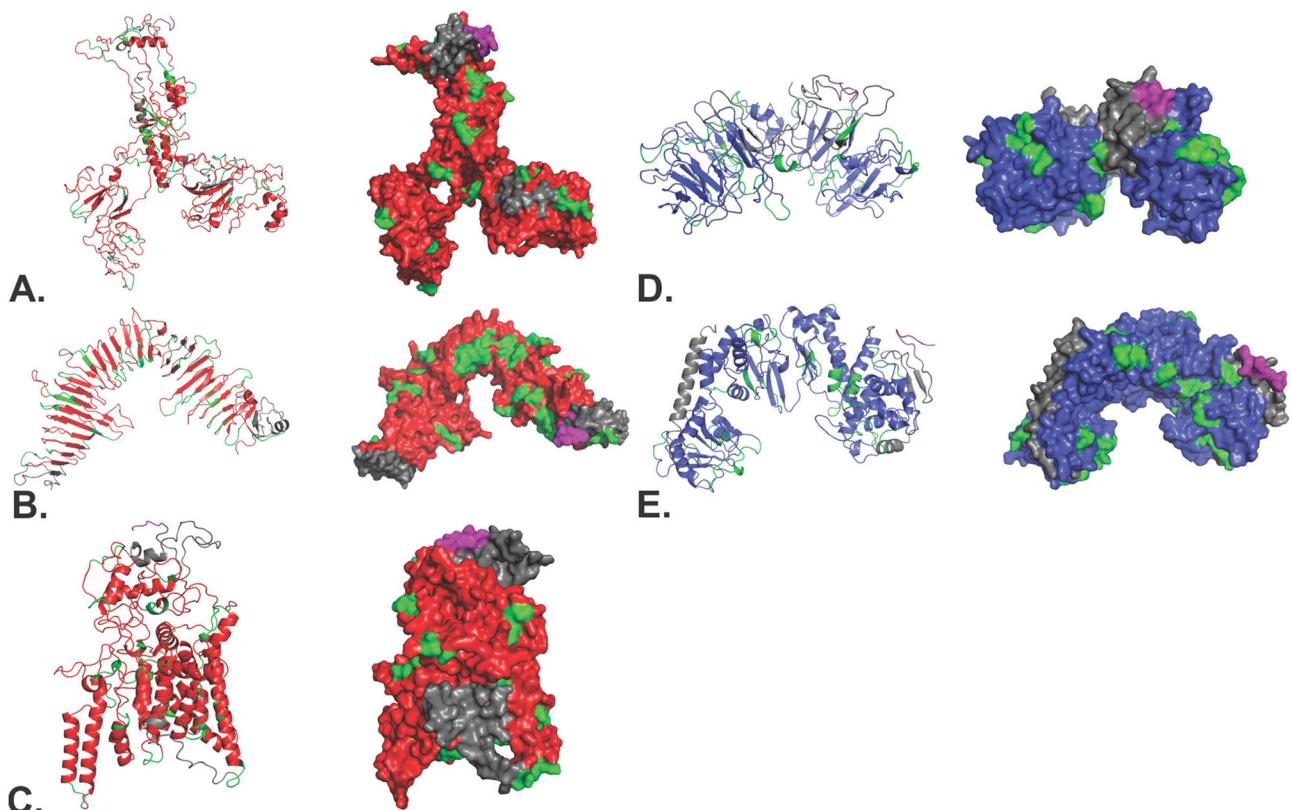
Figure 5. The CTL and HTL Ag-Patches were utilized to design Multi-Patch Vaccine. Short peptide linkers GGGGS and EAAK were used to fuse the Ag-Patches (anti-genic patches) and the adjuvants, respectively. The CTL MPV constructs includes (A) CTL-MPV-1, (B) CTL-MPV-2 and (C) CTL-MPV-3. The HTL MPV construct includes (D) HTL-MPV-1 and (E) HTL-MPV-2 ([Supplementary Table S6](#)). CTL and HTL Patches are shown in red and blue, respectively; adjuvants, linkers and 6xHis tags are shown in gray, green and magenta, respectively.

of the regions of MPVs bound to TLR3-ECD are stable nature in nature. The B-factor analysis has been represented by the VIBGYOR color presentation with blue representing low B-factor and red representing high B-factor. Since the entire

regions of the MPVs in complex with TLR3-ECD are in blue, hence stable complex formation tendency for all the CTL and HTL MPVs with the ectodomain of the human TLR3 receptor is suggested.

Table 3. Physicochemical property analysis based on the amino acid sequences of all the designed three CTL and two HTL Multi-Patch Vaccines.

MPVs	Length	Molecular weight	Theoretical pI	Expected half-life	Aliphatic Index	Grand average of Hydropathicity (GRAVY)	Instability index score	
CTL-MPV-1	842 aa	89.96 kDa	9.64	<i>E. coli</i> : >10 Hr Yeast: >20 Hr Mammalian cell: 30 Hr	70.97	-0.088	44.43	-
CTL-MPV-2	675 aa	71.75 kDa	9.64	<i>E. coli</i> : >10 Hr Yeast: >20 Hr Mammalian cell: 30 Hr	53.51	-0.274	53.37	
CTL-MPV-3	689 aa	74.70 kDa	9.09	<i>E. coli</i> : >10 Hr Yeast: >20 Hr Mammalian cell: 30 Hr	100.86	0.445	39.68	
HTL-MPV-1	644 aa	66.36 kDa	9.35	<i>E. coli</i> : >10 Hr Yeast: >20 Hr Mammalian cell: 30 Hr	80.57	0.094	44.18	
HTL-MPV-2	757 aa	80.60 kDa	9.06	<i>E. coli</i> : >10 Hr Yeast: >20 Hr Mammalian cell: 30 Hr	93.05	0.419	45.58	

**Figure 6.** Tertiary structure modelling of CTL and HTL Multi-Patch Vaccines. Tertiary structural models of CTL and HTL MPVs have been shown in both cartoon and surface presentations. (A) CTL-MPV-1, (B) CTL-MPV-2, (C) CTL-MPV-3, (D) HTL-MPV-1 and (E) HTL-MPV-2. The Ag-Patches are shown in Red (CTL Ag-Patches) and blue (HTL Ag-Patches). Linkers are in green, adjuvants are in gray and 6xHis tag is shown in magenta.

All the CTL and HTL MPVs have shown to form prominent and close binding site region with TLR3-ECD as represented by binding patches shown in magenta (Figure 7: A-II, B-V, C-VIII, D-XI, E-XIV). The complex forms multiple hydrogen bonds in the interaction interface of MPVs and the ectodomain cavity region of TLR3 receptor (Figure 7: A-III, B-VI, C-IX, D-XII, E-XV). The residues involved in hydrogen bond formation for CTL-MPV-1:TLR3-ECD are GLN325:SER342, GLY423:THR391, GLY425:THR391, SER427:THR400, GLU308:SER420, LYS36:LEY595, GLU42:THR679, TYR88:LYS619, ARG85:LYS619,

ARG85:PHE644. The residues involved in hydrogen bond formation for CTL-MPV-2:TLR3-ECD are TRP524:ASN667, SER542:THR638, ASP518:LYS613, SER572:LYS589, GLU620:TYR465. The residues involved in hydrogen bond formation for CTL-MPV-3:TLR3-ECD are LYS677:LEU340, ARG676:ILE338, ASP388:SER571. The residues involved in hydrogen bond formation for HTL-MPV-1:TLR3-ECD are HIS203:ASN517, SER189:ASN541, GLY187:HIS539, SER232:LYS619, GLY92:ASN678, GLY71:THR623, ASN379:HIS684, ASN379:HIS682, ALA363:HIS682, GLN462:VAL658, SER481:SER694. The residues involved in

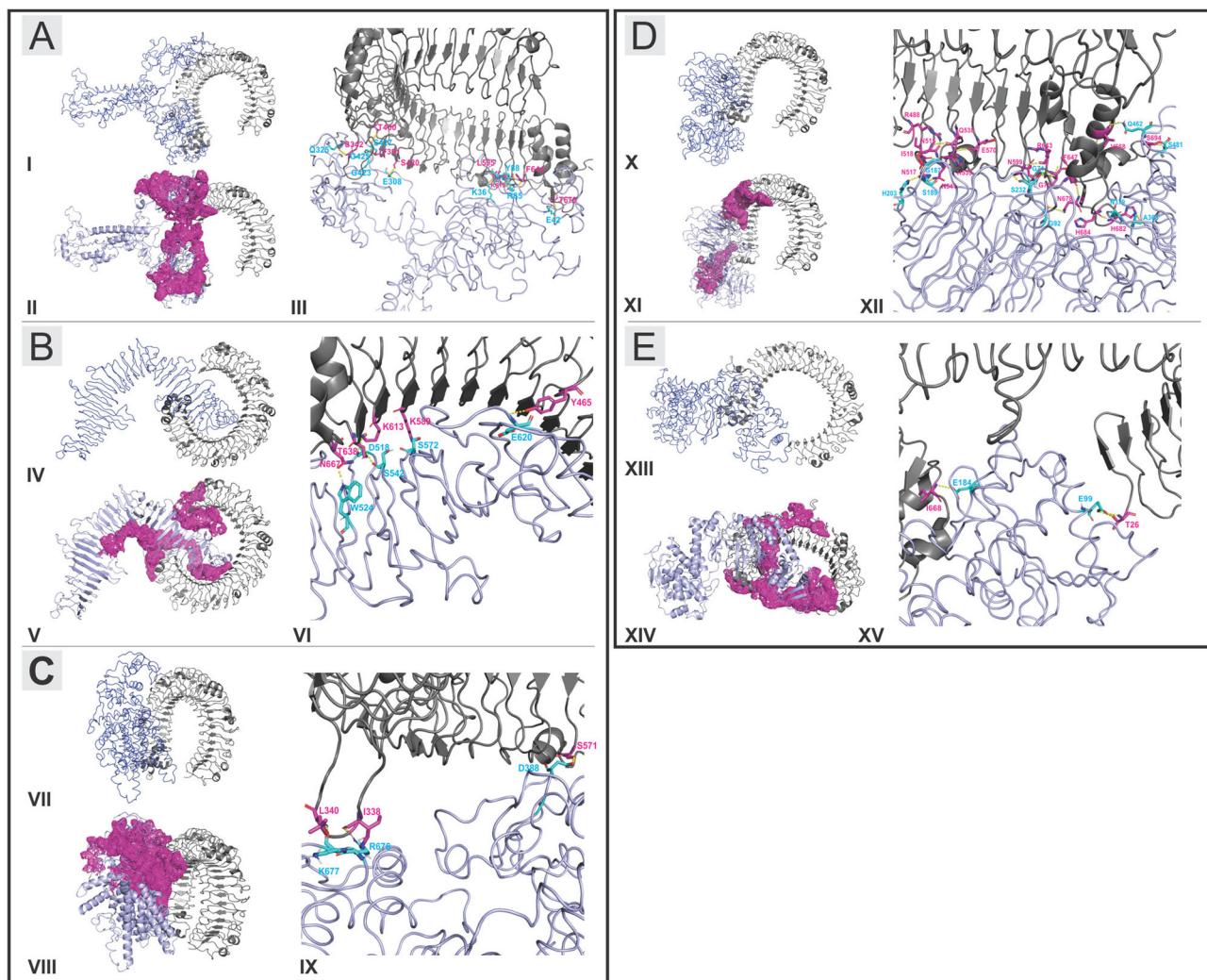


Figure 7. Molecular docking study of CTL and HTL MPVs with TLR3-ECD complexes. (A, B, C, D, E): Complex formation by the MPVs (CTL-MPV-1, CTL-MPV-2, CTL-MPV-3, HTL-MPV-1, HTL-MPV-2) and TLR3-ECD, respectively. (A-I, B-IV, C-VII, D-X, E-XIII): B-factor analysis presentation for the docked MPVs to the TLR3-ECD. The presentation is in VIBGYOR color, with blue showing low B-factor and red show high B-factor. Here, most of the MPV regions are in blue showing low B-factor and hence suggesting a stable complex formation with TLR3-ECD. (A-II, B-V, C-VIII, D-XI, E-XIV): Binding site formed by the MPVs and TLR3-ECD at the molecular interaction interface, represented in magenta-binding patches. All the CTL and HTL MPVs have shown to form prominent and close binding site region with TLR3-ECD. (A-III, B-VI, C-IX, D-XII, E-XV): Representation of hydrogen bond formation is shown. The residues involved in hydrogen bond formation from MPVs are shown in cyan, and the residues from TLR3-ECD are shown in magenta. The hydrogen bonds are shown by yellow dots.

hydrogen bond formation for HTL-MPV-2:TLR3-ECD are GLU184:ILE668 and GLU99:THR26.

MD simulations study of MPVs and immune receptor complexes. All the three CTL and two HTL MPVs in complex with TLR3-ECD were further subjected to MD simulation analysis to investigate the stability of the molecular interaction between the MPVs and the TLR3-ECD. All the complexes have shown a very convincing and reasonably stable RMSD values for $\text{C}\alpha$, Backbone and all-atom of the complexes, which further stabilizes toward the end of MD simulation [(CTL-MPV-1:TLR3-ECD, RMSD: ~4 to ~13 Å) (CTL-MPV-2:TLR3-ECD, RMSD: ~4 to ~11 Å) (CTL-MPV-3:TLR3-ECD, RMSD: ~2.5 to ~5.5 Å) (HTL-MPV-1:TLR3-ECD, RMSD: ~3 to ~6.5 Å) (HTL-MPV-2:TLR3-ECD, RMSD: ~2.5 to ~8 Å)] (**Supplementary Figure S4: A-E**). The RMSD values for all the MPVs:TLR3-ECD complexes were stable to the above-mentioned RMSD for a given time window of 20 nanoseconds at stable temperature (~278 K) and pressure (~1 atm). Molecular docking and MD simulation study of all the MPVs and TLR3-ECD complexes

suggests a stable complex formation tendency for all the MPVs with TLR3-ECD.

Further, most of the amino acid residues of the CTL and HTL MPVs in complex with TLR3-ECD have shown to have root-mean-square fluctuation (RMSF) from their initial confirmation, for all the amino acid residues of the MPVs, in acceptable range [(CTL-MPV-1:TLR3-ECD, RMSF: ~2 to ~18 Å) (CTL-MPV-2:TLR3-ECD, RMSF: ~2 to ~14 Å) (CTL-MPV-3:TLR3-ECD, RMSF: ~2 to ~8 Å) (HTL-MPV-1:TLR3-ECD, RMSF: ~2 to ~10.5 Å) (HTL-MPV-2:TLR3-ECD, RMSF: ~2 to ~13 Å)] (**Supplementary Figure S4: F-J**). These results suggest that all the MPVs and TLR3-ECD complexes to have a stable nature.

Furthermore, the radius of gyration analysis for all the MPVs and TLR3-ECD complexes was performed. The Radius of gyration (R_g) indicates the compactness of the complex conformation. R_g is concerned with how regular secondary structures are compactly packed in tertiary structure of protein complex throughout the MD simulation. The variation in R_g for all the MPVs and TLR3-ECD complexes remains in

acceptable range throughout the MD simulation [(CTL-MPV-1:TLR3-ECD, Rg: ~44.8 to ~47.0 Å) (CTL-MPV-2:TLR3-ECD, Rg: ~46.5 to ~51 Å) (CTL-MPV-3:TLR3-ECD, Rg: ~36.3 to ~37.4 Å) (HTL-MPV-1:TLR3-ECD, Rg: ~40.2 to ~42.50 Å) (HTL-MPV-2:TLR3-ECD, Rg: ~44.6 to ~47.75 Å)] ([Supplementary Figure S4](#): K-O). This suggests that the MPVs:TLR3-ECD complexes remain structurally compact throughout the 20-ns MD simulation analysis.

Analysis of MPVs cDNA for expression in human host cell line. All the three CTL and two HTL MPV construct were further analyzed for feasibility for expression *in vitro*. Codon-optimized cDNA (complementary DNA) for all the CTL and HTL MPVs was generated for expression in the mammalian host cell line (Human) by the Java Codon Adaptation Tool. The generated optimized cDNA for all the MPVs was further analyzed for their expression feasibility *in vitro*, by the GenScript Rare Codon Analysis Tool. The analysis revealed that the codon-optimized cDNA of all the CTL and HTL MPVs satisfies all the crucial parameters such as GC content, CAI (Codon Adaptation Index) score and tandem rare codons for high-level expression in a mammalian cell line (Human) ([Supplementary Table S15](#)). The GC content of a cDNA, the CAI score that indicates the possibility of cDNA expression in a chosen expression system (here, human cell line) and the tandem rare codon frequency that indicates the presence of low-frequency codons in cDNA, all these parameters were in acceptable range for all the cDNA of the MPVs. The tandem rare codons that may hinder the proper expression of the cDNA or even interrupt the translational machinery of the chosen expression system were observed to be 0% in all the MPVs. Hence, as per the GenScript Rare Codon analysis, the cDNA of all the MPVs has a high potential for large-scale expression in the human cell line. The workflow concept, from Ag-Patch identification to codon optimization of cDNA and *in vivo* trial for the here proposed MPVs, is shown in [Supplementary Figure S5](#).

Discussion

The majority of the vaccine design against SARS-CoV-2 are focused on the single protein, protein subunits or the epitopes from SARS-CoV-2 proteins, mostly S, E, M, N and ORF1ab proteins. The recent strategies to design and develop vaccine to combat SARS-CoV-2 involve subunit vaccines or multi-epitope vaccines. The subunit vaccine involves the use of single protein or a multiple subunit of SARS-CoV-2 proteins. On the other hand, the multi-epitope vaccine involves fusion of multiple epitopes identified from the proteome of the SARS-CoV-2 by short peptide linkers. Several subunit and multi-epitope-based vaccine designs have been published claiming potential to activate CD4 and CD8 T-cell immune response driving long-term robust adaptive immunity in the vast majority of the population (Abdelmageed et al., 2020; Ahmed et al., 2020; Akhand et al., 2020; An et al., 2000; Banerjee et al., 2020; Baruah & Bose, 2020; Bhattacharya et al., 2020; Fast & Chen, 2020; Gragert et al., 2013; Gupta et al., 2020; Herst et al., 2020; Ismail et al., 2020; Khan et al., 2020; Lee & Koohy, 2020; Liu et al., 2020; Lu et al., 2014;

Mitra et al., 2020; Nerli & Sgourakis, 2020; Poran et al., 2020; Ramaiah & Arumugaswami, 2020; Saha & Prasad, 2020; Sheikhshahrokh et al., 2020; Singh et al., 2020; Srivastava et al., 2020a; 2020b; Ul Qamar et al., 2020; Vashi et al., 2020; Yarmarkovich, Farrel et al., 2020; Yazdani et al., 2020). Numerous highly antigenic regions have also been reported from SARS-CoV-2 proteins, which have been recognized with a large population coverage by favorable binding with large number of HLA allele distributed among different ethnic human population across the world (Grifoni et al., 2020b; Yarmarkovich, Warrington, et al., 2020).

Although the major focus is on epitope- or subunit-based vaccine design to combat SARS-CoV-2, both the approaches have several limitations. The major limitation with single protein or multiple subunit-based vaccine is the limited efficiency of the vaccines. On the other hand, the epitope-based vaccines have major drawback of frequent mutations in the proteome of the SARS-CoV-2 virus. The multi-epitope-based vaccines have one more major challenge for the presentation of chosen epitopes by the APC. The proteolytic chop down processing by proteasome and lysosome would leave a very narrow chance for the epitopes of the multi-epitope vaccines, to remain intact and successfully presented by APC.

In this study, we have reported a novel method to design a vaccine against SARS-CoV-2 by utilizing multiple antigenic patches from the viral proteins. The Ag-Patches used have been identified by the clusters of overlapping epitopes. The identification of these Ag-Patches was performed by reverse epitomics analysis, of the high scoring CTL and HTL epitopes screened from all the ORF proteins of the SARS-CoV-2 virus. The method is here termed as 'Overlapping-epitope-clusters-to-patches' method. All the screened epitopes were well characterized for their conservancy, immunogenicity, nontoxicity and large population coverage. The clusters of the overlapping epitopes lead us to identify the Ag-Patches. These Ag-Patches from all the ORF proteins of the SARS-CoV-2 proteome were utilized further to design MPV candidate against the SARS-CoV-2 infection.

The designed MPVs from the antigenic patches of SARS-CoV-2 proteins have several advantages over to the subunit and multi-epitope-based vaccines. The Ag-Patches utilized were identified and collected from the entire proteome of the SARS-CoV-2. This would enhance the efficiency of the vaccines and lead the vaccine to be more effective. The MPVs consist of the identified Ag-Patches will have potential to raise multiple epitopes in clusters upon the chop down processing by proteasome and lysosome in the APC. The identified Ag-Patches will also provide larger chance of the epitopes raised after proteasome and lysosomal processing to get presentation by the APC and elicit immune response. Since the Ag-Patches were identified by the large number of epitopes forming clusters, the MPVs designed would have potential to raise a larger number of epitopes upon proteasome and lysosomal processing; hence, a larger number of HLA alleles could be targeted and hence, larger ethnic human population could be covered by the MPVs, in comparison with the limited number of epitope used in multi-epitope vaccines. For instance, the five MPVs designed in

this study used the Ag-Patches, which were identified by 768 (518 CTL and 250 HTL epitopes) overlapping epitope targeting different HLA alleles. Such inclusion of large number of epitopes and targeting large number of HLA alleles is not possible for multi-epitope-based vaccines prepared with limited epitopes.

All the identified Ag-Patches utilized to design MPVs have shown to be highly conserved amongst the protein sequences of SARS-CoV-2 available at NCBI protein database. All the physicochemical properties of SARS-CoV-2 favor their overexpression in vitro. This is further supported by the cDNA analysis of the codon-optimized constructs of all the MPVs for high expression in a mammalian cell line (human). Furthermore, the designed MPVs also show stable binding tendency with the TLR3-ECD, which is an essential criteria for an antigen to be recognized and processed by the human immune system.

Conclusion

In this study, we have identified highly immunogenic novel Ag-Patches (73 CTL and 49 HTL) from the entire proteome of SARS CoV-2. The Ag-Patches were identified by a novel reverse epitomics approach, 'overlapping-epitope-clusters-to-patches' method. These Ag-Patches are highly conserved in nature and found in most of the SARS-CoV-2 protein sequences available at NCBI protein database. These Ag-Patches are identified on the basis of high scoring, immunogenic, overlapping epitopes thoroughly screened from entire proteome of SARS-CoV-2. Further, for the first time we have utilized the identified multiple immunogenic Ag-Patches to design novel MPV, which proposes a new methodology for vaccine design. The MPVs designed against SARS-CoV-2 in our study have potential to give raise up to a total of 768 epitopes (518 CTL and 250 HTL epitopes) targeting a large number of different HLA alleles. Such a large number of epitopes are not possible for multi-epitope-based vaccine design. The large number of epitopes covered implies a large number of HLA alleles targeted. The large number of HLA allele targeted further implies large ethnic human population coverage worldwide. The MPVs, with multiple epitope cluster-based Ag-Patches from entire proteome of SARS-CoV-2, are in better position to raise multiple epitopes in comparison with the MEVs upon proteasome or lysosomal chop down processing by the APC. The designed MPVs against SARS-CoV-2 are validated for stable complex formation with ectodomain of TLR3. The physicochemical properties and the codon-optimized cDNA analysis of all the MPVs designed suggest favored large-scale expression potential. We conclude that the novel MPV designed by the novel Ag-Patches could be a highly potential novel method to combat SARS-CoV-2, with greater effectiveness, high specificity and large human population coverage worldwide.

Acknowledgments

We thank Rajni Kant Dixit, Scientist D, of ICMR-National Institute of Malaria Research (NIMR) for providing computational facility for

experimentations. We also thank Indian Foundation for Fundamental Research (IFFR) for providing the computational and other infrastructural facility for experimentations. We are thankful to the Indian Council for Medical Research (ICMR) for providing SRF fellowship to Sonia Verma (ISRM/11(98)/2019).

Author's contribution

SS carried out the following works: idea conceived, methodology designed and performed, Antigenic Patches (Ag-Patches) identified, MPVs designed and scientific writing; S.V. performed MD simulation; S.S., S.V., M.K., D.A., A.K.S., M. Kolbe, S. Singh., A.K., S.A.N., H-J.S., K.V., K.C.P and B.R carried out scientific reporting and revising the manuscript.

Disclosure statement

Two patents have been filed from the report. All rights reserved.

Funding

IFFR (Indian Foundation For Fundamental Research)

ORCID

Sukrit Srivastava  <http://orcid.org/0000-0003-0295-0272>
 Mohit Kamthania  <http://orcid.org/0000-0002-6361-7257>
 Ajay Kumar Saxena  <http://orcid.org/0000-0001-6967-030X>
 Michael Kolbe  <http://orcid.org/0000-0002-1219-0249>
 Sarman Singh  <http://orcid.org/0000-0002-0749-9647>

References

- Abdelmageed, M. I., Abdelmoneim, A. H., Mustafa, M. I., Elfadol, N. M., Murshed, N. S., Shantier, S. W., & Makhawi, A. M. (2020). Design of multi epitope-based peptide vaccine against E protein of human 2019-nCoV: An immunoinformatics approach. *Biomed Research International*, 2020, 1–21. <https://doi.org/10.1155/2020/2683286>
- Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses*, 12(3), 254. <https://doi.org/10.3390/v12030254>
- Akhand, M. R. N., Azim, K. F., Hoque, S. F., Moli, M. A., Joy, B. D., 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*. <https://doi.org/10.1101/2020.04.15.036285>
- An, L. L., Rodriguez, F., Harkins, S., Zhang, J., & Whitton, J. L. (2000). Quantitative and qualitative analyses of the immune responses induced by a multivalent minigene DNA vaccine. *Vaccine*, 18(20), 2132–2141. [https://doi.org/10.1016/S0264-410X\(99\)00546-0](https://doi.org/10.1016/S0264-410X(99)00546-0)
- Antoniou, A. N., Powis, S. J., & Elliott, T. (2003). Assembly and export of MHC class I peptide ligands. *Current Opinion in Immunology*, 15(1), 75–81. [https://doi.org/10.1016/S0952-7915\(02\)00010-9](https://doi.org/10.1016/S0952-7915(02)00010-9)
- Banerjee, A., Santra, D., & Maiti, S. (2020). Energistics based epitope screening in SARS CoV-2 (COVID 19) spike glycoprotein by Immuno-informatic analysis aiming to a suitable vaccine development. *BioRxiv*. <https://doi.org/10.1101/2020.04.02.201725>
- Baruah, V., & Bose, S. (2020). Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *Journal of Medical Virology*, 92(5), 495–500. <https://doi.org/10.1002/jmv.25698>
- Bell, J. K., Botos, I., Hall, P. R., Askins, J., Shiloach, J., Segal, D. M., & Davies, D. R. (2005). The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), 10548–10553. <https://doi.org/10.1073/pnas.0503700102>

- Sciences of the United States of America*, 102(31), 10976–10980. <https://doi.org/10.1073/pnas.0505077102>
- Bhattacharya, M., Sharma, A. R., Patra, P., Ghosh, P., Sharma, G., Patra, B. C., Lee, S. S., & Chakraborty, C. (2020). Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-CoV-2): Immunoinformatics approach. *Journal of Medical Virology*, 92(6), 618–631. <https://doi.org/10.1002/jmv.25736>
- Biragyn, A., Surenhu, M., Yang, D., Ruffini, P. A., Haines, B. A., Klyushnenkova, E., Oppenheim, J. J., & Kwak, L. W. (2001). Mediators of innate immunity that target immature, but not mature, dendritic cells induce antitumor immunity when genetically fused with nonimmunogenic tumor antigens. *Journal of Immunology (Baltimore, Md: 1950)*, 167(11), 6644–6653. <https://doi.org/10.4049/jimmunol.167.11.6644>
- Bui, H. H., Sidney, J., Dinh, K., Southwood, S., Newman, M. J., & Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics*, 7, 153. <http://tools.iedb.org/population/> <https://doi.org/10.1186/1471-2105-7-153>
- Bui, H. H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*, 8, 361. <http://tools.iedb.org/conservancy/> <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., Kesmir, C., & Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. *PLOS Computational Biology*, 8(1), 361. <http://tools.iedb.org/immunogenicity/>
- Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Gohlke, H., & Goetz, A. W. (2014). The FF14SB force field. *Amber*, 14, pp.29–31.
- Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*, 65(10), 1357–1369. <https://doi.org/10.1016/j.addr.2012.09.039>
- Delneste, Y., Beauvillain, C., & Jeannin, P. (2007). [Innate immunity: structure and function of TLRs]. *Medecine Sciences: M/S*, 23(1), 67–73. <https://doi.org/10.1051/medsci/200723167>
- Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, 8, 30. <https://doi.org/10.1186/1745-6150-8-30>
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8(1), 4. <http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>
- Duhovny, D., Nussinov, R., & Wolfson, H. J. (2002). Efficient unbound docking of rigid molecules<SE-END> [Paper presentation].</SE-END>. Proceedings of the 2nd Workshop on Algorithms in Bioinformatics(WABI). In Gusfield (Ed.), *Lecture Notes in Computer Science* 2452 (pp. 185–200). Springer Verlag.
- Duits, L. A., Nibbering, P. H., van Strijen, E., Vos, J. B., Mannesse-Lazeroms, S. P., van Sterkenburg, M. A., & Hiemstra, P. S. (2003). Rhinovirus increases human β-defensin-2 and -3 mRNA expression in cultured bronchial epithelial cells. *FEMS Immunology and Medical Microbiology*, 38(1), 59–64. [https://doi.org/10.1016/S0928-8244\(03\)00106-8](https://doi.org/10.1016/S0928-8244(03)00106-8)
- Farina, C., Krumbholz, M., Giese, T., Hartmann, G., Aloisi, F., & Meinl, E. (2005). Preferential expression and function of Toll-like receptor 3 in human astrocytes. *Journal of Neuroimmunology*, 159(1–2), 12–19. <https://doi.org/10.1016/j.jneuroim.2004.09.009>
- Fast, E., & Chen, B. (2020). Potential T-cell and B-cell Epitopes of 2019-nCoV. *bioRxiv*. <https://doi.org/10.1101/2020.02.19.955484>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein identification and analysis tools on the ExPASy server* (pp. 571–607). Humana Press. <https://web.expasy.org/protparam/>
- Gragert, L., Madbouly, A., Freeman, J., & Maiers, M. (2013). Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Human Immunology*, 74(10), 1313–1320. <https://doi.org/10.1016/j.humimm.2013.06.025>
- Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & Sette, A. (2020a). Candidate targets for immune responses to 2019-Novel Coronavirus (nCoV): Sequence homology-and bioinformatic-based predictions. *CELL-HOST-MICROBE-D-20-00119*.
- Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & Sette, A. (2020b). A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host & Microbe*, 27(4), 671–680.e2. <https://doi.org/10.1016/j.chom.2020.03.002>
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. (2013). *In silico* approach for predicting toxicity of peptides and proteins. *PLoS One*, 8(9), e73957. http://crdd.osdd.net/raghava/tox-inpred/multi_submit.php <https://doi.org/10.1371/journal.pone.0073957>
- Gupta, E., Mishra, R. K., & Niraj, R. R. K. (2020). Identification of potential vaccine candidates against SARS-CoV-2, a step forward to fight novel coronavirus 2019-nCoV: A reverse vaccinology approach. *bioRxiv*. <https://doi.org/10.1101/2020.04.13.039198>
- Hajighahramani, N., Nezafat, N., Eslami, M., Negahdaripour, M., Rahmatabadi, S. S., & Ghasemi, Y. (2017). Immunoinformatics analysis and in silico designing of a novel multi-epitope peptide vaccine against *Staphylococcus aureus*. *Infection, Genetics and Evolution*, 48, 83–94. <https://doi.org/10.1016/j.meegid.2016.12.010>
- Herst, C. V., Burkholz, S., Sidney, J., Sette, A., Harris, P. E., Massey, S., Brasel, T., Cunha-Neto, E., Rosa, D. S., Chao, W. C. H., Carback, R., Hodge, T., Wang, L., Ciotlos, S., Lloyd, P., & Rubsamen, R. (2020). An effective CTL peptide vaccine for ebola zaire based on survivors' CD8+ targeting of a particular nucleocapsid protein epitope with potential implications for COVID-19 vaccine design. *Vaccine*, 38(28), 4464–4475. <https://doi.org/10.1016/j.vaccine.2020.04.034>
- Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., Buus, S., & Nielsen, M. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*, 61(1), 1–13. <https://doi.org/10.1007/s00251-008-0341-z>
- Hoover, D. M., Rajashankar, K. R., Blumenthal, R., Puri, A., Oppenheim, J. J., Chertov, O., & Lubkowski, J. (2000). The structure of human beta-defensin-2 shows evidence of higher order oligomerization. *The Journal of Biological Chemistry*, 275(42), 32911–32918. PDB ID: 1FD3. <https://doi.org/10.1074/jbc.M006098200>
- Hu, W., Li, F., Yang, X., Li, Z., Xia, H., Li, G., Wang, Y., & Zhang, Z. (2004). A flexible peptide linker enhances the immunoreactivity of two copies HBsAg preS1 (21–47) fusion protein. *Journal of Biotechnology*, 107(1), 83–90. <https://doi.org/10.1016/j.jbiotec.2003.09.009>
- Ismail, S., Ahmad, S., & Azam, S. S. (2020). Immuno-informatics characterization SARS-CoV-2 spike glycoprotein for prioritization of epitope based multivalent peptide vaccine. *bioRxiv*. <https://doi.org/10.1101/2020.04.05.206005>
- Kahan, B. D. (2003). Individuality: The barrier to optimal immunosuppression. *Nature Reviews Immunology*, 3(10), 831–838. <https://doi.org/10.1038/nri1204>
- Khan, A., Alam, A., Imam, N., Siddiqui, M. F., & Ishrat, R. (2020). Design of an epitope-based peptide vaccine against the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2): A vaccine informatics approach. *bioRxiv*. <https://doi.org/10.1101/2020.05.03.074930>
- 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. <http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE> <https://doi.org/10.1093/nar/gks493>
- Kohlgraf, K. G., Pingel, L. C., Dietrich, D. E., & Brogden, K. A. (2010). Defensins as anti-inflammatory compounds and mucosal adjuvants. *Future Microbiology*, 5(1), 99–113. <https://doi.org/10.2217/fmb.09.104>
- Krieger, E., & Vriend, G. (2015). New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry*, 36(13), 996–1007. <https://doi.org/10.1002/jcc.23899>
- Kringelum, J. V., Lundsgaard, C., Lund, O., & Nielsen, M. (2012). Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Computational Biology*, 8(12), e1002829. <https://doi.org/10.1371/journal.pcbi.1002829>
- Lee, C. H., & Koohy, H. (2020). *In silico* identification of vaccine targets for 2019-nCoV. *F1000Research*, 9, 145. <https://doi.org/10.12688/f1000research.225072>

- Liu, G., Carter, B., Bricken, T., Jain, S., Viard, M., Carrington, M., & Gifford, D. K. (2020). Computationally optimized SARS-CoV-2 MHC Class I and II vaccine formulations predicted to target human haplotype distributions. *Cell Systems*, 11(2), 131–144.e6. <https://doi.org/10.1016/j.cels.2020.06.009>
- Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., & Richardson, D. C. (2003). Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3), 437–450. <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php> <https://doi.org/10.1002/prot.10286>
- Lu, Y.-C., Yao, X., Crystal, J. S., Li, Y. F., El-Gamil, M., Gross, C., Davis, L., Dudley, M. E., Yang, J. C., Samuels, Y., Rosenberg, S. A., & Robbins, P. F. (2014). Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 20(13), 3401–3410. <https://doi.org/10.1158/1078-0432.CCR-14-0433>
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- Melief, C. J. (2005). Escort service for cross-priming. *Nature Immunology*, 6(6), 543–544. <https://doi.org/10.1038/ni0605-543>
- Mitra, D., Shekhar, N., Pandey, J., Jain, A., & Swaroop, S. (2020). Multi-epitope based peptide vaccine design against SARS-CoV-2 using its spike protein. *bioRxiv*. <https://doi.org/10.1101/2020.04.23.2055467>
- Morla, S., Makhija, A., & Kumar, S. (2016). Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene*, 584(1), 1–6. <http://www.jcat.de/> <https://doi.org/10.1016/j.gene.2016.02.047>
- Nagpal, G., Gupta, S., Chaudhary, K., Dhanda, S. K., Prakash, S., & Raghava, G. P. (2015). VaccineDA: Prediction, design and genome-wide screening of oligodeoxynucleotide-based vaccine adjuvants. *Scientific Reports*, 5, 12478. <http://crdd.osdd.net/raghava/ifnepitope/scan.php> <https://doi.org/10.1038/srep12478>
- NCBI: Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome; gi|1798174254|ref|NC_045512.2|;[https://www.ncbi.nlm.nih.gov/projects/sviewer/?id=NC_045512&tracks=\[key:sequence_track,name:Sequence,display_name:Sequence,id:STD649220238,annots:Sequence>ShowLabel:false,ColorGaps:false,shown:true,order:1\]\[key:gene_model_track,name:Genes,display_name:Genes,id:STD3194982005,annots:Unnamed,Options:ShowAllButGenes,CDSProductFeats:true,NtRuler:true,AaRuler:true,HighlightMode:2,ShowLabel:true,shown:true,order:9\]&v=1:29903&c=null&select=null&slim=0](https://www.ncbi.nlm.nih.gov/projects/sviewer/?id=NC_045512&tracks=[key:sequence_track,name:Sequence,display_name:Sequence,id:STD649220238,annots:Sequence>ShowLabel:false,ColorGaps:false,shown:true,order:1][key:gene_model_track,name:Genes,display_name:Genes,id:STD3194982005,annots:Unnamed,Options:ShowAllButGenes,CDSProductFeats:true,NtRuler:true,AaRuler:true,HighlightMode:2,ShowLabel:true,shown:true,order:9]&v=1:29903&c=null&select=null&slim=0)
- Nerli, S., & Sgourakis, N. G. (2020). Structure-based modeling of SARS-CoV-2 peptide/HLA-A02 antigens. *bioRxiv*. <https://doi.org/10.1101/2020.03.23.2004176>
- Nielsen, M., Lundegaard, C., & Lund, O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics*, 8(1), 238. <https://doi.org/10.1186/1471-2105-8-238>
- Peters, B., Bulik, S., Tampe, R., Van Endert, P. M., & Holzhutter, H. G. (2003). Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *The Journal of Immunology*, 171(4), 1741–1749. <https://doi.org/10.4049/jimmunol.171.4.1741>
- Ponomarenko, J. V., Bui, H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, 9, 514. <http://tools.iedb.org/elliPro/> <https://doi.org/10.1186/1471-2105-9-514>
- Poran, A., Harjanto, D., Malloy, M., Rooney, M. S., Srinivasan, L., & Gaynor, R. B. (2020). Sequence-based prediction of vaccine targets for inducing T cell responses to SARS-CoV-2 utilizing the bioinformatics predictor RECON. *bioRxiv*. <https://doi.org/10.1101/2020.04.06.207805>
- Ramaiah, A., & Arumugaswami, V. (2020). Insights into cross-species evolution of novel human coronavirus 2019-nCoV and defining immune determinants for vaccine development. *bioRxiv*. <https://doi.org/10.1101/2020.01.29.925867>
- Ramakrishnan, C., & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformations: II. Allowed conformations for a pair of peptide units. *Biophysical Journal*, 5(6), 909–933. [https://doi.org/10.1016/S0006-3495\(65\)86759-5](https://doi.org/10.1016/S0006-3495(65)86759-5)
- 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. <https://zhanglab.ccmb.med.umich.edu/I-TASSER/> <https://doi.org/10.1038/nprot.2010.5>
- Saha, R., & Prasad, B. V. (2020). In silico approach for designing of a multi-epitope based vaccine against novel Coronavirus (SARS-CoV-2). *bioRxiv*. <https://doi.org/10.1101/2020.03.31.2017459>
- Saha, S., & Raghava, G. (2006). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, 34(Web Server issue), W202–W209. <http://crdd.osdd.net/raghava/algpred/submit.html> <https://doi.org/10.1093/nar/gkl343>
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, 33(Web Server issue), W363–W367. <http://biinfo3.cs.tau.ac.il/PatchDock/> <https://doi.org/10.1093/nar/gki481>
- Sheikhshahrokh, A., Ranjbar, R., Saeidi, E., Dehkordi, F. S., Heiat, M., Ghasemi-Dehkordi, P., & Goodarzi, H. (2020). Frontier therapeutics and vaccine strategies for sars-cov-2 (COVID-19): A review. *Iranian Journal of Public Health*, 49, 18–29. <https://doi.org/10.18502/ijph.v49i51.3666>
- Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A., & Peters, B. (2008). Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immune Research*, 4(1), 2. <https://doi.org/10.1186/1745-7580-4-2>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539. <https://www.ebi.ac.uk/Tools/msa/clustalo/> <https://doi.org/10.1038/msb.2011.75>
- Singh, A., Thakur, M., Sharma, L. K., & Chandra, K. (2020). Designing a multi-epitope peptide-based vaccine against SARS-CoV-2. *bioRxiv*. <https://doi.org/10.1101/2020.04.15.2040618>
- Srivastava, S., Kamthania, M., Kumar Pandey, R., Kumar Saxena, A., Saxena, V., Kumar Singh, S., Kumar Sharma, R., & Sharma, N. (2019). Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics. *Journal of Biomolecular Structure and Dynamics*, 37(16), 4345–4360. <https://doi.org/10.1080/07391102.2018.1548977>
- Srivastava, S., Kamthania, M., Singh, S., Saxena, A. K., & Sharma, N. (2018). Structural basis of development of multi-epitope vaccine against middle east respiratory syndrome using in silico approach. *Infection and Drug Resistance*, 11, 2377–2391. <https://doi.org/10.2147 IDR.S175114>
- Srivastava, S., Verma, S., Kamthania, M., Kaur, R., Badyal, R. K., Saxena, A. K., Shin, H. J., Kolbe, M., & Pandey, K. C. (2020a). Structural basis for designing multiepitope vaccines against COVID-19 infection: In silico vaccine design and validation. *JMIR Bioinformatics and Biotechnology*, 1(1), e19371. <https://doi.org/10.2196/19371>
- Srivastava, S., Verma, S., Kamthania, M., Kaur, R., Badyal, R. K., Saxena, A. K., Shin, H. J., Kolbe, M., & Pandey, K. (2020b). Structural basis to design multi-epitope vaccines against Novel Coronavirus 19 (COVID19) infection, the ongoing pandemic emergency: An in silico approach. *bioRxiv*. <https://doi.org/10.1101/2020.04.01.2019299>
- Srivastava, S. (2020c). A novel method for designing Multi-Patch Vaccine and diagnostic kit. Patent No: 202011037585.
- Srivastava, S. (2020d). A novel Multi-Patch Vaccine design to combat SARS-CoV-2 and a method to prepare thereof. Patent No: 202011037939.
- Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U., Braxenthaler, M., Gallazzi, F., Protti, M. P., Sinigaglia, F., & Hammer, J. (1999). Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nature Biotechnology*, 17(6), 555–561. <https://doi.org/10.1038/9858>
- Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., Kloetzel, P. M., Rammensee, H. G., Schild, H., & Holzhutter, 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*, 62(9), 1025–1037. <http://tools.iedb.org/processing/> <https://doi.org/10.1007/s00018-005-4528-2>
- Totura, A. L., Whitmore, A., Agnihothram, S., Schäfer, A., Katze, M. G., Heise, M. T., & Baric, R. S. (2015). Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *mBio*, 6(3), e00638–e00615. <https://doi.org/10.1128/mBio.00638-15>
- Toukmaji, A., Sagui, C., Board, J., & Darden, T. (2000). Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. *The Journal of Chemical Physics*, 113(24), 10913–10927. <https://doi.org/10.1063/1.1324708>
- Ul Qamar, M. T., Rehman, A., Ashfaq, U. A., Awan, M. Q., 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. *bioRxiv*. <https://doi.org/10.1101/2020.02.28.970343>
- Vashi, Y., Jagrit, V., & Kumar, S. (2020). Understanding the B and T cells epitopes of spike protein of severe respiratory syndrome coronavirus-2: A computational way to predict the immunogens. *bioRxiv*. <https://doi.org/10.1101/2020.04.08.013516>
- Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., & Peters, B. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, 11(1), 568. <https://doi.org/10.1186/1471-2105-11-568>
- Wang, Z., & Xu, J. (2013). Predicting protein contact map using evolutionary and physical constraints by integer programming. *Bioinformatics (Oxford, England)*, 29(13), i266–i273. <http://tools.iedb.org/mhcii/> <https://doi.org/10.1093/bioinformatics/btt211>
- WHO Coronavirus Disease (COVID-19) Dashboard; as on 5th September, (2020). <https://covid19.who.int/>
- WHO Weekly Epidemiological Record (WER); as on 4th September, (2020). <https://www.who.int/wer/2020/wer9536/en/>; <https://apps.who.int/iris/bitstream/handle/10665/334140/WER9536-eng-fre.pdf?ua=1>
- Wilson, S. S., Wiens, M. E., & Smith, J. G. (2013). Antiviral mechanisms of human defensins. *Journal of Molecular Biology*, 425(24), 4965–4980. <https://doi.org/10.1016/j.jmb.2013.09.038>
- Wu, X., Wu, S., Li, D., Zhang, J., Hou, L., Ma, J., Liu, W., Ren, D., Zhu, Y., & He, F. (2010). Computational identification of rare codons of *Escherichia coli* based on codon pairs preference. *Bmc Bioinformatics*, 11(1), 61. <https://www.genscript.com/tools/rare-codon-analysis> <https://doi.org/10.1186/1471-2105-11-61>
- Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, 101(10), 2525–2534. <https://zhanglab.ccmb.med.umich.edu/ModRefiner/> <https://doi.org/10.1016/j.bpj.2011.10.024>
- Yang, D., Biragyn, A., Kwak, L. W., & Oppenheim, J. J. (2002). Mammalian defensins in immunity: More than just microbicidal. *Trends in Immunology*, 23(6), 291–296. [https://doi.org/10.1016/S1471-4906\(02\)02246-9](https://doi.org/10.1016/S1471-4906(02)02246-9)
- Yarmarkovich, M., Farrel, A., Sison, A., III, Di Marco, M., Raman, P., Parris, J. L., Monos, D., Lee, H., Stevanovic, S., & Maris, J. M. (2020). Immunogenicity and immune silence in human cancer. *Frontiers in Immunology*, 11, 69. <https://doi.org/10.3389/fimmu.2020.00069>
- Yarmarkovich, M., Warrington, J. M., Farrel, A., & Maris, J. M. (2020). Identification of SARS-CoV-2 Vaccine epitopes predicted to induce long-term population-scale immunity. *Cell Reports Medicine*, 1(3), 100036. <https://doi.org/10.1016/j.xcrm.2020.100036>
- Yazdani, Z., Rafiei, A., Yazdani, M., & Valadan, R. (2020). Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2: An *in silico* analysis. *bioRxiv*. <https://doi.org/10.1101/2020.04.20.051557>