



# Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics

Sukrit Srivastava, Mohit Kamthania, Rajesh Kumar Pandey, Ajay Kumar Saxena, Vaishali Saxena, Santosh Kumar Singh, Rakesh Kumar Sharma & Nishi Sharma

To cite this article: Sukrit Srivastava, Mohit Kamthania, Rajesh Kumar Pandey, Ajay Kumar Saxena, Vaishali Saxena, Santosh Kumar Singh, Rakesh Kumar Sharma & Nishi Sharma (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, DOI: [10.1080/07391102.2018.1548977](https://doi.org/10.1080/07391102.2018.1548977)

To link to this article: <https://doi.org/10.1080/07391102.2018.1548977>



[View supplementary material](#)



Published online: 16 Jan 2019.



[Submit your article to this journal](#)



Article views: 1297



[View related articles](#)



[View Crossmark data](#)



Citing articles: 17 [View citing articles](#)

RESEARCH ARTICLE



## Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics

Sukrit Srivastava<sup>a,b</sup> Mohit Kamthania<sup>a,c</sup>, Rajesh Kumar Pandey<sup>d</sup>, Ajay Kumar Saxena<sup>b</sup>, Vaishali Saxena<sup>b</sup>, Santosh Kumar Singh<sup>d</sup>, Rakesh Kumar Sharma<sup>a</sup> and Nishi Sharma<sup>a</sup>

<sup>a</sup>Department of Biotechnology, Mangalayatan University, Aligarh, India; <sup>b</sup>Molecular Medicine Lab, School of Life Science, Jawaharlal Nehru University, New Delhi, India; <sup>c</sup>Department of Biotechnology, Faculty of Life Sciences, Institute of Applied Medicines and Research, Ghaziabad, India; <sup>d</sup>Center of Experimental Medicine & Surgery, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

Communicated by Ramaswamy H. Sarma

### ABSTRACT

Severe acute respiratory syndrome (SARS) is endemic in South China and is continuing to spread worldwide since the 2003 outbreak, affecting human population of 37 countries till present. SARS is caused by the severe acute respiratory syndrome Coronavirus (SARS-CoV). In the present study, we have designed two multi-epitope vaccines (MEVs) composed of cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL) and B cell epitopes overlap, bearing the potential to elicit cellular as well as humoral immune response. We have used truncated (residues 10–153) *Onchocerca volvulus* activation-associated secreted protein-1 as molecular adjuvants at N-terminal of both the MEVs. Selected overlapping epitopes of both the MEVs were further validated for stable molecular interactions with their respective human leukocyte antigen class I and II allele binders. Moreover, CTL epitopes were further studied for their molecular interaction with transporter associated with antigen processing. Furthermore, after tertiary structure modelling, both the MEVs were validated for their stable molecular interaction with Toll-like receptors 2 and 4. Codon-optimized cDNA of both the MEVs was analysed for their potential high level of expression in the mammalian cell line (Human) needed for their further *in vivo* testing. Overall, the present study proposes *in silico* validated design of two MEVs against SARS composed of specific epitopes with the potential to cause a high level of SARS-CoV specific cellular as well as humoral immune response.

**Abbreviations:** Antigen-presenting cell: APCs; AutoDock: ADT; Centers for Disease Control and Prevention: CDC; Cluster of Differentiation: CD4+ or CD8+; Codon Adaptation Index: CAI; Cytotoxic T Lymphocyte: CTL; Ectodomain: ECD; Envelope: E; Global Distance Test: GDT; Helper T Lymphocyte: HTL; Human Leukocyte Antigen: HLA; Immune Epitope Database And Analysis Resource: IEDB; Interferon Gamma: IFN- $\gamma$ ; Lamarckian Genetic Algorithm: LGA; major histocompatibility complex: MHC; Membrane: M; Molecular Dynamics: MD; Motif-Emerging with Classes-Identification: MERCI; multi-epitope vaccines: MEVs; Multiple Sequence Alignment: MSA; Nanoseconds: ns; National Center for Biotechnology Information: NCBI; Neural Network-based method: NN; Non-Human Primates: NHPs; Nucleocapsid: N; *Onchocerca volvulus* activation-associated secreted protein-1: Ov-ASP-1; Open Reading Frame: ORF; Optimized Potential for Liquid Simulations-all-atom force field: OPLS-AA; Pathogenesis-Related-1: PR-1; Peripheral Blood Mononuclear Cells: PBMCs; Picoseconds: ps; Protein Data Bank: PDB; Qualitative Model Energy ANalysis: QMEAN; Receptor Binding Domain: RBD; root-mean-square deviation: RMSD; root-mean-square fluctuation: RMSF; Severe Acute respiratory syndrome: SARS; Severe Acute respiratory syndrome Coronavirus: SARS-CoV; Spike: S; Stabilization Matrix alignment method: SMM; Support Vector Machine: SVM; Toll-Like Receptors: TLRs; Transporter associated with antigen processing: TAP; un-normalized global distance test: uGDT

### ARTICLE HISTORY

Received 4 August 2018  
Accepted November 2018

### KEYWORDS

Severe acute respiratory syndrome; severe acute respiratory syndrome coronavirus; epitope; multi-epitope vaccines; human transporter associated with antigen processing; Toll-like receptors; molecular docking; molecular dynamics simulation

## Introduction

Severe acute respiratory syndrome coronavirus (SARS-CoV) causes a severe form of respiratory disorder called severe acute respiratory syndrome (SARS). SARS infection is highly contagious and is characterized by breathing difficulties, dry cough, pneumonia and high fever. SARS is endemic in south China; in 2003 outbreak, it caused sickness in 8098 people

and death of 774. Eventually, SARS has now spread around the world, infecting people in 37 countries (Booth et al., 2003; CDC Report; Severe Acute Respiratory Syndrome (SARS) 2018; Leung et al., 2003; Zhong et al., 2003). Though there is an urgent need to develop a specific and safe SARS vaccine, till date, no specific vaccine is available. With the situation that the pathogenesis mechanism of SARS-CoV is still largely

unknown, an immunoinformatics approach to thoroughly study, design and develop epitope-based-specific vaccine would be an essential step forward.

Proteome of SARS-CoV consists of several potential drug targets and vaccine candidate proteins. The present study covers screening of potential epitopes from 11 SARS-CoV proteins. These proteins are critically involved in virus–host interaction and pathogenesis. The spike (S) protein of SARS-CoV plays a key role in membrane fusion and viral entry into the host cell (Du et al., 2009; Liao et al., 2015). The envelope (E) protein also plays a determining role in virus–host cell interaction and viral pathogenesis (Jimenez-Guardeño et al., 2014). The membrane (M) protein mediates assembly and budding of viral particles and hence is crucial for viral pathogenesis (Voß et al., 2009). The nucleocapsid (N) protein packages the viral genome into a helical ribonucleocapsid and hence fundamentally involved in viral self-assembly and pathogenesis (Chang et al., 2014; Wei et al., 2012). The ORF (open reading frame) 3a protein is involved in viral replication and hence is a potential target for drug and vaccine design and development (Åkerström et al., 2007; Lu et al., 2009; Zhong et al., 2006). The ORF 3 b protein plays a crucial role in the upregulation of transcriptional activity during pathogenesis, and hence 3 b is also a potential drug and vaccine design target (Åkerström et al., 2007; Varshney et al., 2012). The ORF 7a protein plays an important role during the viral replication cycle (Åkerström et al., 2007; Vasilenko et al., 2010). The ORF 7 b protein plays a crucial role in virus replication as well as enhances virulence (DeDiego et al., 2008; Pfefferle et al., 2009). The ORF 8a protein not only enhances viral replication but also induces host cell apoptosis (Chen et al., 2007). The ORF 8 b protein is involved in replication and induces DNA synthesis (Law et al., 2006). The ORF 9b protein has been observed to be involved in suppressing innate immunity (Shi et al., 2014). All the above-discussed proteins of SARS-CoV are involved crucially in either viral proliferation or host cell pathogenesis and hence are potential targets for drug and vaccine design.

In the present study, we have proposed design for two multi-epitope vaccines (MEVs) composed of screened Cytotoxic T lymphocyte (CTL) epitopes, Helper T lymphocyte (HTL) epitopes and B cell epitopes with potential to elicit humoral as well as cellular immune response. To enhance immune response, truncated (residues 10–153) *Onchocerca volvulus* activation-associated secreted protein-1 (Ov-ASP-1) was used as an adjuvant at N-terminal of both MEVs. Ov-ASP-1 was chosen for its potential to activate antigen-processing cells (APCs) (Guo et al., 2015; MacDonald et al., 2005). To design MEVs, CTL, HTL and B cell epitopes were screened from the earlier discussed 11 SARS-CoV proteins. Screened epitopes were further studied for overlapping consensus regions. Epitopes showing partial or complete overlapping region, and the epitopes with the highest human leukocyte antigen (HLA) allele binders were selected for detailed further studies. The selected CTL and HTL epitopes were studied for their molecular interaction with their respective HLA allele binders by molecular docking and dynamics simulation. Moreover, the selected CTL epitopes were also

validated for their molecular interaction with TAP (transporter associated with antigen processing) transporter cavity to analyse their smooth passage from cytoplasm to the endoplasmic reticulum lumen (Antoniou et al., 2003). Tertiary models of both the designed MEVs were generated and validated by *in silico* methods. Both the MEVs were further screened for discontinuous B cell epitopes as well as IFN- $\gamma$  (interferon gamma) inducing epitopes. Signalling by multiple Toll-like receptors (TLRs) is a necessary component of the innate immune response for SARS-CoV infection. Since rOv-ASP-1 primarily binds APCs among human PBMCs (peripheral blood mononuclear cells) and trigger pro-inflammatory cytokine production via TLR2 and TLR4, hence both the MEVs were further analysed for their molecular interaction with TLR2 and TLR4 by molecular docking and dynamics simulation studies (Delneste et al., 2007; He et al., 2009; Totura et al., 2015; Wang & Liu, 2016). Further, the optimized cDNA of both MEVs was generated and analysed for high expression in the mammalian cell line (human) using *in silico* methods. Workflow chart: *Supplementary Figure S1*.

## Methodology

*SARS-CoV proteins chosen for potential epitope screening.* SARS-CoV proteins that are critically involved in different stages of viral pathogenesis including virus and host cell interaction, viral multiplication and host cell disruption were chosen. In the present study, 11 SARS-CoV proteins were chosen for epitope screening. They include Spike (S) glycoprotein, Envelope (E), Nucleocapsid (N) protein, Membrane (M) protein, ORFs (ORF 3a, -3 b, -7a, -7 b, -8a, -8 b and 9b). Full-length protein sequence of above-mentioned SARS-CoV proteins was retrieved from NCBI database (National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/protein>)). A total of 1182 protein sequences of chosen SARS-CoV proteins belonging to different strains and regions of origins were retrieved. To perform structural-based epitope screenings, available 3D structures of chosen SARS-CoV proteins were retrieved from Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/home/home.do>), and for those with no 3D structure available at PDB, 3D homology models were generated by Swiss model (<http://swissmodel.expasy.org/>) (Arnold et al., 2006) (*Supplementary Table S1*).

## Screening of potential epitopes

### Prediction of T cell epitope

*CTL epitope screening.* IEDB (Immune Epitope Database And Analysis Resource) tool ‘Proteasomal cleavage/TAP transport/MHC class I combined predictor’ (<http://tools.iedb.org/processing/>) was used to screen the CTL epitopes (Hoof et al., 2009; Peters et al., 2003; Tenzer et al., 2005). The total score generated for every CTL epitope gives a combined scoring of the proteasomal, TAP (N-terminal interaction), MHC (major histocompatibility complex) and processing analysis scoring. Scoring is performed by six different methods, the Consensus, NN-align, SMM-align, Combinatorial library,

Sturniolo and the NetMHCIIpan. Further, the IC(50) (nM) value of the screened epitopes was also predicted by this IEDB tool. Epitopes IC(50) values are predicted as high (<50 nM), intermediate (<500 nM) and of least affinity (<500 nM) for different epitopes. Moreover, the immunogenicity of all the screened CTL epitopes was also obtained using 'MHC I Immunogenicity' (Calis et al., 2013) tool of IEDB (<http://tools.iedb.org/immunogenicity/>). The tool predicts immunogenicity of a particular peptide on the basis of physicochemical properties of the constituent amino acid and their position within the epitope amino acid sequence.

**HTL epitopes screening.** The IEDB tool 'MHC-II Binding Predictions' (<http://tools.iedb.org/mhcii/>) was used to screen HTL epitopes from above-mentioned 11 SARS-CoV proteins. The tool involves three different methods (combinatorial library, SMM\_align and Sturniolo) to generate percentile rank for each epitope (Nielsen et al., 2007; Sidney et al., 2008; Sturniolo et al., 1999; Wang et al., 2010). The tool compares the score of epitope against the scores of other random five million 15 mer peptides from SWISSPROT database. The rank for the consensus of all three methods was generated in terms of the percentile rank.

**Population coverage by the screened CTL and HTL epitopes.** Human population coverage by the screened and shortlisted 32 CTL and 36 HTL epitopes was performed by the 'Population Coverage' tool of IEDB (<http://tools.iedb.org/population/>) (Bui et al., 2006). The population coverage by the combined use of CTL and HTL epitopes would have larger world population coverage.

### Prediction of B cell epitope

**Protein sequence-based B cell linear epitope prediction.** Linear B cell epitopes screening was performed by 'B Cell Epitope Prediction Tools' tool of IEDB server (<http://tools.iedb.org/bcell/>). Six different methods available at IEDB server were used. These methods includes 'Bepipred Linear Epitope Prediction' (propensity scale method like hidden Markov model), 'Chou & Fasman Beta-Turn Prediction', 'Emini Surface Accessibility Prediction', 'Karplus & Schulz Flexibility Prediction', 'Kolaskar & Tongaonkar Antigenicity' and 'Parker Hydrophilicity Prediction' (Chou & Fasman, 1978; Emini et al., 1985; Karplus & Schulz, 1985; Kolaskar & Tongaonkar, 1990; Larsen et al., 2006; Parker et al., 1986). These methods are based mainly on propensity scale method as well as physicochemical properties of SARS-CoV antigenic protein sequence.

**Protein structure-based linear and discontinuous B cell epitope prediction.** Structure-based B cell epitopes from SARS-CoV proteins were screened using DiscoTope2.0 (DiscoTope: Structure-based Antibody Prediction tool; <http://tools.iedb.org/discolope/>) and Ellipro (Ellipro: Antibody Epitope Prediction tool; <http://tools.iedb.org/ellipro/>) tools available at IEDB (Kringelum et al., 2012; Ponomarenko et al., 2008). DiscoTope2.0 screens discontinuous epitopes while Ellipro screens both discontinuous and linear epitopes. DiscoTope2.0 is based on the number of contact of one C $\alpha$  with other C $\alpha$  within a range of 10 Å in the 3D structure of the protein and the propensity of a residue to be a part of an epitope.

ElliPro method is based on the location of the residue within an ellipsoid covering the percentage of inner core residues of the protein. The Protrusion Index score of a residue lying outside of the ellipsoid covering 90% of the inner core residues will be 0.9. The discontinuous epitopes prediction by ElliPro is based on the distance  $R$  (Å) between the centre of mass of two residues lying outside the largest possible ellipsoid. The larger value of  $R$  indicates larger discontinuous epitopes being predicted.

### Characterization of screened epitopes

**Epitope conservation analysis.** Conservancy analysis of all the shortlisted CTL and HTL epitopes was performed by 'Epitope Conservancy Analysis' tool of IEDB (Bui, Sidney, Li, Fusseeder, & Sette, 2007) (<http://tools.iedb.org/conservancy/>). The tool computes the percentage of epitope sequence conservancy amongst a given set of protein sequences at a given identity level. In our study, we analysed epitopes for 100% conservancy/identity level amongst numerous SARS-CoV protein sequences from different strains and origins retrieved from NCBI protein database ([Supplementary Table S1](#)). With this analysis, we could elucidate the percentage of SARS-CoV protein sequences carrying the screened epitopes with full-length conservancy (100% identity). The higher the percentage of SARS-CoV strains carrying the screened epitope in their protein sequences, the greater would be its effectiveness as a vaccine candidate epitope ([Supplementary Tables S2, S3 and S5](#)).

**Toxicity prediction for the screened epitopes.** The ToxinPred tool ([http://crdd.osdd.net/raghava/toxinpred/multi\\_submit.php](http://crdd.osdd.net/raghava/toxinpred/multi_submit.php)) (Gupta et al., 2013) was used for the prediction of possible toxicity of all the screened CTL, HTL and B cell epitopes. The tool allows identifying highly toxic or non-toxic short peptide sequences. Toxicity analysis was performed by 'SVM (Swiss-Prot) based' (support vector machine) method.

**Epitope overlapping residue analysis.** Multiple sequence alignment (MSA) analysis was performed to analyse the overlapping residue sequence amongst the screened CTL, HTL and B cell epitopes. The analysis was performed by Clustal Omega tool (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) of EBI (European Bioinformatics Institute) (Sievers et al., 2011). MSA by Clustal Omega virtually aligns a multiple number of protein sequences and generates accurate alignment.

### Selection of epitope for molecular interaction study with HLA allele binders and TAP transporter

Few CTL and HTL epitopes were chosen for further analysis of their molecular interaction with their respective HLA allele binders and TAP transporter cavity. The chosen epitopes either had sequence overlap amongst all three types of epitopes (CTL, HTL and B cell) or had complete overlap amongst any two types of epitopes, or had the highest number of their respective HLA allele binders.

## Molecular interaction analysis of selected epitopes with their respective HLA allele binders

**3D structural modelling of HLA alleles and selected epitopes.** Homology models of HLA class I and II alleles showing binding affinity with selected epitopes were done using Swiss-model (Arnold et al., 2006). Amino acid sequences of HLA class I and II alleles were retrieved from the Immuno Polymorphism Database (IPD-IMGT/HLA) (<https://www.ebi.ac.uk/ipd/imgt/hla/allele.html>) (Robinson et al., 2014). Template showing high sequence identity was chosen for homology modelling. The models generated were validated by QMEAN (Qualitative Model Energy ANalysis) for their quality. QMEAN value is a composite (both global and local (i.e. per residue structures) quality estimate of a tertiary protein model (Berkert et al., 2008). Out of the generated models, models with acceptable QMEAN value with a cut-off of -4.0 were chosen for further studies (Supplementary Table S8).

Tertiary structures of the selected epitopes were generated by the 'Natural Peptides Module for Beginners' module of PEPstrMOD ([http://osddlinux.osdd.net/raghava/peptrmod/nat\\_ss.php](http://osddlinux.osdd.net/raghava/peptrmod/nat_ss.php)) (Singh et al., 2015). The tertiary model prediction was done with a simulation time window of 100 picoseconds (ps) in a peptide environment of vacuum.

**Molecular docking and molecular dynamics (MD) simulation study of selected epitopes and HLA alleles binders.** Molecular interaction between selected epitopes and their respective HLA alleles binders was done by molecular docking. Molecular docking was performed by AutoDock Tool 4.2 (ADT) and AutoDock Vina (Morris et al., 2009; Trott & Olson, 2010). For molecular docking, AutoDock Tool 4.2 was utilized to prepare the epitope ligand and the HLA allele receptor. Ligand epitopes and receptor HLA allele molecules were prepared and saved in pdbqt format with polar hydrogen added and all HOH molecules removed. Grid for docking was prepared to cover entire HLA molecule to search the most potential site of interaction and fitting for ligand epitope. Autodock Vina was utilized to run molecular docking by the Lamarckian Genetic Algorithm (LGA), and the epitope–HLA allele complexes with low binding free energy were generated (Morris et al., 1998).

Further to analyse the stability of the complex formed by molecular docking, MD simulation was performed for epitope–HLA allele complexes using Gromacs 5.1.4 with Optimized Potentials for Liquid Simulations-all-atom force field (OPLS-AA) (Abraham et al., 2015; Jorgensen et al., 1996). MD simulation was performed by placing the epitope–HLA allele complexes at least 1.0 nm from the edge of a cubic box filled with water as solvent and with eight negative ions neutralizing the net charge on the protein complex. The protein complex within the cubic box was further energy-minimized and equilibrated at an invariable temperature of 300 K and a pressure of 1 bar. Further, the MD simulation was run for 2–3 ns to attain a stable root-mean-square deviation (RMSD) relative to the epitope–HLA allele complex structure present in the minimized, equilibrated system. Further, the radiiuses of gyration  $R_g$  as well as the root-mean-square fluctuation (RMSF) for every atom of the epitope–HLA allele complexes were also analysed.

## Molecular docking study of selected CTL epitopes and TAP transporter cavity

The AutoDock Vina tool was used for molecular docking study between the shortlisted CTL epitopes and TAP transporter cavity (Morris et al., 2009; Trott & Olson, 2010). To be more accurate in our study, the available cryo-EM structure of TAP (PDB ID: 5u1d) (Oldham, Grigorieff, & Chen, 2016) was used.

## Design, characterization and molecular interaction analysis of MEVs with TLR2 and TLR4 immunoreceptors

**Design of MEVs.** The two MEVs were designed consisting of the screened and shortlisted CTL and HTL epitopes. The short stretches of amino acids EAAAK and GGGGS were used as rigid and flexible short peptide linkers, respectively (Figure 1(A,B)). The peptide linker involving EAAAK has very favourable property to effectively separate bi-functional domains. The peptide linker involving GGGGS allows interaction between domains, hence improving and restoring folding (Arai, Ueda, Kitayama, Kamiya, & Nagamune, 2001; Bai & Shen, 2006; Chen, Zaro, & Shen, 2013; Hajighahramani et al., 2017; Hu et al., 2004; Huston et al., 1988). Truncated Ov-ASP-1 protein (residues 10–153) was used as an adjuvant at N-terminal of MEVs, to enhance the immune response. Ov-ASP-1 was used as adjuvant has already been reported to induce high anti-RBD (Receptor Binding Domain) of SARS antibody titres in non-human primates (NHPs). At the same time, it has also been reported earlier that the truncated rOv-ASP-1 containing the core pathogenesis-related-1 (PR-1) domain exhibits adjuvanticity similar to that of the full-length Ov-ASP-1 (Guo et al., 2015; MacDonald et al., 2005). Hence, truncated rOv-ASP-1 was chosen here as adjuvant for the design of MEVs.

## Characterization of designed MEVs

**MEV IFN- $\gamma$  inducing epitope prediction.** IFN- $\gamma$  plays an important role in both adaptive and innate immune responses. IFN- $\gamma$  stimulates both macrophages and natural killer cells. MEV epitopes having the potential to induce the release of IFN- $\gamma$  from CD4+ T cells were predicted using the tool 'IFNepitope' (<http://crdd.osdd.net/raghava/ifnepitope/scan.php>) by 'Motif and SVM hybrid' method, i.e. applying MERCI: Motif-EmeRging with Classes-Identification, and SVM: support vector machine approach. A dataset of IFN- $\gamma$  inducing and non-inducing MHC class-II binders is used by IFNepitope tool for IFN- $\gamma$  epitope prediction (Dhanda, Vir, & Raghava, 2013).

**Allergenicity and antigenicity of designed MEVs.** Both CTL and HTL MEVs were analysed for their allergenicity using the tool AlgPred (<http://crdd.osdd.net/raghava/algpred/submit.html>) (Saha & Raghava, 2006). Allergenicity prediction by AlgPred is based on the similarity of the query protein with already known epitope. Antigenicity of both the MEVs was analysed by VaxiJen (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) (Doytchinova & Flower, 2007). The

(A) CTL epitopes		(B) HTL epitopes	
EAAAK	MEAAAK		
AGFLVLYALYLFLQG	YLYALYLF		
GGGGS	GGGGS		
TMSLYMWPBKFTTS	SLYMASPKF		
GGGGS	GGGGS		
SPLFLVALVFLL	LTMASPKF		
GGGGS	GGGGS		
PLFLVALYLFLIC	PLFLVALYLFLIC		
GGGGS	GGGGS		
LFLIVAVLFLCF	ALTCISTH		
GGGGS	GGGGS		
FLIVALFLFLFT	TLFLFLFL		
GGGGS	GGGGS		
LIVALFLFLCTI	SLCSOCTV		
GGGGS	GGGGS		
TLDFYCLFLFLFL	ALLARCYWL		
GGGGS	GGGGS		
LDFYCLFLFLFL	TYFLVPLFLV		
GGGGS	GGGGS		
IDFYCLFLFLFLV	DPYLCFLAFLFLV		
GGGGS	GGGGS		
FYLCFLAFLFLV	FYLCFLAFLFLV		
GGGGS	GGGGS		
YLCFLAFLFLV	YLCFLAFLFLV		
GGGGS	GGGGS		
LCFLAFLFLVLL	LCFLAFLFLVLL		
GGGGS	GGGGS		
CFLAFLFLVLL	TPSGTWTLY		
GGGGS	GGGGS		
FLAFLFLVLL	LQDLEFLFLY		
GGGGS	GGGGS		
MKLLVLTCSLCS	IFFKQSYIF		
GGGGS	GGGGS		
ALGKVAPPHWHTWV	NAFNCNTFEY		
GGGGS	GGGGS		
VVPALFLIDPQIQL	LYN1STFSFF		
GGGGS	GGGGS		
NSVLLFLAFVFLVY	STFSTFCY		
GGGGS	GGGGS		
SVLLFLAFVFLVY	CYMLNDYGF		
GGGGS	GGGGS		
VLLFLAFVFLVY	WFLNDYGFY		
GGGGS	GGGGS		
LFLFLAFVFLVY	RYSHTDNVVF		
GGGGS	GGGGS		
LFLAFVFLVY	LITDDMMIAY		
GGGGS	GGGGS		
NLVIGFLFLWML	LTONDIAHY		
GGGGS	GGGGS		
GTRNNINNATVQL	QIPFANQIMAY		
GGGGS	GGGGS		
MFLFLFLTSQSD	IFFMOMAY		
GGGGS	GGGGS		
GYOPYRVLFLFELL	FAMONAYRF		
GGGGS	GGGGS		
YQPYRVVLSFELLIN	AYPREGVVF		
GGGGS	GGGGS		
PYRVVLSFELLINAP	FREGEVVF		
GGGGS	GGGGS		
YRVVLSFELLINAPA	KWPKWVWLG		

**Figure 1.** Design of multi-epitope vaccines. Selected (A) CTL and (B) HTL epitopes were linked by short peptide linker GGGGS. The Ov-ASP-1 (residues 10–153) was used as an adjuvant at N-terminal of both MEVs, linked by a short rigid peptide linker EAAAK. Epitopes from different proteins are coloured in different colours.

VaxiJen prediction for antigenicity is based on alignment-free approach involving physicochemical properties of the query proteins sequence.

*Physicochemical property analysis of designed MEVs.* Physicochemical properties of both the MEVs were analysed by the ProtParam tool (<https://web.expasy.org/protparam/>) (Gasteiger et al., 2005). ProtParam analysis generates the empirical parameters based on the protein's amino acid sequence.

### Tertiary structure modelling and refinement of MEVs

The tertiary structure of both CTL and HTL MEVs was generated by homology modelling using the RaptorX structure prediction server (<http://raptorg.uchicago.edu/StructurePrediction/predict/>) (Källberg et al., 2012). The quality assessment of generated homology models of CTL and HTL MEVs was done by their *P*-value. The *P*-value for a predicted model is the probability of the model being worse than the best. *P*-value indicates the relative quality of the model in terms of modelling error, combining global distance test (GDT) and un-normalized GDT (uGDT) indicating the error at each residue. The smaller *P*-value indicates the greater quality of a predicted model.

Refinement of the generated MEVs models was performed by ModRefiner (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>) (Xu & Zhang, 2011) and GalaxyRefine (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) (Ko, Park, Heo & Seok, 2012). ModRefiner refines the initial model for both local and global distortions by a two-step atomic-level energy minimization. Further, GalaxyRefine refines the submitted tertiary structure by repeated structure perturbation as well as with subsequent structural relaxation by MD simulation. To avoid breaks in model structures, GalaxyRefine uses triaxial loop closure method. The MolProbity value generated for a refined model is the log-weighted combination of clash score, the percentage of Ramachandran not favoured residues and the percentage of bad side-chain rotamers.

*Validation of CTL and HTL MEVs refined models.* Refined 3D models of CTL and HTL MEVs were validated by RAMPAGE analysis (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) (Lovell et al., 2003). The generated Ramachandran plot

shows the energetically allowed and disallowed residues with their dihedral angles psi ( $\psi$ ) and phi ( $\phi$ ).

*Discontinuous B-cell epitope prediction from MEVs.* Both the designed and modelled CTL and HTL MEVs were analysed for the presence of potential discontinuous B cell epitopes. The prediction was carried out by ElliPro tool of IEDB (Ponomarenko et al., 2008).

### Molecular docking and MD simulation analysis of MEVs with TLR2 and TLR4 immunoreceptors

Molecular interaction analysis of the designed CTL and HTL MEVs with immunological receptors TLR2 and TLR4 was performed. Molecular docking study was performed for both MEVs with TLRs using PatchDock server (<http://bioinfo3d.cs.tau.ac.il/PatchDock/>) (Duhovny et al., 2002; Schneidman-Duhovny et al., 2005). Tertiary structures of human TLRs ectodomain (ECD) were retrieved from PDB databank (PDB ID: 2Z7X (TLR2), 4G8A (TLR4)) (Jin et al., 2007; Ohto et al., 2012). The PatchDock algorithm first computes the molecular surface and then detects geometric patches (concave, convex and flat surface pieces), and the patches with 'hot spot' residues are retained. The candidate complexes with acceptable penetrations of the atoms of the ligand to the atoms of the receptor are chosen and ranked according to a geometric shape complementarity score.

The MD simulation was performed to analyse the stable complex formation for MEVs-TLRs complexes by Gromacs 5.1.4 using Optimized Potential for Liquid Simulations-all-atom force field (OPLS-AA) (Abraham et al., 2015; Jorgensen et al., 1996). MD simulation was performed by placing the MEV-TLR protein complexes at least 3.0 nm from the edge of a cubic box filled with water as solvent and with eight negative ions neutralizing the net charge on the protein complex. The protein complex within the cubic box was further energy-minimized and equilibrated at an invariable temperature of 300 K and a pressure of 1 bar. Further, the MD simulation was run for up to 6 ns to attain stable RMSD relative to the MEV-TLR complex structure present in the minimized, equilibrated system. Further, the radiiuses of gyration  $R_g$  as

well as the RMSF (fluctuation) for every atom of the MEV-TLR complex were also analysed.

### cDNA analysis of CTL and HTL MEVs

Mammalian cell line expression system (Human) optimized cDNA for both CTL and HTL MEVs was generated by Codon Usage Wrangler Tool (<http://www.mrc-lmb.cam.ac.uk/ms/methods/codon.html>). The generated cDNA was further analysed for high expression in mammalian cell line by GenScript Rare Codon Analysis Tool (<https://www.genscript.com/tools/rare-codon-analysis>). The analysis was based on the GC content, Codon Adaptation Index (CAI) and the Tandem rare codon frequency (Morla et al., 2016; Nezafat et al., 2017; Wu et al., 2010).

## Results and discussion

### Screening of potential epitopes

#### Prediction of T cell epitope

**CTL epitope screening.** The CTL epitopes with highest 'Total Score', and acceptably low IC(50) (nM) value of epitope–HLA class I allele pairs and with the larger number of HLA class I allele binders, were chosen. The screened epitopes have shown predicted affinity with mostly A and B types of HLA class I alleles. Immunogenicity of the chosen CTL epitopes was also determined; the higher immunogenicity score shows the greater immunogenic potential of the epitope (Supplementary Tables S2 and S6). From 11 SARS-CoV proteins, a total of 32 CD8+ T cell epitopes were chosen.

**HTL epitopes screening.** The HTL epitopes were chosen on the basis of 'Percentile rank' and the higher number of HLA class II allele binders. Small-numbered percentile rank shows higher peptide–HLA allele binding affinity. The screened epitopes have shown predicted affinity with mostly DPA1, DPB1, DQA1, DQB1, DRB1, DRB3 and DRB5 HLA class II allele types. From 11 SARS-CoV proteins, a total of 36 CD4+ T cell epitopes were chosen (Supplementary Tables S3 and S7).

Several screened CTL and HTL epitopes by *in silico* methods in our study show a match with previous experimental studies as indicated by IEDB epitope ID (Supplementary Tables S2 and S3).

**Population coverage by the screened CTL and HTL epitopes.** Most of the world geographical regions were included in this study. From this study, we may conclude that by joint administration of all the chosen CTL and HTL epitopes, 85.21% of the world population could be covered with a standard deviation of 25.88 (Supplementary Table S4).

**Prediction of B cell epitope.** Protein sequence-based B cell linear epitope prediction. By Bepipred Linear Epitope Prediction method, a total of 78 B cell epitopes with the length between 4 and 20 amino acids were screened from 11 SARS-CoV proteins. Out of these 78 epitopes, 14 epitopes have shown to have significant amino acid sequence overlap with shortlisted CTL and HTL epitopes. Moreover, these 14 epitopes also show an overlap of amino acid sequence with the B cell epitopes predicted by other five B cell epitope prediction methods based on different physicochemical properties (Supplementary Table S5) (Figures 2 and 3). Hence, these 14 B cell epitopes were shortlisted.

Overlapping regions of predicted B cell Bepipred linear Epitope and epitopes predicted by other protein sequence and structure based methods		Overlapping regions of sequence based B cell linear epitope prediction					Overlapping regions of structure based B cell epitope prediction		
		Chou & Fasman Beta-Turn	Emini Surface Accessibility	Karplus & Schulz Flexibility	Kolaskar & Tongaonkar Antigenicity	Parker Hydrophilicity	DiscoTope: discontinuous epitopes	ElliPro	Linear
3a	65 AISPKFTTSL 74	65-74		67-74	73-74	65-74			
7b	1 MNELTLL 7	1-7	1-7	1-7		1-7			
	37 VQTCTPNVTINCQDPAGGAL 56	37-56		37-51		37-56		46-51	46-51
8b	1 MDPNOTNIVP 10	1-10		1-7	6-7	1-10		9-10	9-10
9b	29 DAMQQGQNSADPKV 42	29-42	33-41	33-38	40-42	29-42	38-40	29-41	29-41
9b	145 DHIGTRNPNN 155	145-155		145-155		145-155			
Spike	74 HTFGNPVIPFK 84	74-84		74-84		74-84		74-84	74-84
Spike	88 YFAATEKSNV 97	88-97		88-97		88-97		88-97	88-97
Spike	481 YGFYTTGIGYQ 492	481-492		482-492		481-492	484-492	481-492	481-492
Spike	508 ATVCGPPLSTD 518	508-518		509-518	508-514	508-518		514-518	508-518
Spike	620 RIYSTGNNVFQ 631	620-631		620-631		620-631			
Spike	680 SLGADSSIAYSNNNTIAIP 697	680-697		680-697		680-697	685-688	683-697	683-697
Spike	787 ILPDPLKPTKRS 798	787-798	787-798	790-798	787-791	787-798	792-794	787-797	787-798
Spike	1068 KAYFP 1072	1068-1072	1068-1072	1068-1072		1068-1072		1068-1072	1068-1072

**Figure 2. Overlapping regions amongst the B cell epitopes predicted by the BepiPred method and seven other different B cell epitopes prediction methods.** B cell epitopes predicted by the BepiPred method and other different sequence-based (Chou, Emini, Karplus, Kolaskar, and Parker.) and tertiary structure-based (DiscoTope and ElliPro) prediction methods were found to have significant consensus overlapping regions. Consensus overlapping regions of BepiPred epitopes are underlined by the different colour, corresponding to respective prediction method.

Clusters of overlapping CTL, HTL & B cell epitopes	
Orf 3a	YLALYIYL MEAOFLY QFLYVALIYFLOCI AQFLYVALIYFQC 101-118 MEAGFLYVALIYFQC
Orf 3b	LYMAISPKF SLYMAISPKF TMSLYMAISPKEITS AISPEKFTS 59-74 TMSLYMAISPKEITS
Orf 7a	SPLFLIVAALVFLIL PLFLIVAALVFLILC LFPLFLIVAALVFLILCF FLIVAALVFLILCFT LIVAALVFLILCFTI 99-117 SPLFLIVAALVFLILCFTI
Orf 7b	TUDFYLCFL CFLAFLLFLVLMIL LCFLAFLLFLVLMIL YLCLAFLLFLVLMIL EVLCFLAFLLFLVLMIL DPLFCFLAFLLFLVLMIL IDFYLCFLAFLLFLVLMIL TUDFYLCFLAFLLFLVLMIL MNELTU 1-27 MNELTUDFYLCFLAFLLFLVLMILII
Orf 8a	SLCSICITV MKLLIVLTCISLSC 1-19 MKLLIVLTCISLSCSICITV
Orf 8b	ALGKVLPFHRWHTMV VQTCTPNVTINCQDPAGGL 23-63 ALGKVLPFHRWHTMVQTCVPNVTINCQDPAGGLARCWYL
Orf 9b	VVPPALHLDPQIQL MDPNQTNVWP 1-50 MDPNQTNVWPALHLDPQIQLTIRMEDAMGQQGNSADPKVYPIILR DAMGQQGNSADPKVYPIILR KVYPILRL
Envelope	LAFPVFLVLTIAIT LAFPVFLVLTIAIT LAFPVFLVLTIAIT VLLFLAFVFLVLTIAIT SVLLFLAFVFLVLTIAIT NSVLLFLAFVFLVLTIAIT 15-35 NSVLLFLAFVFLVLTIAIT
Membrane	WNLVIGFLFLAWIML WNLVIGFLFLAWIML 19-34 WNLVIGFLFLAWIML MWLSYFVASE 90-102 MWLSYFVASEFILE
Nucleocapsid	GTRNPNNNAATVQL DHIGTRNPNNNAATVQL 145-162 DHIGTRNPNNNAATVQL
Spike	IPFKDGFIY HTFGNPVPIP HTFGNPVPIP 74-97 HTFGNPVPIP YFAATEKSNV LYNSTFFSTFKCY 355-367 LYSTFFSTFKCY STFFSTFKCY WPLNDYGFY CYWPLNDYGF YRVVVLSEELLNAPA PYRVVVLSEELLNAP CYRVVVLSEELLNAP YOPYRVVVLSEELLN GYOPYRVVVLSEELLN 474-518 CYWPLNDYGPYTTGIGYQ ATVCGPKLSTD 618-631 RIYSTGNVNF RIYSTGNVNFQ RIYSTGNVNFQ
Spike	LGADSSIA SLGADSSIASSNNNTIAIP SLGADSSIASSNNNTIAIP 680-697 ILPDPLPKTKRS ILPDPLPKTKRSFIEDLFF RSFIEDLFF 846-855 LTDDMIAAY LTDDMIAAY 877-888 FAMQMAYRF IPPFAMQMAYRF QIPPFAMQMAYRF 1068-1079 KAYEP KAYEP FPRECVVF AYFPREGVFV KAYEP 1068-1079 KAYEPREGVFV

**Figure 3.** Overlapping CTL, HTL and B cell epitopes. Multiple sequence alignment using Clustal Omega at EBI was used to identify the consensus overlapping regions of CTL (red), HTL (blue) and B cell epitopes (green) amongst selected epitopes. Epitopes with overlapping region amongst all three types of epitopes, epitopes with full sequence overlap and epitopes with the highest number of HLA allele binders were chosen for further studies (encircled).

**Protein structure-based linear and discontinuous B cell epitope prediction.** Both the discontinuous and linear epitopes were predicted by protein structure-based methods, DiscoTope 2.0 and Ellipro. Sequence-based Bepipred Linear Epitopes have shown to be in significant amino acid sequence overlap with screened structure-based epitopes (Figure 2). This result confirms that the shortlisted 14 Bepipred Linear Epitopes are the potential B cell epitopes predicted from 11 SARS-CoV proteins.

### Characterization of screened epitopes

**Epitope conservation analysis.** The conservancy analysis of chosen 32 CTL, 36 HTL and 14 BepiPred linear B cell epitopes was performed. This study shows that the 100% amino acid sequence conservancy of CTL epitopes varies from 60.67 to 100%, and that of HTL epitopes varies from 60 to 100%, with

exception for HTL epitopes from ORF 7a protein having 25% conservancy; likewise, conservancy for B cell epitope varies from 60.67 to 100% (Supplementary Tables S2, S3 and S5). With these results, we may conclude that the amino acid sequences of the chosen CTL, HTL and B cell epitopes are highly conserved.

**Toxicity prediction for the screened epitopes.** Toxicity analysis result reveals that all the shortlisted CTL, HTL and B cell epitopes are non-toxic in nature (Supplementary Tables S2, S3 and S5). This analysis was based on the toxic peptide dataset of ToxinPred consisting of 1805 toxic peptides.

**Epitope overlapping residue analysis.** CTL, HTL and B cell shortlisted epitopes were analysed to have significant overlapping amino acid sequence region. The study was performed by the MSA analysis using Clustal Omega tool (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) of EBI (European Bioinformatics Institute) (Figure 3).

### Selection of epitope for molecular interaction study with HLA allele binders and TAP transporter

CTL and HTL epitopes showing to have amino acid sequence overlap with all three (CTL, HTL and B cell) types of epitopes or having complete sequence in consensus with any of the other two types of epitopes, or having highest number of HLA allele binders, were chosen for their molecular interaction with their respective HLA allele binders and TAP transporter cavity (Figure 3). These include CTL epitopes: <sup>5</sup>TLIDFYLCFL<sup>14</sup>, <sup>11</sup>SLCSCICTV<sup>19</sup>, <sup>61</sup>SLYMAISPKE<sup>70</sup>, <sup>62</sup>LYMAISPKE<sup>70</sup>, <sup>101</sup>MEAQFLYLY<sup>109</sup>, <sup>107</sup>YLYALIYFL<sup>115</sup>, <sup>474</sup>CYWPLNDYGF<sup>483</sup>, <sup>476</sup>WPLNDYGFY<sup>484</sup>; and HTL epitopes: <sup>1</sup>MKLLIVLTCISLCSC<sup>15</sup>, <sup>5</sup>TLIDFYLCFLAFLLF<sup>19</sup>, <sup>59</sup>TMSLYMAISPKFTTS<sup>73</sup>, <sup>103</sup>AQFLYLYALIYFLQC<sup>117</sup>, <sup>104</sup>QFLYLYALIYFLQCI<sup>118</sup>, <sup>490</sup>GYQPYRVVVSFELL<sup>504</sup> and <sup>491</sup>YQPYRVVVSFELLN<sup>505</sup>.

### Molecular interaction analysis of selected epitopes with their respective HLA allele binders

**Molecular docking and MD simulation study of selected epitopes and HLA allele binders.** Molecular docking study of all selected CTL and HTL epitopes with their respective HLA allele binders was performed. The study shows a significantly favourable molecular interaction with negative binding energy and one or more than one hydrogen bond formation (Figure 4(A,B)). To analyse the stability of the molecular interaction of the docking complexes, MD simulation was performed. Analysis time window was kept at 3 ns at the reasonably invariable temperature (~300 K) and pressure (~1 bar). The result shows a reasonably stable RMSD throughout the analysis window, with fluctuation between ~0.3 and 0.8 nm, indicating stable complex formation (Figure 5(A,B)). Furthermore, the stable  $R_g$  (radius of gyration) (Supplementary Figure S2(A,B)) and RMSF for all the atoms in the epitope–HLA allele complexes (Supplementary Figure S3(A,B)) indicate that the complexes remain very stable and are in their folded form. B-factor analysis of the epitope–HLA allele complexes shows that most of the regions of complexes are stable (blue) with a very small region being acceptably fluctuating (yellow and orange) (VIBGYOR colour presentation) (Supplementary Figure S4(A,B)).

### Molecular docking study of selected CTL epitopes and TAP transporter cavity

Molecular docking study was performed to analyse the molecular interaction of CTL epitopes with the TAP cavity. The results show a favourable interaction of CTL epitopes at different sites of the TAP cavity. Two sites of CTL epitope interaction within the TAP cavity are shown in Figure 6. One of the CTL epitope interaction sites within the TAP cavity is close to the cytoplasm and another site is close to the ER lumen. The results show a significant molecular interaction of CTL epitopes within the TAP cavity having significantly negative binding energy with one or more than one hydrogen bond formation. From this study, we may predict a smooth passage for selected CTL epitopes through the TAP transporter cavity favouring the transport of epitopes from the cytoplasm to the ER lumen (Figure 6).

### Characterization and molecular interaction analysis of MEVs with TLR2 and TLR4 immunoreceptors

#### Characterization of designed MEVs

**MEV IFN- $\gamma$  inducing epitope prediction.** The IFN epitope screening reveals CTL and HTL MEVs to contain 592 and 848 INF- $\gamma$  epitopes, respectively. Out of all the scanned INF- $\gamma$  epitopes, a total of 68 epitopes from CTL MEV and 31 epitopes from HTL MEV were found to be POSITIVE epitopes, with the score of 1 or more than 1 (Supplementary Table S9) (Figure 7(C,G)).

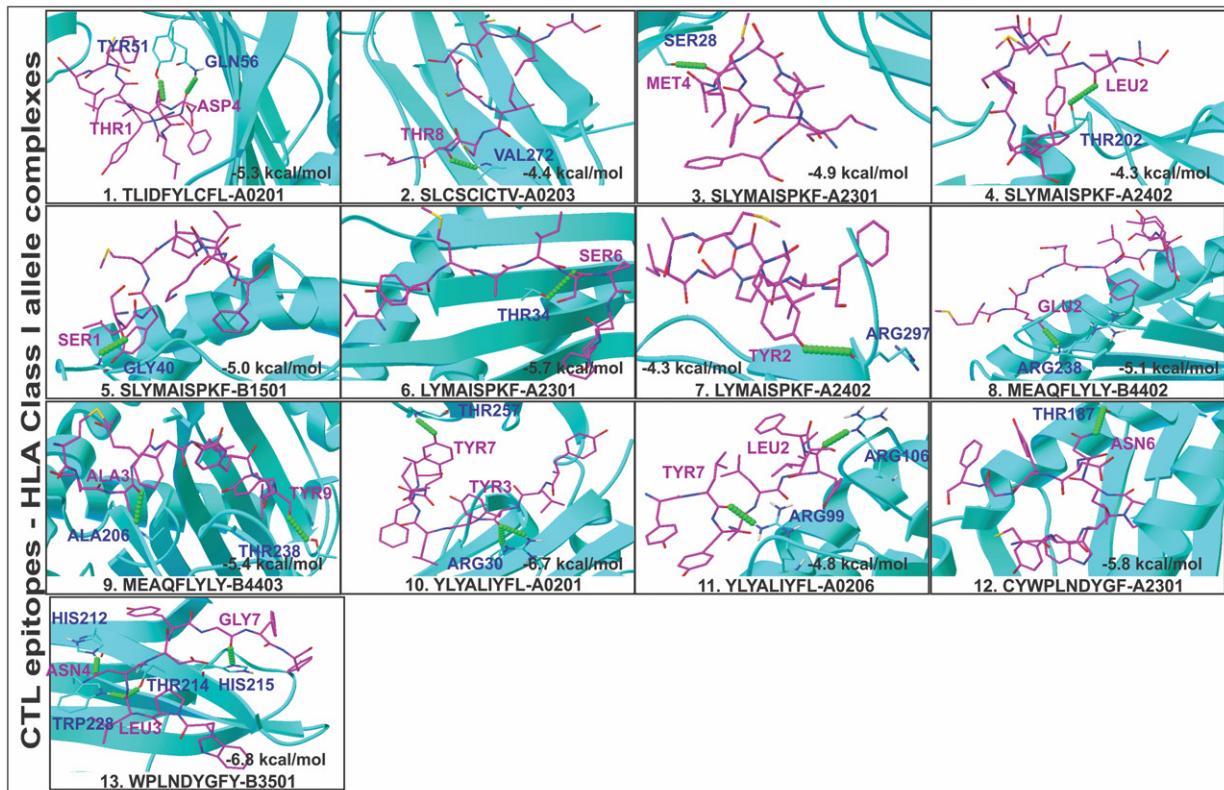
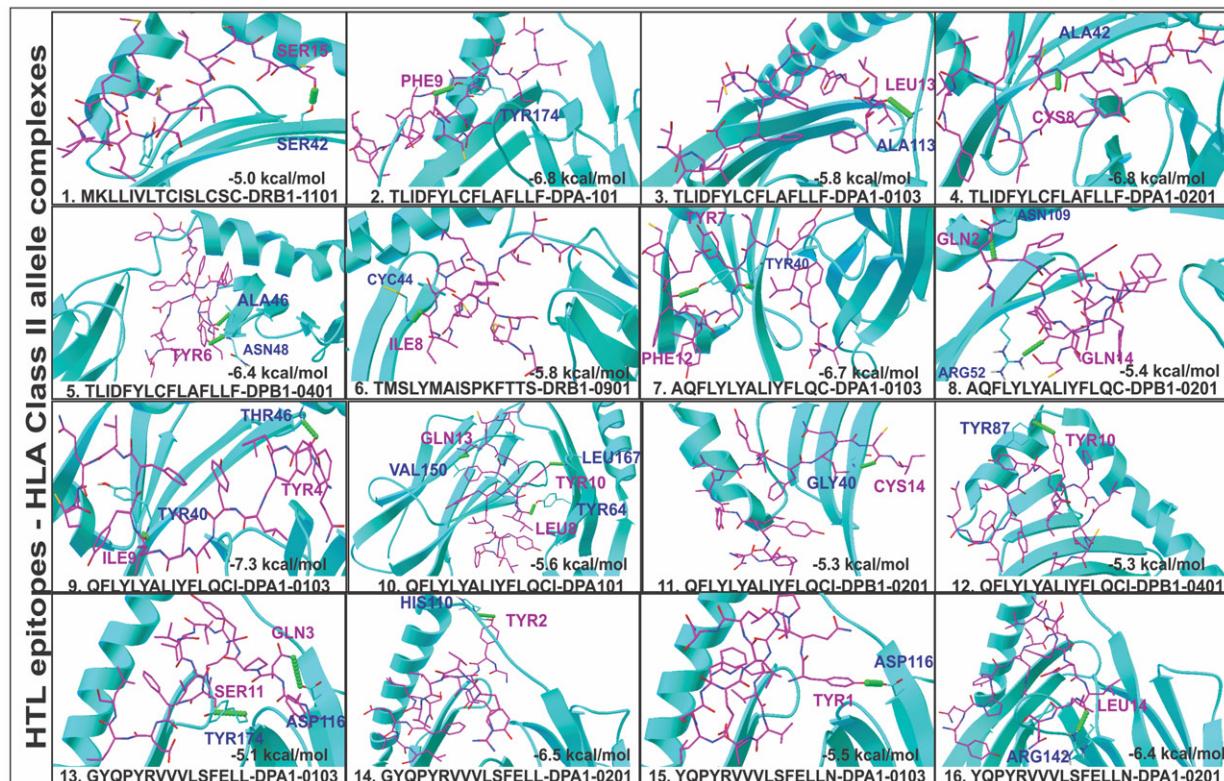
**Allergenicity and antigenicity of designed MEVs.** AlgPred analysis score of -0.71920478 and -1.1649757 for CTL and HTL MEVs, respectively (default threshold being -0.4), predicts both the MEVs to be NON ALLERGEN. The antigenicity analysis for both the MEVs revealed them to be ANTIGENS with a prediction score of 0.4796 (CTL MEV) and 0.5600 (HTL MEV), default threshold being 0.4 for viral proteins. Hence, both the CTL and HTL MEVs are predicted to be non-allergic as well as antigenic in nature.

**Physicochemical property analysis of designed MEVs.** ProtParam analysis of MEVs reveals that the CTL MEV is composed of 608 amino acids, has 62.45 kD molecular weight and poses 8.85 theoretical pI. The half-life of CTL MEV in mammalian reticulocytes, yeast and *Escherichia coli* was predicted to be 20 h, 30 min and 10 h, respectively; the aliphatic index was found to be 58.91, and the grand average of hydropathicity (GRAVY) was found to be 0.015, indicating the CTL MEV to have globular and hydrophilic nature. The instability index of CTL MEV was 44.58, indicating theoretically close to stable in nature.

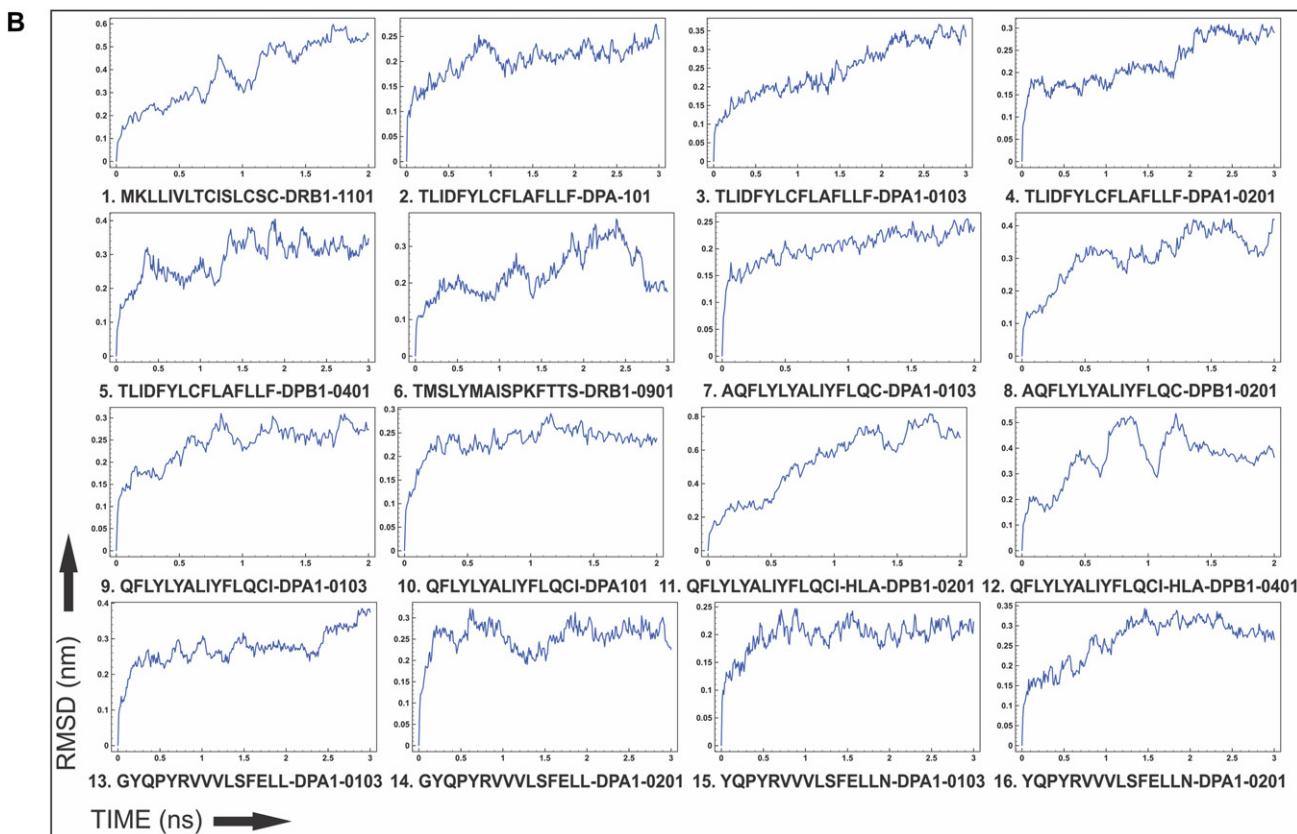
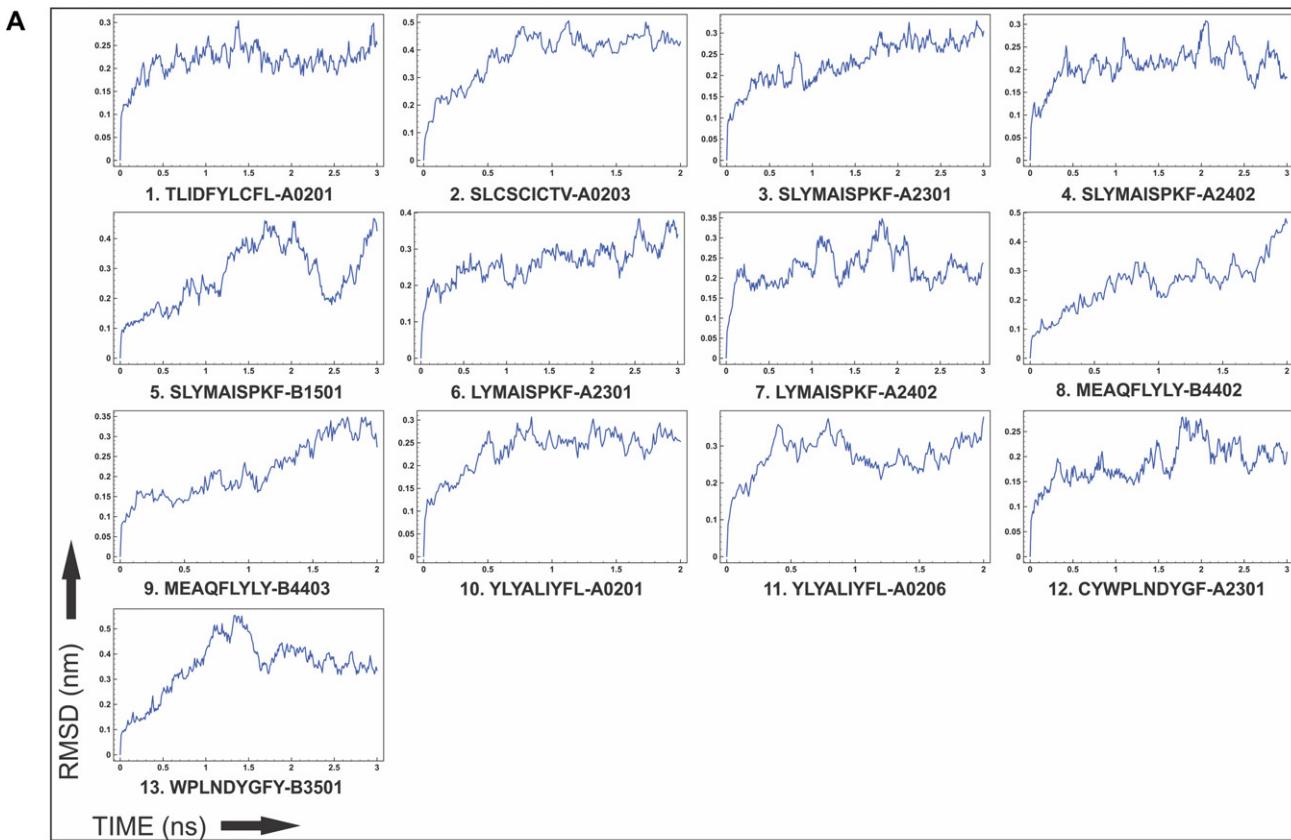
Likewise, the ProtParam study of designed HTL MEV has revealed it to be composed of 864 amino acids, has 89.62 kDa molecular weight and 8.72 theoretical pI. The half-life of HTL MEV in mammalian reticulocytes, yeast and *E. coli* was predicted to be 20 h, 30 min and 10 h, respectively; the aliphatic index was found to be 127.69, and the grand average of hydropathicity (GRAVY) was 1.145, both indicating HTL MEV to have globular and hydrophilic nature. The instability index of HTL MEV was 35.62, indicating it to be theoretically stable in nature.

#### Tertiary structure modelling and refinement of MEVs

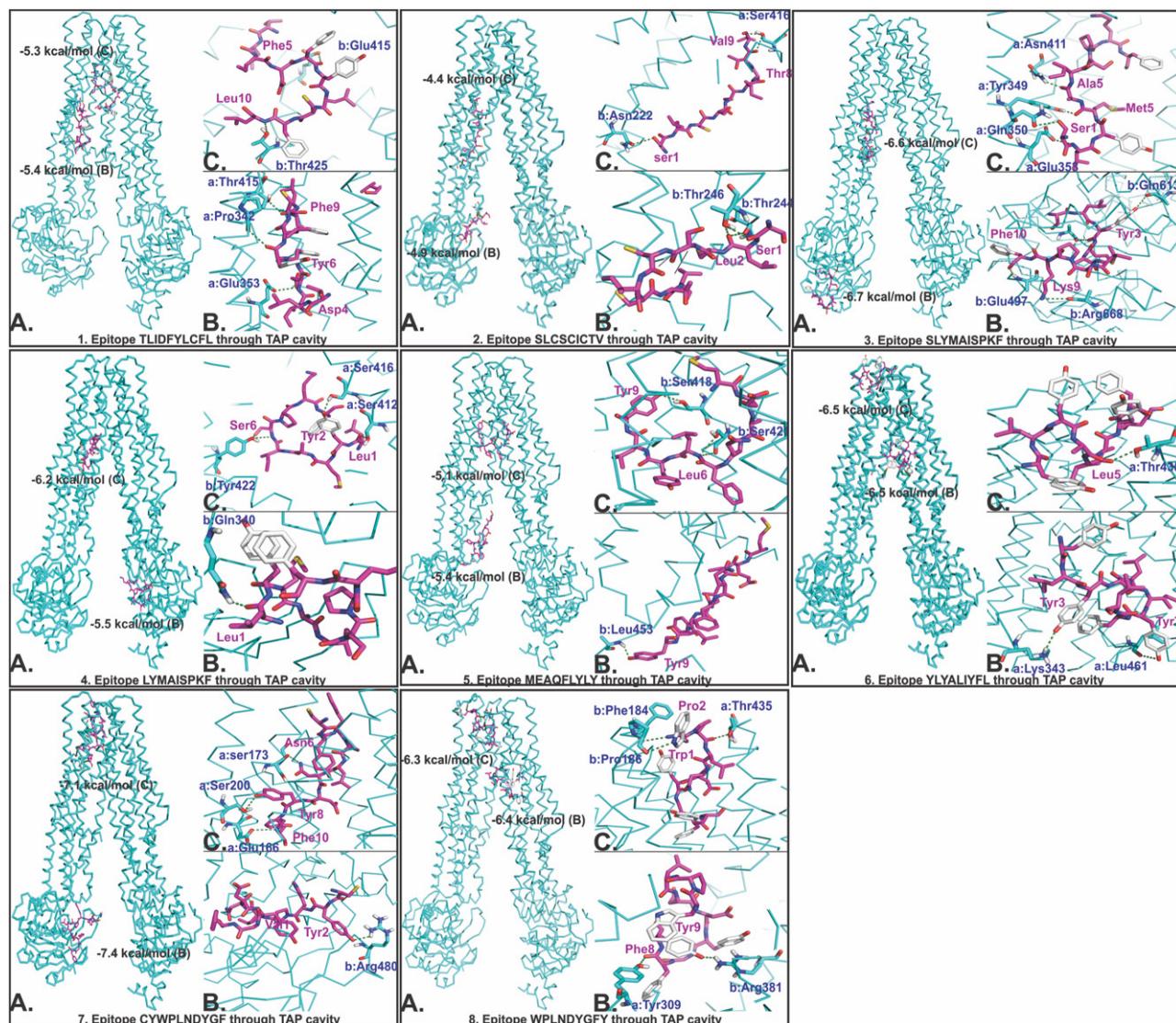
Amino acid sequences of both the CTL and HTL MEVs were subjected to tertiary structural homology modelling by RaptorX. The generated CTL MEV model was found to have 11% helix, 21%  $\beta$ -sheet and 67% coil; the model was 28% exposed, 33% medium and 37% buried in its folded conformation. The CTL MEV model was found to have three domains: first domain (1–199 aa, templates – 1u53A, 3nt8A, 2vznA, 1qnxA); second domain (200–438 aa, templates – 1vh4A, 4dn7A); and third domain (439–608 aa, template – 4lcl:A) (Figure 7(A)). Likewise, the HTL MEV model generated by RaptorX consists of 52% helix, 10%  $\beta$ -sheet and 36% coil; the generated HTL MEV model was 32% exposed, 26% medium and 40% buried in its folded conformation. The HTL MEV model has three domains: first domain (1–205 aa, template – 2yevA, 3eh3A, 3ayfA, 1m56A, 4xydA); second domain

**A****B**

**Figure 4.** (A). Molecular docking analysis of screened CTL epitopes with their respective HLA allele binders. Molecular docking of CTL epitopes (magenta sticks) with their respective HLA allele binders (cyan cartoon). Molecular docking study shows the docked complexes of CTL epitopes and HLA class I alleles to have significantly negative binding energy along with hydrogen bonds (green balls) formation in the complex interface. (B). Molecular docking analysis of screened HTL epitopes with their respective HLA allele binders. Molecular docking of CTL epitopes (magenta sticks) with their respective HLA allele binders (cyan cartoon). Molecular docking study shows the docked complexes of HTL epitopes and HLA class II alleles to have significantly negative binding energy along with hydrogen bonds (green balls) formation in the complex interface.



**Figure 5.** (A). Molecular dynamics simulation analysis of screened CTL epitopes and HLA allele complexes. Molecular dynamics simulation analysis of the CTL epitopes and their respective HLA class I allele binder complexes. The study reveals a stable behaviour of the CTL-HLA allele complexes throughout a 3-ns time window with reasonably invariable RMSD. (B). Molecular dynamics simulation analysis of screened CTL epitopes and HLA allele complexes. Molecular dynamics simulation analysis of the HTL epitopes and their respective HLA class II allele binder complexes. The study reveals a stable behaviour of the HTL-HLA allele complexes throughout a 3-ns time window with reasonably invariable RMSD.



**Figure 6.** Molecular docking analysis of CTL epitopes within the TAP transporter cavity. Molecular interaction of CTL epitopes (magenta sticks) within the TAP cavity (cyan ribbon) is shown in detail. For every panel of epitope–TAP complex, (A) shows the binding of epitope at two different sites of TAP cavity, (B) and (C) show detailed molecular interaction between epitopes and TAP cavity residues at two respective sites of TAP cavity; (a, b) show chains A and B of TAP transporter. H-bonds are shown in green dots. Binding energy is shown in kcal/mol.

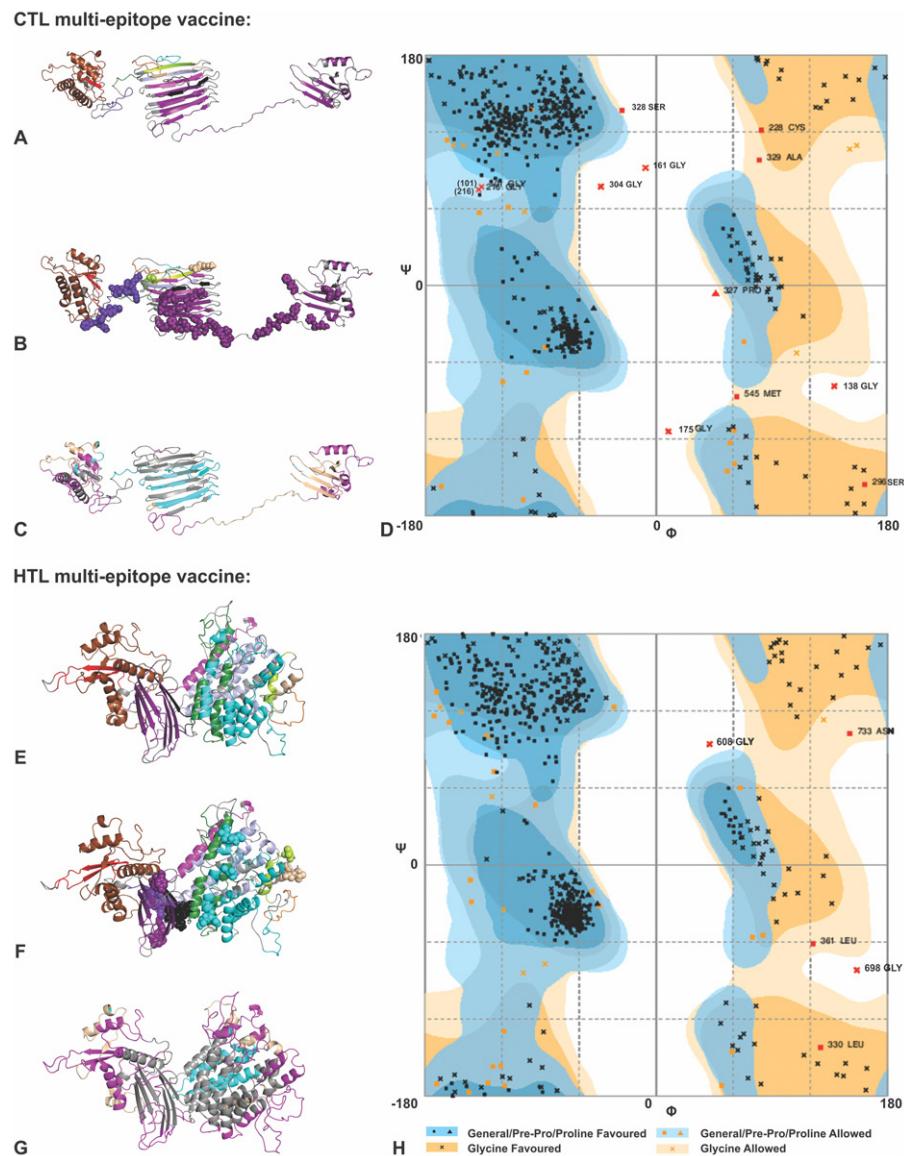
(206–731 aa, templates – 6any:A); and third domain (732–864 aa, template – 1vh4A, 4dn7A) (Figure 7(E)). The *P*-values for the best template used for homology modelling of CTL (best template: 1u53A) and HTL (best template: 6anyA) MEVs were 1.77e-08 and 1.97e-10, respectively. Good quality, mostly alpha proteins should have a *P*-value of less than 10<sup>-3</sup> and mostly beta proteins should have a *P*-value of less than 10<sup>-4</sup>. Hence, both the homology models of CTL and HTL MEVs are predicted to be of good quality. Since the constituting epitopes of both CTL and HTL MEVs show sequence overlap with several B cell linear epitopes (Figure 2), both the generated MEVs tertiary models also carry overlapping B cell linear epitopic regions (Figure 7(B,F)).

The structural accuracy of initial CTL and HTL MEV tertiary models including joining of the gaps was achieved by ModRefiner refinement. GalaxyRefine was used further for overall refinement of both the MEV models. Refined model 1 was chosen for both the MEVs, on the basis that they scored best. The refined model of CTL MEV had Rama favoured,

93.7%; GDT-HA, 0.9457; RMSD, 0.425; MolProbity, 2.326; Clash score, 26.5; and Poor rotamers, 0.9. Likewise, the refined model 1 of HTL MEV had Rama favoured, 95%; GDT-HA, 0.9447; RMSD, 0.430; MolProbity, 2.377; Clash score, 27.3; and Poor rotamers, 1.4. These values are the significantly improved values of different parameters after refinement of the tertiary models (Supplementary Table S10).

*Validation of CTL and HTL MEVs refined models.* Analysis of energetically allowed regions by RAMPAGE revealed that the refined CTL MEV model has 94.6% (573/608) residues in favoured region, 3.5% (21/608) residues in allowed region and only 2.0% (12/608) residues in outlier region, while the refined HTL MEV model has 95.8% (826/864) of residues in favoured region, 3.6% (31/864) residues in allowed region and only 0.6% (5/864) residues in outlier region (Figure 7(D,H)).

*Discontinuous B-cell epitope prediction from MEVs.* The structure-based B-cell discontinuous epitopes screening from both designed CTL and HTL MEVs 3D models was performed by ElliPro tool. The CTL MEV model was found to have eight



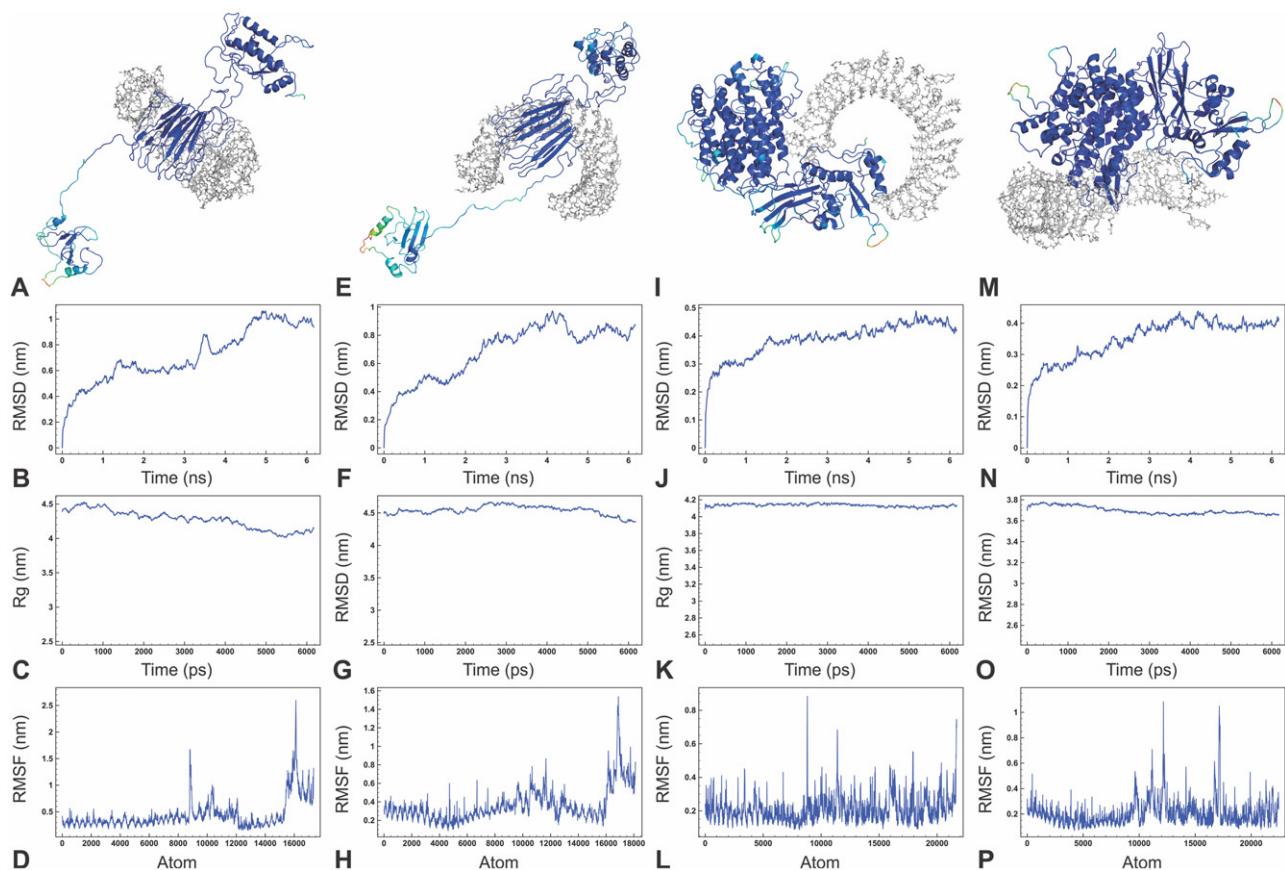
**Figure 7.** Tertiary structure modelling of CTL and HTL multi-epitope vaccines. (A) & (E) Tertiary structural models of CTL and HTL MEVs, respectively, showing epitopes from different proteins in different colours (epitope colours correspond to, as in Figure 6). (B) & (F) Overlapping B cell linear epitope regions present in CTL and HTL MEVs are shown in spheres. From (C) CTL and (G) HTL MEVs, INF- $\gamma$  inducing epitopes are shown in cyan, discontinuous B cell epitopes are shown in magenta and the regions common amongst INF- $\gamma$  and discontinuous B cell epitopes are shown in wheat colour. (D) & (H) RAMPAGE analysis of the refined CTL and HTL MEV models, respectively.

discontinuous epitopes and HTL MEV was found to have three discontinuous epitopes. The Protrusion Index (PI) score for CTL MEV epitopes ranges from 0.534 to 0.844 and that of HTL MEV ranges from 0.69 to 0.817 (Supplementary Table S11) (Figure 7(C,G)). The higher PI score indicates the greater potential of the B cell discontinuous epitope.

#### Molecular docking and dynamics simulation analysis of MEVs with TLR2 and TLR4 immunoreceptors

Both the CTL and HTL MEV refined models were subjected for molecular docking study with the ectodomain of human immunological receptors, TLR2 and TLR4, by using PatchDock tool. The highest scoring docking complex conformations were chosen for further study. The higher scoring indicates higher geometric shape complementarity fitting of

ligand (MEVs) and immunoreceptors (TLRs). The docking complexes show a fitting conformation of both the MEVs within ectodomains of the TLRs (Figure 8(A,E,I,M)). Further, the MD simulation analysis of the docked MEVs-TLRs complexes has shown a very convincing and reasonably stable RMSD value between  $\sim$ 0.4 and 1.0 nm for a given time window of 6 ns at a reasonably invariable temperature ( $\sim$ 300 K) and pressure ( $\sim$  1 bar) (Figure 8(B,F,J,N)). Molecular docking and MD simulation study of all the four MEVs-TLRs complexes indicate a stable complex formation. All the complexes have also shown to have a reasonably invariant  $R_g$  (radius of gyration) (Figure 8(C,G,K,O)), as well as RMSF for all the atoms in all four complexes (Figure 8(D,H,L,P)). These results indicate all the complexes to have a stable nature with proper molecular folding. The B-factor of all four MEVs-TLRs complexes also indicates the complexes to be



**Figure 8.** Molecular docking and dynamics simulation analysis of CTL and HTL complexes with TLR2 and TLR4. Complexes (A) CTL MEV-TLR2 (E), CTL MEV-TLR4, (I) HTL MEV-TLR2, and (M) HTL MEV-TLR4 are shown with CTL and HTL MEVs in cartoon (VIBGYOR) and TLR2 and TLR4 in line (grey). B-factor of CTL and HTL MEVs showing that the stable region is in blue and the region in red being unstable (VIBGYOR presentation). In all the four complexes of designed MEVs with TLRs, most of the regions are in blue and very small region is in yellow and orange, indicating MEVs-TLRs complexes to be very stable in nature. Panels B, F, J and N show acceptably invariable RMSD of MD simulation for respective MEV-TLR complexes. Panels C, G, K and O show reasonably invariable  $R_g$  (radius of gyration) of respective MEV-TLR complexes. Panels D, H, L and P show reasonably invariable RMS fluctuation for all the atoms of respective MEV-TLR complexes.

stable (blue) with a very small region being acceptably fluctuating (yellow and orange) as shown by VIBGYOR presentation in Figure 8(A,E,I,M).

#### cDNA analysis of CTL and HTL MEVs

Codon-optimized cDNA of both CTL and HTL MEVs were generated by Codon Usage Wrangler tool with mammalian host cell line (Human) being the choice of expression system. Further, the GenScript Rare Codon Analysis has revealed that the GC content of optimized CTL MEV cDNA was 68.32% and the CAI score was 1.00 with 0% tandem rare codons. Likewise, the GC content of optimized HTL MEV cDNA was 67.16% and the CAI score was 1.00 with 0% tandem rare codons. Ideally for high expression, GC content should be 30–70%; the CAI score indicating possibility of cDNA expression in the chosen expression system should be between 0.8 and 1.0, and the tandem rare codon frequency indicating the percentage of low-frequency codons present in cDNA for the chosen expression system should be <30%. The tandem rare codons may hamper the expression of cDNA or even interrupt the translational machinery. Hence as per above analysis of different parameters, the optimized cDNA of both MEVs is predicted to be highly expressing in mammalian host cell line (Human).

#### Conclusions

SARS MEVs design proposed in the present study is composed of CTL, HTL and potential B cell epitopes overlaps with potential to elicit humoral as well as cellular responses. Truncated Ov-ASP-1 (residues 10–153) was chosen as the molecular adjuvant for both MEVs. The CTL and HTL epitopes chosen for MEV design were *in silico* validated for their stable molecular interaction with their respective HLA allele binders; moreover, for CTL epitopes, their molecular interaction within the TAP transporter cavity was also studied. Average world population coverage by the combined use of all the chosen CTL and HTL epitopes was analysed to be as high as 85.21%. Along with CTL and HTL epitopes, both the designed and modelled MEVs have also shown to be the carrier of potential IFN- $\gamma$  epitopes as well as B cell discontinuous epitopes. Furthermore, both the CTL and HTL MEVs tertiary models show to have a significant and stable molecular interaction with immunoreceptors TLR2 and TLR4. Since both the MEVs contain CTL, HTL as well as B cell epitopes, administration of both the MEVs is predicted to cause humoral as well as cell-mediated immune response. The codon-biased cDNA of both the CTL and HTL MEVs for expression in the mammalian host cell line (Human) was analysed to be favourable for high-level expression. Hence, both

the MEVs could be cloned, expressed and tried for *in vivo* validations and animal trials at the laboratory level.

## Acknowledgements

We acknowledge the Advance Instrumentation Research Facility (AIRF) of the Jawaharlal Nehru University, New Delhi, for extending their advanced computational facility to conduct experiments.

## Disclosure statement

The authors declare to have no competing interests.

## Author contributions

Protocol design: S.S., Methodology performed by: S.S., M.K. Result interpretation and scientific writing: S.S., M.K., R.K.P., A.K.S., V.S., S.K.S., R.K.S. and N.S.

## ORCID

Sukrit Srivastava  <http://orcid.org/0000-0003-0295-0272>

## References

- Abraham, M. J., Murtola, T., Schulz, R., Pa Li, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from lap-tops to supercomputers. *SoftwareX*, 1–2, 19–25. doi:[10.1016/j.softx.2015.06.001](https://doi.org/10.1016/j.softx.2015.06.001)
- Åkerström, S., Mirazimi, A., & Tan, Y. J. (2007). Inhibition of SARS-CoV replication cycle by small interference RNAs silencing specific SARS proteins, 7a/7b, 3a/3b and S. *Antiviral Research*, 73(3), 219–227.
- 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.
- Arai, R., Ueda, H., Kitayama, A., Kamiya, N., & Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering*, 14(8), 529–532. doi:[10.1093/protein/14.8.529](https://doi.org/10.1093/protein/14.8.529)
- Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)*, 22(2), 195–201.
- Bai, Y., & Shen, W. C. (2006). Improving the oral efficacy of recombinant granulocyte colony-stimulating factor and transferrin fusion protein by spacer optimization. *Pharmaceutical Research*, 23(9), 2116–2121. doi:[10.1007/s11095-006-9059-5](https://doi.org/10.1007/s11095-006-9059-5)
- Benkert, P., Tosatto, S. C., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, 71(1), 261–277. doi:[10.1002/prot.21715](https://doi.org/10.1002/prot.21715)
- Booth, C. M., Matukas, L. M., Tomlinson, G. A., Rachlis, A. R., Rose, D. B., Dwosh, H. A., ... Detsky, A. S. (2003). "Clinical features and short-term outcomes of 144 patients with SARS in the greater Toronto area". *JAMA*, 289(21), 2801–2809.
- 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.
- 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
- Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., ... Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. *PloS Computational Biology*, 8(1), 361.
- CDC Report; Severe Acute Respiratory Syndrome (SARS) (2018). SARS Basics Fact Sheet. Atlanta, GA: Centers for Disease Control and Prevention; as cited on 28 July. Retrieved from <https://www.cdc.gov/sars/about/fs-sars.html>
- Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., & Huang, T. H. (2014). The SARS coronavirus nucleocapsid protein – forms and functions. *Antiviral Research*, 103, 39–50. doi:[10.1016/j.antiviral.2013.12.009](https://doi.org/10.1016/j.antiviral.2013.12.009)
- Chen, C. Y., Ping, Y. H., Lee, H. C., Chen, K. H., Lee, Y. M., Chan, Y. J., ... Chen, Y. M. A. (2007). Open reading frame 8a of the human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induces apoptosis. *The Journal of Infectious Diseases*, 196(3), 405–415. doi:[10.1086/519166](https://doi.org/10.1086/519166)
- Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews*, 65(10), 1357–1369. doi:[10.1016/j.addr.2012.09.039](https://doi.org/10.1016/j.addr.2012.09.039)
- Chou, P. Y., & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology and Related Areas of Molecular Biology*, 47, 45–148.
- DeDiego, M. L., Pewe, L., Alvarez, E., Rejas, M. T., Perlman, S., & Enjuanes, L. (2008). Pathogenicity of severe acute respiratory coronavirus deletion mutants in hACE-2 transgenic mice. *Virology*, 376(2), 379–389. doi:[10.1016/j.virol.2008.03.005](https://doi.org/10.1016/j.virol.2008.03.005)
- Delneste, Y., Beauvillain, C., & Jeannin, P. (2007). Innate immunity: Structure and function of TLRs. *Médecine/Sciences*, 23(1), 67–73. [InsertedFromOnline]
- Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, 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.
- Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B. J., & Jiang, S. (2009). The spike protein of SARS-CoV – a target for vaccine and therapeutic development. *Nature Reviews Microbiology*, 7(3), 226. doi:[10.1038/nrmicro2090](https://doi.org/10.1038/nrmicro2090)
- Duhovny, D., Nussinov, R., & Wolfson, H. J. (2002). Efficient unbound docking of rigid molecules. In Gusfield (Ed.), *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics (WABI) Rome, Italy, Lecture Notes in Computer Science* (Vol. 2452, pp. 185–200). Berlin, Germany: Springer Verlag
- Emini, E. A., Hughes, J. V., Perlow, D. S., & Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology*, 55(3), 836–839.
- 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). New York, NY: Humana Press.
- Guo, J., Yang, Y., Xiao, W., Sun, W., Yu, H., Du, L., ... Zhou, Y. (2015). A truncated fragment of Ov-ASP-1 consisting of the core pathogenesis-related-1 (PR-1) domain maintains adjuvanticity as the full-length protein. *Vaccine*, 33(16), 1974–1980. doi:[10.1016/j.vaccine.2015.02.053](https://doi.org/10.1016/j.vaccine.2015.02.053)
- Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., & Raghava, G. P. and Open Source Drug Discovery Consortium, (2013). *In silico* approach for predicting toxicity of peptides and proteins. *PLoS One*, 8(9), e73957.
- 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. doi:[10.1016/j.meegid.2016.12.010](https://doi.org/10.1016/j.meegid.2016.12.010)
- He, Y., Barker, S. J., MacDonald, A. J., Yu, Y., Cao, L., Li, J., ... Jiang, S. (2009). Recombinant Ov-ASP-1, a Th1-biased protein adjuvant derived from the helminth *Onchocerca volvulus*, can directly bind and activate antigen-presenting cells. *The Journal of Immunology*, 182(7), 4005–4016. doi:[10.4049/jimmunol.0800531](https://doi.org/10.4049/jimmunol.0800531)
- Hoof, I., Peters, B., Sidney, J., Pedersen, L. E., Sette, A., Lund, O., ... Nielsen, M. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*, 61(1), 1. doi:[10.1007/s00251-008-0341-z](https://doi.org/10.1007/s00251-008-0341-z)
- Hu, W., Li, F., Yang, X., Li, Z., Xia, H., Li, G., ... 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.

- Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M. S., Novotný, J., Margolies, M. N., ... Crea, R. (1988). Protein engineering of antibody binding sites: Recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 85(16), 5879–5883. doi:10.1073/pnas.85.16.5879
- Jimenez-Guardeño, J. M., Nieto-Torres, J. L., DeDiego, M. L., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodríguez, C., & Enjuanes, L. (2014). The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS Pathogens*, 10(8), e1004320. doi:10.1371/journal.ppat.1004320
- Jin, M. S., Kim, S. E., Heo, J. Y., Lee, M. E., Kim, H. M., Paik, S. G., ... Lee, J. O. (2007). Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell*, 130(6), 1071–1082.
- Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45), 11225–11236. doi:10.1021/ja9621760
- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, 7(8), 1511. doi:10.1038/nprot.2012.085
- Karplus, P. A., & Schulz, G. E. (1985). Prediction of chain flexibility in proteins. *Naturwissenschaften*, 72(4), 212–213. doi:10.1007/BF01195768
- Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Research*, 40(W1), W294–W297. doi:10.1093/nar/gks493
- Kolaskar, A. S., & Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters*, 276(1-2), 172–174.
- 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. doi:10.1371/journal.pcbi.1002829
- Larsen, J. E., Lund, O., & Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immune Research*, 2, 2.
- Law, P. Y. P., Liu, Y. M., Geng, H., Kwan, K. H., Waye, M. M. Y., & Ho, Y. Y. (2006). Expression and functional characterization of the putative protein 8b of the severe acute respiratory syndrome-associated coronavirus. *FEBS Letters*, 580(15), 3643–3648. doi:10.1016/j.febslet.2006.05.051
- Leung, T. F., Wong, G. W. K., Hon, K. L. E., & Fok, T. F. (2003). Severe Acute respiratory syndrome (SARS) in children: epidemiology, presentation and management. *Paediatric Respiratory Reviews*, 4(4), 334–339. doi:10.1016/S1526-0542(03)00088-5
- Liao, Y., Zhang, S. M., Neo, T. L., & Tam, J. P. (2015). Tryptophan-dependent membrane interaction and heteromerization with the internal fusion peptide by the membrane proximal external region of SARS-CoV spike protein. *Biochemistry*, 54(9), 1819–1830. doi:10.1021/bi501352u
- Lovell, S. C., Davis, I. W., Arendall, W. B., III, De Bakker, P. I., Word, J. M., Prisant, M. G., ... Richardson, D. C. (2003). Structure validation by  $\text{C}\alpha$  geometry:  $\phi$ ,  $\psi$  and  $\text{C}\beta$  deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3), 437–450. doi:10.1002/prot.10286
- Lu, B., Tao, L., Wang, T., Zheng, Z., Li, B., Chen, Z., ... Wang, H. (2009). Humoral and cellular immune responses induced by 3a DNA vaccines against severe acute respiratory syndrome (SARS) or SARS-like coronavirus in mice. *Clinical and Vaccine Immunology*, 16(1), 73–77. doi:10.1128/CVI.00261-08
- MacDonald, A. J., Cao, L., He, Y., Zhao, Q., Jiang, S., & Lustigman, S. (2005). rOv-ASP-1, a recombinant secreted protein of the helminth *Onchocerca volvulus*, is a potent adjuvant for inducing antibodies to ovalbumin, HIV-1 polypeptide and SARS-CoV peptide antigens. *Vaccine*, 23(26), 3446–3452. doi:10.1016/j.vaccine.2005.01.098
- Morla, S., Makhija, A., & Kumar, S. (2016). Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene*, 584(1), 1–6.
- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14), 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791. doi:10.1002/jcc.21256
- Nezafat, N., Eslami, M., Negahdaripour, M., Rahbar, M. R., & Ghasemi, Y. (2017). Designing an ancient multi-epitope oral vaccine against *Helicobacter pylori* using immunoinformatics and structural vaccinology approaches. *Molecular Biosystems*, 13(4), 699–713. doi:10.1039/c6mb00772d. doi:10.1039/C6MB00772D
- Nielsen, M., Lundsgaard, 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. doi:10.1186/1471-2105-8-238
- Ohto, U., Yamakawa, N., Akashi-Takamura, S., Miyake, K., & Shimizu, T. (2012). Structural analyses of human Toll-like receptor 4 polymorphisms D299G and T399I. *Journal of Biological Chemistry*, 287(48), 40611–40617. doi:10.1074/jbc.M112.404608
- Oldham, M. L., Grigorieff, N., & Chen, J. (2016). Structure of the transporter associated with antigen processing trapped by herpes simplex virus. *eLife*, 5, e21829.
- Parker, J. M., Guo, D., & Hodges, R. S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, 25(19), 5425–5432.
- Peters, B., Bulik, S., Tampe, R., Van Endert, P. M., & Holzhtutter, H. G. (2003). Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *Journal of Immunology (Baltimore, MD: 1950)*, 171(4), 1741–1749.
- Pfefferle, S., Krähling, V., Ditt, V., Grywna, K., Mühlberger, E., & Drosten, C. (2009). Reverse genetic characterization of the natural genomic deletion in SARS-CoV strain Frankfurt-1 open reading frame 7b reveals an attenuating function of the 7b protein *in-vitro* and *in-vivo*. *Virology Journal*, 6(1), 131. doi:10.1186/1743-422X-6-131
- 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.
- Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flück, P., Parham, P., & Marsh, S. G. (2014). The IPD and IMGT/HLA database: Allele variant databases. *Nucleic Acids Research*, 43(Database issue), D423–D431.
- Saha, S., & Raghava, G. (2006). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, 34(Web Server issue), W202–W209.
- 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.
- Shi, C. S., Qi, H. Y., Boulanian, C., Huang, N. N., Abu-Asab, M., Shelhamer, J. H., & Kehrl, J. H. (2014). SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. *The Journal of Immunology*, 193(6), 3080–3089. doi:10.4049/jimmunol.1303196
- 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. doi:10.1186/1745-7580-4-2
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539. doi:10.1038/msb.2011.75
- Singh, S., Singh, H., Tuknait, A., Chaudhary, K., Singh, B., Kumaran, S., & Raghava, G. P. S. (2015). PEPrStrMOD: Structure prediction of peptides containing natural, non-natural and modified residues. *Biology Direct*, 10, 73.
- Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U., ... 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. doi:10.1038/9858
- Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., ... Holzhtutter, 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: Cmls*, 62(9), 1025–1037.
- Totura, A. L., Whitmore, A., Agnihotram, 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.
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2): 455–461. doi:[10.1002/jcc.21334](https://doi.org/10.1002/jcc.21334)
- Varshney, B., Agnihotram, S., Tan, Y. J., Baric, R., & Lal, S. K. (2012). SARS coronavirus 3b accessory protein modulates transcriptional activity of RUNX1b. *PloS One*, 7(1), e29542. doi:[10.1371/journal.pone.0029542](https://doi.org/10.1371/journal.pone.0029542)
- Vasilenko, N., Moshynskyy, I., & Zakhartchouk, A. (2010). SARS coronavirus protein 7a interacts with human Ap4A-hydrolase. *Virology Journal*, 7(1), 31
- Voß, D., Pfefferle, S., Drosten, C., Stevermann, L., Traggiai, E., Lanzavecchia, A., & Becker, S. (2009). Studies on membrane topology, N-glycosylation and functionality of SARS-CoV membrane protein. *Virology Journal*, 6(1), 79. doi:[10.1186/1743-422X-6-79](https://doi.org/10.1186/1743-422X-6-79)
- 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
- Wang, Y., & Liu, L. (2016). The membrane protein of severe acute respiratory syndrome coronavirus functions as a novel cytosolic pathogen-associated molecular pattern to promote beta interferon induction via a Toll-like-receptor-related TRAF3-independent mechanism. *MBio*, 7(1), e01872–e01815.
- Wei, W. Y., Li, H. C., Chen, C. Y., Yang, C. H., Lee, S. K., Wang, C. W., ... Lo, S. Y. (2012). SARS-CoV nucleocapsid protein interacts with cellular pyruvate kinase protein and inhibits its activity. *Archives of Virology*, 157(4), 635–645. doi:[10.1007/s00705-011-1221-7](https://doi.org/10.1007/s00705-011-1221-7)
- Wu, X., Wu, S., Li, D., Zhang, J., Hou, L., Ma, J., ... He, F. (2010). Computational identification of rare codons of *Escherichia coli* based on codon pairs preference. *BMC Bioinformatics*, 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. doi:[10.1016/j.bpj.2011.10.024](https://doi.org/10.1016/j.bpj.2011.10.024)
- Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, L. L. M., Xie, Z. H., Chan, K. H., ... Guan, Y. (2003). Epidemiological and aetiological studies of patients with severe acute respiratory syndrome (SARS) from Guangdong in February 2003. *Lancet*, 362(9393), 1353–1358. doi:[10.1016/S0140-6736\(03\)14630-2](https://doi.org/10.1016/S0140-6736(03)14630-2)
- Zhong, X., Guo, Z., Yang, H., Peng, L., Xie, Y., Wong, T. Y., ... Guo, Z. (2006). Amino terminus of the SARS coronavirus protein 3a elicits strong, potentially protective humoral responses in infected patients. *Journal of General Virology*, 87(2), 369–373. doi:[10.1099/vir.0.81078-0](https://doi.org/10.1099/vir.0.81078-0)