In Silico design and characterization of multiepitopes vaccine for SARS-CoV2 from its Spike proteins

Gunderao H Kathwate*

Department of Biotechnology, Savitribai Phule Pune University, Pune

*Corresponding author:

Gunderao H Kathwate

Department of Biotechnology, Savitribai Phule Pune University, Pune

Email: santoshkathwate@gmail.com

Orchid id: 0000-0002-3649-8192

Abstract

COVID 19 is disease caused by novel corona virus, SARS-CoV2 originated in China most probably of Bat origin. Till date, no specific vaccine or drug has been discovered to tackle the infections caused by SARS-CoV2. In response to this pandemic, we utilized bioinformatics knowledge to develop efficient vaccine candidate against SARS-CoV2. Designed vaccine is rich in effective BCR and TCR epitopes screened from the sequence of S-protein of SARS-CoV2. Bioinformatics tools were utilized for prediction, refinement and validation of tertiary structure of designed vaccine. Protein-Protein interaction prediction of TLR2/4 and designed vaccine indicate effective binding. Designed multiepitopes vaccine has induce cell mediated and humoral immunity along with increased interferon gamma response. Macrophages and dendritic cells were also found increased over the vaccine exposure. *In silico* codon optimization and cloning in expression vector indicate that vaccine can be efficiently expressed in E. coli. In conclusion, predicted vaccine is a good antigen, probable no allergen and has potential to induce cellular and humoral immunity.

Key words: COVID19, SARS-CoV-2, multi-epitope vaccine, Immune simulation, *in silico* cloning.

Introduction

In December 2019 group of patients from Wuhan city of China was found to have pneumonia like symptoms and diagnosed with the infection of beta coronavirus (Guo, 2020). Later it spread across the china and now spread all over the globe. These infections were named as COVID 19 disease and the virus as SARS-CoV2 by WHO (WHO, 2020). On 30 January 2020 due to its spread more than 120 countries declared COVID 19 pandemic as a public health emergency of international concern. Novelty of SARS-CoV2 is its rapid spread may be due asymptomatic patients (Gandhi, 2020; Bai, 2020) and highly sophisticated, time consuming diagnostic methods (Chan, 2020; Yelin, 2020; Li, 2020). As of 13 May 2020, 4.28 million confirmed cases with more than 0.2 million deaths worldwide (WHO, 2020). In India due to lockdown positive cases are in control and spread is also marginal, but still after approximately four months case reports are increases. As of 13 May 2020, a total of 71865 confirmed cases with 2415 deaths are reported by ministry of health and family welfare, Govt. of India. Till date, no antiviral drugs available to combat the SARS-CoV2 infections. There are few drug candidates in clinical trials but the process is time consuming (Mitjà 2020). In South Korea, Lopinavir/Ritonavir combination was found significantly effective in lowering of viral load to no detectible or little SAR-CoV2 titer (Kim, 2020). But in another group study same combination was totally ineffective beyond standard care (Cao, 2020). Another drug pair drug, hydroxychloroquine, an antimalarial drug and azithromycin reported to found effectively associated with reduction of viral load (Gautret 2020, Gautret, 2020). But QT interval prolongation that may cause life threating arrhythmia (Chorin, 2020) and in large scale study both drugs and drug alone compared with neither drugs was not associated with mortality (Rosenberg 2020, Molina, 2020). In a drug repurposing study, remdesivir, lopinavir, emetine, and homoharringtonine have significantly inhibited replication of SAR-CoV (Choy, 2020). But this was *in vitro* study, clinical trial report are underway. Similar determined attempts are going through the globe. To break the chain of infection prevention is much better option than treatments. Several evidences suggest and support the importance of vaccine to eliminate COVID 19 epidemic (Chen 2020; Bi 2020; Prem 2020). Various epidemiological surveys provide the evidences that naturally acquired immunity can eliminate the SARS-CoV2 titer (Prompetchara 2020; Shi 2020). More than 80% patients

develop mild and 14% develop severe symptoms of COVID19 with zero case fatality rate (Spychalski 2020).

Presently there is no vaccine available against COVID 19 disease. Efforts are being taken for the development of effective vaccine all over the globe (Ahmed, 2020; Kim, 2020, Tai, 2020; Wang, 2020). All these vaccines are under development (WHO, 2020) and may require at least 1to 1.5 year to hit the market (Spinney, 2020). Bioinformatics approach can accelerate the process by predicting potential candidate peptides. Vaccines against MERS, Ebola, and human papilloma virus were designed and successfully developed by using bioinformatics approaches ().

SARS-CoV belong to beta coronoviridae family which includes four endemic viz HCoV-HKU1, HCoV-OC43, HCoV-229E, HCoV-NL63, two epidemic viruses like Middle East Respiratory Syndrome virus (MERS), and SARS (Osman 2020). This is a non-segmented positive strand RNA virus with envelop. Proteins are categorized into structural, non-structural and accessory proteins. Structural proteins are involved in protection and bind to the host. Several proteins of SAR-CoV2 are important for virulence and pathogenesis. For example Nucleocapsid (N) protein N, is essential for RNA binding and its replication and transcription (Snijder 2016). Envelope (E) and membrane (M) proteins and instrumental in virus assembly and virulence promotion (Ruch and Machamer 2012; Alsaadi and Jones 2019). E and M proteins are effective in induction of immune response (Ahmed 2020). Spike (S), a structural protein is responsible for binding to receptor on host cell, Angiotensin converting enzyme 2 (Hoffmann, 2020). Thereby proteolytic cleavage by TMPRSS2 allows subsequent entry into cell through endocytosis (Glowacka 2011). S protein also found to be involved in activation of T cell response (). Entire S protein expressed in chimpanzee adeno 38 (ChAd)-vector showed protection against SARS-CoV2 in mice and *rhesus* macaques (Doremalen 2020). Single dose of ChAdOx1 nCoV-19 vaccine is sufficient to elicit humoral and cell mediated immune response in both animals. Viral load was found reduced compared to control animals and symptoms of pneumonia were absent. Here we designed a multi-epitopic vaccine from S Protein. This vaccine has all the ideal properties like stability at room temperature, immunogenic, antigenic and non-allergen. All the epitopes are good in stimulation of humeral as well and cell mediated immunity.

Methods

Selection of protein for epitope prediction

Complete Spike protein sequence (P0DTC2) of SARS-CoV2 was downloaded in fasta format from UniProt protein database. This sequence was used for further analysis and obtaining potential epitopes for B and T cell receptors.

T cell Epitope prediction

There are various tools used to predict the epitopes that could be presented to T cell receptors. Potential epitopes of various bacteria and viruses are predicted by using such online epitopes predicting tools. Following tools were used for prediction of TCR (T cell Receptor) epitopes IEDB MHC-I processing predictions, MHC-NP, netCTLpan1.1, RANKPEP, and netMHCpan4.0. All these tools were used to screen the potential peptides as epitopes for TCR. These tools have various methods to predict the epitopes. IEDB (Instructor/Evaluator Database) is a web based epitope analysis resource includes tools for T cell epitope prediction, B cell epitope prediction and other analysis tools like epitope conservancy etc. This resource use methods based on artificial neural network (ANN), stabilized matrix method (SMM), and Combinatorial Peptide Libraries(CombLib), predicts the peptides way that processed naturally and presented by MHC I (Calis et al., 2013).

Analysis of immunological properties

Immunogenicity, Toxicity, Allergen, Conservancy, and Antigencity were analyzed for predicted TCR epitopes. IEDB MHC class I immunogenicity server and conservancy tool were used for determination of immunogenicity and conservancy. Immunogenicity of a peptide MHC complex (http://tools.immuneepitope.org/immunogenicity/) were assessed keeping all the parameters at default. Protein sequence variants (Calis et al., 2013) used setting sequence identity 100% and other parameter default. ToxiPred (http://www.imtech.res.in/raghava/toxinpred/index.html) online tool predicts the toxicity considering physicochemical properties of selected peptides (Gupta et al., 2013). Online server AllergenFP v.1.0 (http://ddgpharmfac.net/AllergenFP/) was used to predict peptides as allergens (Dimitrov et al., 2014). Antigenicity of epitopes was analyzed by online server VaxiJen v2.0 (http://www.ddg-harmfac.net/vaxijen/VaxiJen/VaxiJen.html). Threshold value was set to 0.5 (Doytchinova et al., 2007). This is alignment independent predictor based on auto-cross covariance (ACC) transformation epitopes sequences into uniform vectors of principle amino acid properties. Accuracy of this server varies in between 70 to 89% depending on targeted organism.

Linear B cell receptor epitope prediction

Six different method were used for the prediction of B cell receptor (BCR) epitopes. All these methods generate fragments the protein. For the prediction of B cell epitopes it is necessary to find linear sequence of B cell epitope in the protein sequence. BepiPred linear epitope predication server uses a hidden Markov model and propensity scale method. Similarly other properties are also be consider to predict good B cell epitopes. Those properties are calculated by different methods like Kolaskar–Tongaonkar antigenicity scale provide physiology of the aminoacid residues, Emini Surface accessible score for accessible surface of the epitope, Secondary structure of epitopes also has role in antigenicity. Karplus-Schulz flexibility score and Chou-Fasman β turns methods were used for flexibility and β turns prediction respectively. Parker hydrophilicity prediction method was used for determination of *in silico* hydrophilicity of peptides

Engineering of multi epitope vaccine sequence

High scored and common peptides predicted by various tools were selected for deriving sequence of potential vaccine candidate. Different epitopes for T cell and B cell receptors were linked together by GPGPG and AAY. To enhance the immunogenicity, OmpA (GenBank: AFS89615.1) protein was chosen as adjuvant and was linked through EAAAK at N terminal site of the vaccine.

Prediction of immunogenic properties of designed vaccine:

Antigenicity of chimeric vaccine was predicted by VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) and AntigenPro. VaxiJen 2.0 is alignment free antigenicity prediction tool utilizes auto cross covariance transformation of protein sequence into uniform vector vectors of principal amino acid properties (Doytchinova et al., 2007). ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) is also online tool utilizes protein microarray dataset for the prediction of antigenicity. Based on cross validation experiment accuracy of this server is estimated to be around 76 % using combined dataset. AllerTop v2.0 and AllergenFP were two online tools utilized for the prediction of allergenicity of chimeric protein. Amino acid E-descriptors, auto- and cross-covariance transformation, and the k nearest neighbours (kNN) machine learning methods are basis of AllerTop v2.0 (http://www.ddg-

pharmfac.net/AllerTOP) for allergenicity prediction (Dimitrov et al., 2014). AllergenFP, is descriptor based fingerprint, alignment free tool for allergenicity prediction. The tools uses four step algorithm. In first step, proteins are described on the basis of their properties like hydrophobicity, size, secondary structures formation and relative abundance. In subsequent step, generated strings are converted into vectors of equal length by ACC. Then vectors are converted into binary fingerprints and compared in terms of the Tanimoto coefficient. Applying this approach to known allergen and non-allergens can identify the 88 % of allergen/non-allergen with Mathew's Correlation coefficient of 0.759 (Dimitrov et al., 2014).

Prediction of solubility and Physiochemical properties:

For Solubility prediction of multi-epitopes chimeric vaccine, PROSO II server (http://mbiljj45.bio.med.uni-muenchen.de:8888/prosoII/prosoII.seam) was utilized (). PROSO II server works on an approach of classifier which utilizes difference between TargetDB and PDB and insoluble proteins of TargetDB. The accuracy of this server is 71% at default threshold 0.6. Protoparam, online web server was exploited for evaluation of physiochemical properties. The properties like amino acid composition, theoretical pI, instability index, in vitro and in vivo half-life, aliphatic index, molecular weight, and grand average of hydropathicity (GRAVY) were evaluated.

Secondary and tertiary structure prediction:

PSIPRED and RaptorX Property online servers were used to determine secondary structure of predicted vaccine. PSIPRED is publicly available webserver, includes two feed forward neural networks works on output obtained from PSI-BLAST. PSIPRED 3.2 attains average Q3 score of 81.6% obtained using very stringent cross approval strategies to assess its performance. RaptorX property is another web based server prediction secondary structure of protein without template (Wang et al., 2016). This server utilize DeepCNF (Deep Convolutional Neural Fields), a new machine learning model that predicts secondary structure and solvent accessibility and disorder regions simultaneously (Wang 2016). It accomplish Q3 score of approximately 84 for 3 state secondary structure and approx. 66% for 3 state solvent accessibility.

Tertiary structure of multi-epitopes chimeric vaccine candidate was built using I TASSER online server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The I-TASSER (Iterative Threading

ASSEmbly Refinement) web server utilize sequence-to-structure-to-function paradigm to build protein structure (Roy et al., 2010). It is top ranked 3D protein structure web server in community wide CASP experiments (Zheng et al., 2019).

Tertiary structure refinement and validation:

Two web bases servers were used to refine the 3D structure of multi-epitopes chimeric protein. Initially Modrefiner server (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) and finally GalaxyRefine server (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) was used. Modrefiner server is an algorithm for atomic level structure refinement utilizes c alpha trace, main chain or atomic model. Output structure is refined in term of accurate position of side chains hydrogen bon network and less atomic overlaps (Xu et al., 2011). On other hand Galaxy server rebuild the 3D structure, performs repacking, and uses molecular dynamic simulations to accomplish overall protein structure relaxation. Structure refined by Galaxy server of best quality in accordance with community-wide CASP10 experiments (Ko et al., 2012).

ProSA-web (https://prosa.services.came.sbg.ac.at/prosa.php), The ERRAT server (http://services.mbi.ucla.edu/ERRAT/) and RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) web servers were utilized for 3D structure validation obtained after galaxy server refinement. ProSA web server calculate the quality score as Z score that should fall in a characteristic range (Wiederstein et al., 2007). Z score obtained for a specific input protein is in context with protein structure available in public domain. ERRATA server analyses non bonded atom-atom interaction the refined 3D structure of protein compared to high resolved crystallographic protein structures (Colovos et al., 1993). Ramachandran plot displays energetically allowed and disallowed dihydral psi and phi angles of amino acids. The plot is calculated based on van der Waal radius of the protein side chain. RAMPAGE server determines Ramachandran plot for a protein that include percent residues in allowed and disallowed regions (Lovell et al., 2000).

Discontinuous B cell epitope prediction:

More than 90% B cell epitopes are discontinuous that is they are present in small segments on linear protein and the brought to close proximity while protein folding. Discontinues B cell epitopes for designed vaccine were predicted by Ellipro online tool

(http://tools.immuneepitope.org/tools/ElliPro) at IEDB. In this tool three algorithms implemented to determine the discontinuous B cell epitopes. 3D structure of input protein is approximated as number of ellipsoid shapes, calculate protrusion index (PI) and clusters neighboring residues. Ellipro defines PI score of each residues based on the center of mass of residue residing outside the largest possible ellipsoid. In consideration with other epitope predicting tools Ellipro gave an AUC value of 0.732, best of all (Ponomarenko et al., 2008).

Protein-protein docking of designed vaccine with TLR2 and TLR4

Molecular docking server HADDOCK (https://bianca.science.uu.nl/haddock2.4/) was used to see the interaction of desined vaccine and TLR4. HADDOCK (High Ambiguity Driven protein-protein DOCKing) is information derived flexible docking server (Zundert et al., 2015). Galaxyrefine 3D model of multi-epitopes chimeric proteins, adjuvant TLR2 (PDB ID: 6NIG) and TLR4 (PDB ID: 4G8A) were uploaded for docking at HDDOCK server keeping all the parameter default. Finally top five models were downloaded and binding affinities were predicted by PRODIGY (PROtein binDIng enerGY prediction) webserver (https://nestor.science.uu.nl/prodigy/) (Xue et al., 2016).

Immune simulation

In silico cloning of vaccine construct was predicted by C-ImmSim server (http://150.146.2.1/C-IMMSIM/index.php). C immSim Server is freely available web based server utilizes position specific scoring matrix (PSSM) for the prediction of immune epitopes and mechine learning techniques for immune interaction (Rapin et al., 2010).

JCat (Java Codon Adaptation Tool) server was utilized for reverse translation and codon optimization. Codon optimization is carried out in order to express the construct in E. coli host. JCat (http://www.prodoric.de/JCat) output display sequence of nucleotide, other important properties of sequence that includes codon adaptaion Index (CAI), percent GC content essential for to assess the protein expression in host (Grote et al., 2005). Finally vaccine construct was cloned in pET30a (+) plasmid vector by adding XhoI and XbaI restriction sites at C and N terminus respectively. Snapgene tool was used to clone the construct to ensure the cloning and expression (Goldberg et al., 2018).

Results

Prediction of B and T cell receptor Epitopes

T cell receptor epitopes

IEDB recommend, MHC-NP, netCTLpan1.1, RANKPEP and netMHCpan3.0 server predicted the potential candidate epitopes from the spike glycoprotein sequence of SARS-CoV19. Epitopes commonly predicated by at least four prediction tools were selected further for analysis parameter. There were four such epitopes with high score and predicted by four different tools (Table 1A). All the four epitopes were immunogenic and antigenic in nature, conserved 100 %, and predicted to be non-allergen (Table 1B). Immunogenicity scores for the predicted epitopes ranges from 0.3858 to 0.0961. All the epitopes were with positive score hence considered for further analysis. All the peptides predicted to be nontoxic by toxipred online toxicity prediction tool.

B cell receptor linear epitopes

After provision of spike protein sequence to the BepiPred server showed average score of 0.407, with 0.695 maximum and 0.163 minimum scores(Figure 1). Other properties like physiology of amino acid residues, accessible surface, hydrophilicity, Fexibility and β turns were also determined (Table 2 and Fig1). Peptides obtained in six different method were arranged according to higher to lower score and 2 % of high score peptides were selected for overlap screening. Regions commonly overlap in at least four prediction methods were selected as potential B epitopes. Regions 250 to 261 and 1246 to 1267 were overlapped regions as per the condition applied. Region 250 to 260 includes BepiPred region 250-260, 251to257 predicted by Karplus-Schulz flexibility and Chau-Fasman β turn prediction methods, and 250 to 257 region predicted by parker hydrophilicity prediction method. Similarly region 1246 to 1267 covers region predicted by BepiPred 2.0 (1252-1267), Emini Surface accessible (1257-1262), Kolaskar and Tongoankar method (1247-1254), Chau-Fasman β turn prediction method (1246-1252) and Parker hydrophilicity method (1256-1262).

Engineering of multi epitope vaccine sequence

Three peptides high scored, predicted for T cell receptors and two peptides for B cell receptors were ligated together by GPGPG or AAY. Additionally OmpA (GenBank: AFS89615.1) was linked at N terminal end of the designed polypeptides by EAAAK linker. For purification

purpose sis histidine residues were added. Final amino acid residues of designed vaccine was 451 when all the five peptides, linkers and adjuvant were ligated (Fig 2).

Prediction of immunogenic properties of designed vaccine

Antigenicity of the designed multi-epitopes vaccine predicted by VaxiJen v2.0 server was 0.7048 with probable antigen annotation for virus as target organism at 0.5 threshold. Antigenicity predicted for same sequence of vaccine by AntigenPro server was 0.765946 with 0.953508 predicted probable solubility upon overexpression. Antigenicity of peptide without adjuvant sequence was 0.671028 with 0.861451 probable solubility of peptides upon overexpression. The results obtained indicates that with and without adjuvant predicted vaccine sequence is potentially antigenic in nature. AllergenFP and AllerTop servers predicted that both vaccine with adjuvant and without adjuvant are probable non allergen with 0.84 and 0.77 Tanimoto similarity indexes respectively.

Prediction physiochemical properties

The computed molecular weight of designed vaccine was 48.49368 KDa and theoretical pI is 7.96. Predicted pI indicates the slight alkalinity of the vaccine (Table 3). The estimated half-lives were 30 hrs in mammalian reticulocytes (in vitro), >20 hours in yeast (in vivo) and >10 hours in Escherichia coli (*in vivo*). Vaccine is highly stable with instability index 24.58 that classify the protein as stable. Estimated aliphatic index 79.02 indicating thermo-stability of the protein (Ikai et al., 1980). The predicted GRAVY score is -0.175 (Table 3). Negative GRAVY score indicates that protein is hydrophilic in nature and will have interaction with water molecules (Kyte et al., 1982).

Secondary structure prediction: Secondary structure predicted by online tool RaptorX property includes 16% alpha helix region, 42% beta sheet region and 40% coil region. Furthermore solvent accessibility of amino acid residues predicted to be 42% exposed, 27% medium exposed and 24% buried. Total 56 residues (12%) are predicted as residues in disordered region. Pictorial representation of secondary structure predicted of final protein by PSIPRED is shown in figure 3.

Tertiary structure prediction refinement and validation

I TASSER predicted the top five 3D structure model of designed vaccine utilizing 10 threading templates. Z score for these 10 templates was ranging from 0.95 to 5.50 indicating good alignment with sequence submitted. C score is critical in the quality of built model which is quantitatively measured. C score ranges in between -5 to 2 signifies the best model quality. C score of top five model ranging from -4.14 to -3.41. Model with high C score i.e -3.41, estimated TM score 0.34±0.11 and estimated RMSD 15.6±3.3 was chosen for further analysis (Fig 4A) (Zhang et al., 2004).

For refinement purpose tertiary structure predicted by I TASSER is initially submitted to ModRefine and finally to GalaxyRefine. Among top five refined models, model 5 found to be the best based on various parameters like GDT-HA 0.8398, RMSD-.696, MolProbit-3.526 and Rama favored 77.1(Fig 4B).

Quality of refined model 5 was validated by Ramachandran plot at RAMPAGE webserver. 79.5 % residues of modelled vaccine are in favored region, 13.1% in allowed region and 7.3% are in outlier region (Fig 4D). Error in the model are analyzed by ProSA web and ERRAT web server. Z score is -1.64 (Fig 4E) and ERRAT score index is 74.387 (Fig 4C). Ramachandran Plot Z score and ERRAT score showed that refined model 5 is of good quality and can be used further.

Dis-continuous epitopes for B cell

Four discontinuous epitopes includes of 265 residues were predicted. The score of the predicted epitopes ranges from 0.58 to 0.749. Shortest and longest discontinuous B cell epitope is of 14 and 97 residues long respectively.

Molecular Docking

HADDOCK web bases server has been utilized to dock the designed vaccine to TLR2 and TLR 4. TLR2/4 eliciting immune response toward designed vaccine was analyzed by conformational change with reference to adjuvant TLR2/4 complexes. Top models from each group with lowest HADDOCK score were selected. HADDOCK score for TLR2/Adjuvant and TLR2/Vaccine were, -89.3 and -95.9 Kcal/mol respectively. In addition relative binding energies were -9.8 and -12.8 for the TLR2/adjuvant and TLR2/vaccine respectively. Similarly HADDOCK scores for TLR4 adjuvant complex and TLR4 vaccine complex were -75.6 and -139.9 Kcal/mol respectively. Relative binding energy of TLR4 adjuvant complex is -10.7 Kcal/mol and -8.2 for

TLR4 vaccine complex. Difference in scores and binding energies of adjuvants and vaccines indicates the change in conformation that may stimulate TLR2/4 receptors. Even there is difference in number interacting residues at the juncture. TLR2/Adjuvant complex have charged-charged 2, charged polar 6, charged-apolar 24, polar-polar 2, polar-apolar 22 and apolar-apolar21. While TLR2/Vaccine have charged-charged 3, charged polar 7, charged-apolar 15, polar-polar 1, polar-apolar 13 and apolar-apolar 119(Fig 5A, B). Similarly TLR4/Adjuvant complex have charged-charged 3, charged polar 10, charged-apolar 21, polar-polar 3, polar-apolar 15 and apolar-apolar20. While TLR vaccine complex have charged-charged 1, charged-polar 9, charged-apolar 10, polar-polar 9, polar-apolar 15 and apolar-apolar 8 interface contacts were observed (Figure 5C, D).

Codon optimization

Enhance expression of nucleotide construct in Escherichia coli, is reverse translated in Jcat online webtool. Total number of nucleotides in the constructs were1350. After codon optimization CAI was 1.0 and average GC content of optimized nucleotide construct was 50.7340 indicating good probable expression of vaccine in E. coli.K12. To clone the gene Xho and XbaI restriction sites are added at N and C terminus of the construct and gene was cloned in pET30a+ plasmid, by using SnapGene software (Fig 6).

Immune simulation

C-ImmSim Immune simulator web server was used for determining ability of vaccine to induce immunity. Results obtained are consistent with the experimental results published elsewhere (Gao et al., 2020). Within first week of vaccine administration primary response was observed by high level of IgM. Secondary and tertiary immune response clearly seen with increase in B cell number, level of IgM, IgG1+IgG2, and IgM+IgG with decreasing concentration of antigen. Different B cell isotypes population were also found increasing indicating isotype switching and memory formation (Fig 6A). B cell activities were also found high, especially B isotype IgM and IgG1, was observed with prominent memory cell formation (Fig 6B). Similarly, cell population of Th and Tc cells are found high along with memory development (Fig 6C, D). In addition, macrophage active was found to be increased consistently after each antigen shots and declined upon antigen clearance (Figure 6E). Another cell from cell mediated immunity, dendritic cells were also found increased (Figure 6F). IFNγ and IL2 expression was found high with low

Simpson Index indicating sufficient immunoglobulin production, suggesting good humoral immune response(Fig 6G). Simpson Index (D) is a measure of diversity. Increase in D over time indicates emergence of different epitope-specific dominant clones of T-cells. The smaller the D value, the lower the diversity.

Discussion

Till date there is no cure available for COVID 19, suggesting urgent requirement of drug or vaccine to control the spread of SAR-CoV2 infections. Advancement in bioinformatics tools, process of vaccine development can be facilitated by identifying potential epitopes for T and B cells that will lead to vaccine for prevention of COVID19. Several reports and animal trials suggest that S-protein can be potential target for vaccine development (Gao et al., 2020; Doremalen et al., 2020; Prompetchara et al., 2020; Xu et al., 2020). There are several reports suggesting the importance of Spike protein in activation of cells of immunity and vaccine development (Wang et al., 2004; Jiang et al., 2005; Tsao et al., 2006). HLA-A2 restricted epitopes from S protein of SARS-CoV, elicit T cell specific immune response in SARS-CoV recovered patients (Wang et al., 2004). Similarly serum samples of SARS-CoV recovered patients were sufficient to neutralize epitope rich region on the spike S2 protein from SARS-CoV (Zhong et al., 2005). Programing of live attenuated form of pathogen is commonly used strategy for the vaccine development. Despite efficacy of such vaccines, reversion of virulence remain a challenge and not utilized in weak immune patients (Plitnick et al., 2013). Here in case SARS-CoV2 due to its high communal transmission vaccine research requires highly skilled workers, sophisticated instrumentation and biosafety level III facility. This may likely the slow down the process of vaccine development, enhanced cost leading to restriction on availability of vaccine for mass population. Advancement in the field of bioinformatics and molecular biology techniques have provided opportunity of development of vaccine with high efficacy, less time for development, low cost of production that may lead to low cost vaccine in time for large population. We designed a multi-epitopes vaccine construct from S-protein of SARS-CoV2. Several online tools were used for the designing of epitopes and thereby a vaccine. From various epitopes predicted by the online server on the basis of common sequence and high score three TCR and two BCR epitopes were selected as part of COVID 19 vaccine. All the five TCR and BCR epitopes were linked by GPGPG and AAY linker peptides (Figure 2). To enhance the

antigenicity of TCR/BCR linked epitopes peptide, adjuvant OmpA were linked by EAAAK linker sequence for a complete vaccine design which was found to restore antigenicity and non-allergen status. (Table 2). For successful vaccine candidate, secondary and tertiary structures characters play an essential role. Secondary structure of designed vaccine contains 16% alpha helixes, 42% beta sheets and 40% coils. Upon refinement tertiary structure of designed vaccine showed improvement to desirable level as indicated by Ramachandran plot. Designed vaccine showed high antigenicity score, no probable allergen.

TLR are essential receptor proteins in activation of innate immune response. That recognize and respond by induction of immune reactions to PAMPs (pathogen associated molecular patterns). Various TLRs have shown to activate immune response to virus through their interaction with nucleic acids and envelopes proteins. TLR3 and TLR7/8 involved in nucleic acids detection, while TLR2 and TLR4 recognize envelope glycoproteins (Xagorari and Chlichlia, 2008). A research report provide evidence of S protein involvement in increased IL8 production through TLR 2 (Dosch et al., 2009). TLR 2 also involved in eliciting innate immune response upon recognition of several other virus components (Finberg et al., 2007). In a comparative binding efficacy of S-proteins to TLRs, TLR 4 showed strongest protein protein interaction with hydrogen and hydrophobic bonds on extracellular domain (Choudhury and Mukherjee, 2020). In molecular docking analysis, TLR2 and TLR4 showed stable protein-protein interaction with designed vaccine compared to adjuvant protein. TLR 4 showed more stable interaction than TLR2, which is consistent with previous report (Choudhury and Mukherjee, 2020). Stable interaction of designed vaccine indicates efficiency of vaccine for activation of TLRs, involved in dendritic cell activation, thereby subsequent antigen processing, and presentation on surface to T cells (Englmeier, 2020; Li et al., 2020). Immune simulation showed typical natural immune response pattern upon multiple exposure to same antigen. Server predicted elevated level of B cell and T cell for longer time on repeated exposure of antigen. Increased level of antiviral cytokine IFNγ and IL2 indicate potential of subsequent activation of T-helper cell thereby high level of Ig production, supporting humoral immune response (Majid and Andleeb 2019).

To validate the designed vaccine in the screening of immune reactivity by serological test, it is necessary to express the construct in preferred host like *E. coli* K12 for recombinant proteins. Codon optimization of was performed to achieve high expression designed vaccine in E coli 12.

Codon adaptability index (CAI=1.0) and GC content (51.66%) of the vaccine construct was optimum for high expression of recombinant protein in host. Immediate step, here after is to express the protein in preferred host and validate results obtained in this report by performing the several immunological assays.

Conclusion

Inadequate number of drugs and time consuming process of drug development, multiplication chain of SARS-CoV2 infection can be control only by vaccine. We utilized tools and techniques of immune-informatics to design a potential vaccine candidate. This vaccine codes epitopes form S protein of SARS-CoV2 virus for T and B cell receptors. Bioinformatics tools have showed effectiveness of the designed vaccine. This is one more time showed that bioinformatics approach can effectively use to develop vaccines in short time and with incurred less cost. Although *in silico* results point out the effectiveness of the vaccine, efficacy need to be analyzed by preforming laboratory experiments and animal model studies.

1. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, Tan KS, Wang DY, Yan Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status. Military Medical Research. 2020 Dec;7(1):1-0.
2. Osman EE, Toogood PL, Neamati N. COVID-19: Living through Another Pandemic. ACS Infectious Diseases. 2020 May 10.
3. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020 Mar 5.
4. Glowacka, I., Bertram, S., Müller, M.A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T.S., He, Y., Gnirss, K. and Niemeyer, D., 2011. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. Journal of virology, 85(9), pp.4122-4134.

5. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. 2020 Mar;12(3):254.
6. Kim E, Erdos G, Huang S, Kenniston TW, Balmert SC, Carey CD, Raj VS, Epperly MW, Klimstra WB, Haagmans BL, Korkmaz E. Microneedle array delivered recombinant coronavirus vaccines: Immunogenicity and rapid translational development. EBioMedicine. 2020 Apr 2:102743.
7. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, Zhou Y, Du L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cellular & molecular immunology. 2020 Mar 19:1-8.
8. Wang N, Shang J, Jiang S, Du L. Subunit vaccines against emerging pathogenic human coronaviruses. Frontiers in Microbiology. 2020 Feb 28;11:298.
9. Snijder EJ, Decroly E, Ziebuhr J. The nonstructural proteins directing coronavirus RNA synthesis and processing. InAdvances in virus research 2016 Jan 1 (Vol. 96, pp. 59-126). Academic Press.
10. Ruch TR, Machamer CE. The coronavirus E protein: assembly and beyond. Viruses. 2012 Mar;4(3):363-82.
11. J Alsaadi EA, Jones IM. Membrane binding proteins of coronaviruses. Future Virology. 2019 Apr;14(4):275-86.
12. Spinney L. When will a coronavirus vaccine be ready? Guardian 17 March, revised 19 March, 2020.
13. World Health Organization. DRAFT landscape of COVID-19 candidate vaccines. https://www.who.int/blueprint/priority-diseases/key-action/novel-coronavirus-landscape-ncov.pdf?ua=1. Accessed 20 March 2020.
14. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19.
15. Chan JF, Yip CC, To KK, Tang TH, Wong SC, Leung KH, Fung AY, Ng AC, Zou Z, Tsoi HW, Choi GK. Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-PCR assay

validated in vitro and with clinical specimens. Journal of Clinical Microbiology. 2020 Apr 23;58(5).

16. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, Wang M. Presumed asymptomatic carrier transmission of COVID-19. Jama. 2020 Apr 14;323(14):1406-7.
17. Yelin I, Aharony N, Shaer-Tamar E, Argoetti A, Messer E, Berenbaum D, Shafran E, Kuzli A, Gandali N, Hashimshony T, Mandel-Gutfreund Y. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. medRxiv. 2020 Jan 1.
18. Li Y, Xia L. Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. American Journal of Roentgenology. 2020 Feb 21:1-7.
19. Kim JY, Choe PG, Oh Y, Oh KJ, Kim J, Park SJ, Park JH, Na HK, Oh MD. The first case of 2019 novel coronavirus pneumonia imported into Korea from Wuhan, China: implication for infection prevention and control measures. Journal of Korean medical science. 2020 Jan 6;35(5).
20. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, Ruan L, Song B, Cai Y, Wei M, Li X. A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19. New England Journal of Medicine. 2020 Mar 18.
21. Gautret P, Lagier JC, Parola P, Meddeb L, Mailhe M, Doudier B, Courjon J, Giordanengo V, Vieira VE, Dupont HT, Honoré S. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. International journal of antimicrobial agents. 2020 Mar 20:105949.
22. Molina JM, Delaugerre C, Goff JL, Mela-Lima B, Ponscarme D, Goldwirt L, de Castro N. No evidence of rapid antiviral clearance or clinical benefit with the combination of hydroxychloroquine and azithromycin in patients with severe COVID-19 infection. Med Mal Infect. 2020 Mar 30;10.
23. Chorin E, Wadhwani L, Magnani S, Dai M, Shulman E, Nadeau-Routhier C, Knots R, Bar-Cohen R, Kogan E, Barbhaiya C, Aizer A. QT Interval Prolongation and Torsade De Pointes in Patients with COVID-19 treated with Hydroxychloroquine/Azithromycin. medRxiv. 2020 Jan 1.
24. Gautret P, Lagier JC, Parola P, Meddeb L, Sevestre J, Mailhe M, Doudier B, Aubry C, Amrane S, Seng P, Hocquart M. Clinical and microbiological effect of a combination of hydroxychloroquine and azithromycin in 80 COVID-19 patients with at least a six-day

follow up: A pilot observational study. Travel Medicine and Infectious Disease. 2020 Apr 11:101663.

25. Rosenberg ES, Dufort EM, Udo T, et al. Association of Treatment With Hydroxychloroquine or Azithromycin With In-Hospital Mortality in Patients With COVID-19 in New York State [published online ahead of print, 2020 May 11]. JAMA. 2020;10.1001/jama.2020.8630. doi:10.1001/jama.2020.8630
26. Choy KT, Wong AY, Kaewpreedee P, Sia SF, Chen D, Hui KP, Chu DK, Chan MC, Cheung PP, Huang X, Peiris M. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. Antiviral research. 2020 Apr 3:104786.
27. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophysical journal. 2011 Nov 16;101(10):2525-34.
28. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic acids research. 2015 Jul 1;43(W1):W174-81.
29. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, Ma K, Xu D, Yu H, Wang H, Wang T. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. Bmj. 2020 Mar 26;368.
30. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, Liu X, Wei L, Truelove SA, Zhang T, Gao W. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. MedRxiv. 2020 Jan 1.
31. Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CA, Munday JD, Abbott S. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health. 2020 Mar 25.
32. Mitjà O, Clotet B. Use of antiviral drugs to reduce COVID-19 transmission. The Lancet Global Health. 2020 May 1;8(5):e639-40.
33. Spychalski P, Błażyńska-Spychalska A, Kobiela J. Estimating case fatality rates of COVID-19. The Lancet Infectious Diseases. 2020 Mar 31.
34. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. Asian Pac J Allergy Immunol. 2020 Mar 1;38(1):1-9.

35. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, Bucci E, Piacentini M, Ippolito G, Melino G. COVID-19 infection: the perspectives on immune responses. 2020; 1451-1454.
36. van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham J, Port J, Avanzato V, Bushmaker T, Flaxman A, Ulaszewska M, Feldmann F. ChAdOx1 nCoV-19 vaccination prevents SARS-CoV-2 pneumonia in rhesus macaques. bioRxiv. 2020 Jan 1.
37. Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. Nucleic acids research. 2019 Jul 2;47(W1):W402-7.
38. Wang S, Li W, Liu S, Xu J. RaptorX-Property: a web server for protein structure property prediction. Nucleic acids research. 2016 Jul 8;44(W1):W430-5.
39. Wang S, Peng J, Ma J, Xu J. Protein secondary structure prediction using deep convolutional neural fields. Scientific reports. 2016 Jan 11;6(1):1-1.
40. Zheng W, Zhang C, Bell EW, Zhang Y. I-TASSER gateway: A protein structure and function prediction server powered by XSEDE. Future Generation Computer Systems. 2019 Oct 1;99:73-85.
41. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010 Apr;5(4):725.
42. Xu D, Zhang Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophysical journal. 2011 Nov 16;101(10):2525-34.
43. Ko J, Park H, Heo L, Seok C. GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research. 2012 May 30;40(W1):W294-7.
44. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein science. 1993 Sep;2(9):1511-9.
45. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007 Jul 1;35(suppl_2):W407-10.
46. Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 50, 437e450.

47. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, Peters B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC bioinformatics. 2008 Dec 1;9(1):514.
48. Van Zundert GC, Rodrigues JP, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond AS, van Dijk M, De Vries SJ, Bonvin AM. The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. Journal of molecular biology. 2016 Feb 22;428(4):720-5.
49. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. Bioinformatics. 2016 Dec 1;32(23):3676-8.
50. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research. 2005 Jul 1;33(suppl_2):W526-31.
51. Goldberg MF, Roeske EK, Ward LN, Pengo T, Dileepan T, Kotov DI, Jenkins MK. Salmonella persist in activated macrophages in T cell-sparse granulomas but are contained by surrounding CXCR3 ligand-positioned Th1 cells. Immunity. 2018 Dec 18;49(6):1090-102.
52. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One. 2010;5(4).
53. Ikai A. Thermostability and aliphatic index of globular proteins. The Journal of Biochemistry. 1980 Oct;88(6):1895-8.
54. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. Journal of molecular biology. 1982 May 5;157(1):105-32.
55. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins: Structure, Function, and Bioinformatics. 2004 Dec 1;57(4):702-10.
56. Gao Q, Bao L, Mao H, Wang L, Xu K, Yang M, Li Y, Zhu L, Wang N, Lv Z, Gao H. Development of an inactivated vaccine candidate for SARS-CoV-2. Science. 2020 May 6.

57. van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham J, Port J, Avanzato V, Bushmaker T, Flaxman A, Ulaszewska M, Feldmann F. ChAdOx1 nCoV-19 vaccination prevents SARS-CoV-2 pneumonia in rhesus macaques. bioRxiv. 2020 Jan 1.
58. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, Zhong W, Hao P. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Science China Life Sciences. 2020 Mar;63(3):457-60.
59. Dosch SF, Mahajan SD, Collins AR. SARS coronavirus spike protein-induced innate immune response occurs via activation of the NF-κB pathway in human monocyte macrophages in vitro. Virus research. 2009 Jun 1;142(1-2):19-27.
60. Tsao YP, Lin JY, Jan JT, Leng CH, Chu CC, Yang YC, Chen SL. HLA-A∗ 0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. Biochemical and biophysical research communications. 2006 May 26;344(1):63-71.
61. Wang B, Chen H, Jiang X, Zhang M, Wan T, Li N, Zhou X, Wu Y, Yang F, Yu Y, Wang X. Identification of an HLA-A* 0201–restricted CD8+ T-cell epitope SSp-1 of SARS-CoV spike protein. Blood. 2004 Jul 1;104(1):200-6.
62. Jiang S, He Y, Liu S. SARS vaccine development. Emerging infectious diseases. 2005 Jul;11(7):1016.
63. Holmes KV. SARS-associated coronavirus. New England Journal of Medicine. 2003 May 15;348(20):1948-51.
64. Wang, Y.D., Sin, W.Y.F., Xu, G.B., Yang, H.H., Wong, T.Y., Pang, X.W., He, X.Y., Zhang, H.G., Ng, J.N.L., Cheng, C.S.S. and Yu, J., 2004. T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. Journal of virology, 78(11), pp.5612-5618.
65. Zhong, X., Yang, H., Guo, Z.F., Sin, W.Y.F., Chen, W., Xu, J., Fu, L., Wu, J., Mak, C.K.G., Cheng, C.S.S. and Yang, Y., 2005. B-cell responses in patients who have recovered from severe acute respiratory syndrome target a dominant site in the S2 domain of the surface spike glycoprotein. Journal of virology, 79(6), pp.3401-3408.
66. Plitnick LM. Global regulatory guidelines for vaccines. In Nonclinical Development of Novel Biologics, Biosimilars, Vaccines and Specialty Biologics 2013 Jan 1 (pp. 225-241). Academic Press.

67. Michel-Todó L, Reche PA, Bigey P, Pinazo MJ, Gascón J, Alonso-Padilla J. In silico design of an epitope-based vaccine ensemble for Chagas disease. Frontiers in Immunology. 2019 Nov 22;10:2698.
68. Xagorari A, Chlichlia K. Toll-like receptors and viruses: induction of innate antiviral immune responses. The open microbiology journal. 2008;2:49.
69. Nain Z, Karim MM, Sen MK, Adhikari UK. Structural basis and designing of peptide vaccine using PE-PGRS family protein of Mycobacterium ulcerans—An integrated vaccinomics approach. Molecular Immunology. 2020 Apr 1;120:146-63.
70. Choudhury A, Mukherjee S. In silico studies on the comparative characterization of the interactions of SARS-CoV-2 spike glycoprotein with ACE-2 receptor homologs and human TLRs. Journal of Medical Virology. 2020 May 8.
71. Englmeier L. A theory on SARS-CoV-2 susceptibility: Reduced tlr7-activity as a mechanistic link between men, obese and elderly.
72. Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, Pan P, Wang W, Hu D, Liu X, Zhang Q. Coronavirus infections and immune responses. Journal of medical virology. 2020 Apr;92(4):424-32.
73. Majid M, Andleeb S. Designing a multi-epitopic vaccine against the enterotoxigenic Bacteroides fragilis based on immunoinformatics approach. Scientific Reports. 2019 Dec 24;9(1):1-5.
74. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, Keşmir C, Peters B. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS computational biology. 2013 Oct;9(10).
75. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP, Open Source Drug Discovery Consortium. In silico approach for predicting toxicity of peptides and proteins. PloS one. 2013;8(9).
76. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics. 2007 Dec 1;8(1):4.
77. Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics. 2014 Mar 15;30(6):846-51.

78. Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine. 2007 Jan 15;25(5):856-66.
79. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v. 2—a server for in silico prediction of allergens. Journal of molecular modeling. 2014 Jun 1;20(6):2278.

Figure and Table legends

Table 1: A. Common peptides predicted by at least four TCR epiptope prediction tools; B. immunogenic properties of that peptides. In case of immunogenic properties peptide number 3 has negative score for antigenicity hence neglected for further analysis.

Table 2: B cell receptor epitope prediction, A: Common B cell receptor epitopes predicted by various methods B: Score obtained B cell receptor epitopes for Spike glycoprotein of SARS-CoV2

Table 3. Analysis of the physicochemical and immunological properties of the designed vaccine for SARS-CoV2

Figure1: Graphical representation of prediction of B cell epitopes Immunogenic peptide predicted by various tools probable epitopes are scored above the threshold, showed here as yellow. A: Parker hydrophilicity prediction, B:Chau-Fasman β turn prediction C:Karplus-Schulz flexibility, D: Emini Surface accessible , E: Kolaskar and Tongoankar, F: Bepipred Linear Epitope Prediction 2.0

Figure 2. Schematic presentation of the final multi-epitope vaccine peptide. The 599-amino acid long peptide sequence containing an adjuvant (Red) at the amino terminal end linked with the multi-epitope sequence through an EAAAK linker (black). BCR epitopes (yellow) are linked using GPGPG linkers (cyan) while the TCR epitopes (blue) are linked with the help of AAY linkers (green). A 6x-His tag is added at the Carboxy terminus for purification and identification purposes.

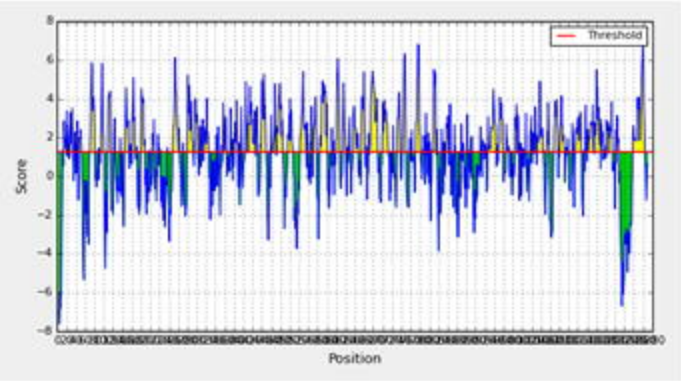
Figure 3: Graphical representation of secondary structure characters; 42% beta sheet region, 40% coil region and 16% alpha helix region

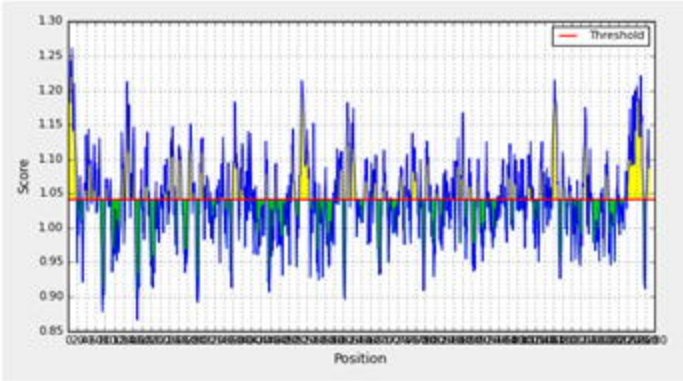
Figure 4: Vaccine tertiary structure modelling, refinement and validation A: tertiary structure of ITASSER modelled multi-epitopes Vaccine along with OmpA as adjuvant; B: Refined 3D model (green) after GalaxyRefine superimposed on ITASSER crude model (violet) C: Errata score for refined model with graphical representation, D: Ramachandran plot for refined 3D model of vaccine, E: Z-score -1.64 by ProSA webtool.

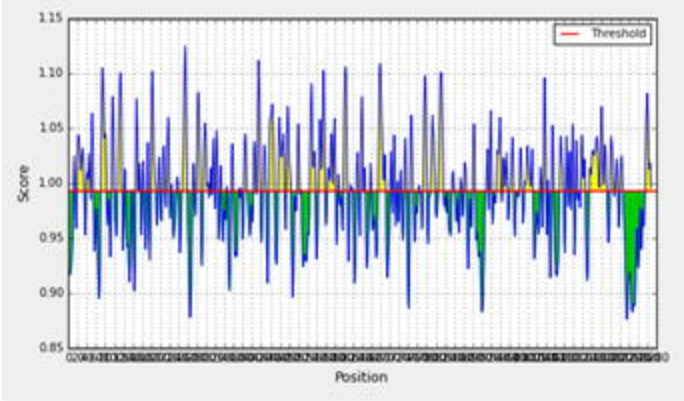
Figure 5: Haddock docking cartoon models showing interaction A: TLR2-Vaccine; TLR2 in red and Vaccine in blue; B:TLR2-Adjuvant; TLR2 in red and Adjuvant in blue; C: TLR4-Vaccine; TLR4 in red and Vaccine in red; D: TLR4-Adjuvant; TLR4 in red and Adjuvant in blue

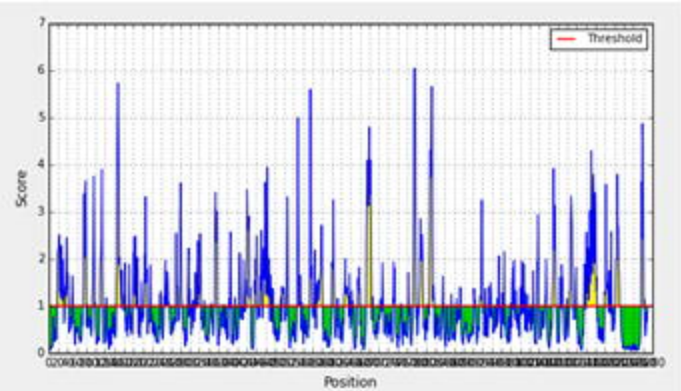
Figure 6: *In silico* cloning of designed vaccine construct in pET30a+ expression vector where yellow arc segment represent cloned construct and remaining segment represent vector backbone.

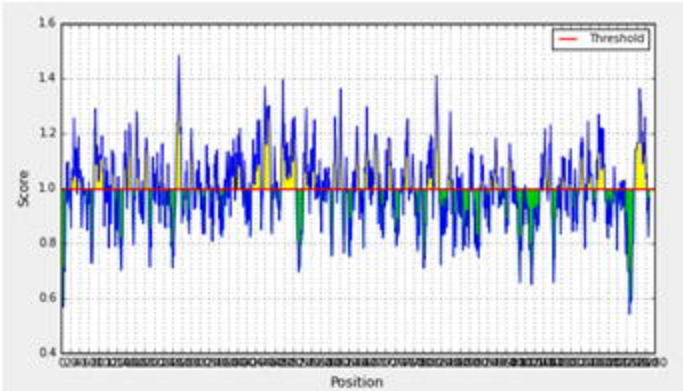
Figure 7: Immune response simulation by designed vaccine, A: Immune response upon antigen exposure; B: B cell population per state(cells/mm^3); C: TH cell population state (cells/mm^3); D: TC cell population state (cells/mm^3); E: activity of macrophage population in three subsequent immune responses; F: Dendritic cell population state (cells/mm^3) upon antigen exposure; G:Cytokine induced by three injection of vaccine after one week interval, The main plot shows cytokine levels after the injections. The insert plot shows IL-2 level with the Simpson index, D indicated by the dotted line.

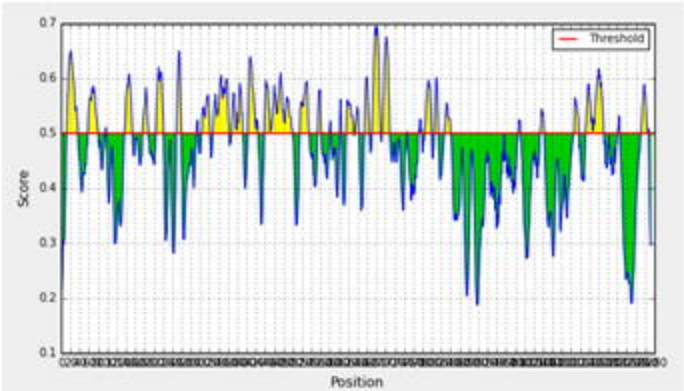

Threshold
Score
Position

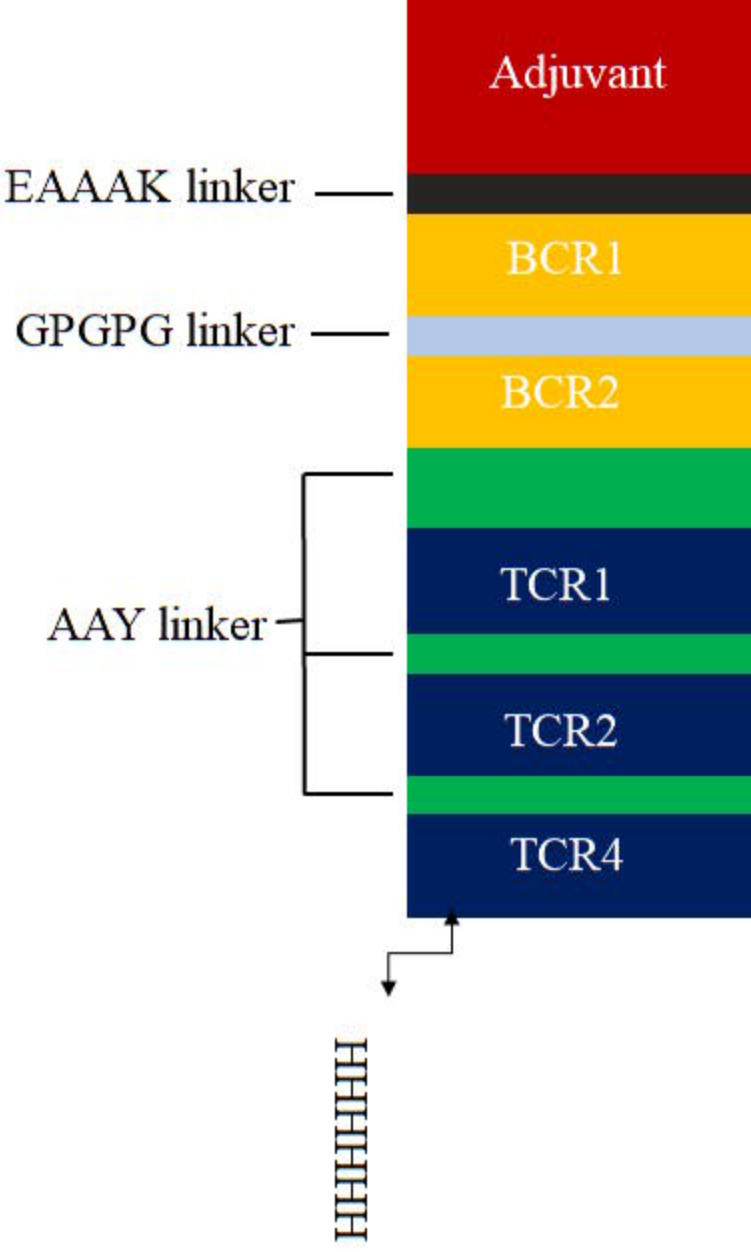

Threshold
Score
Position

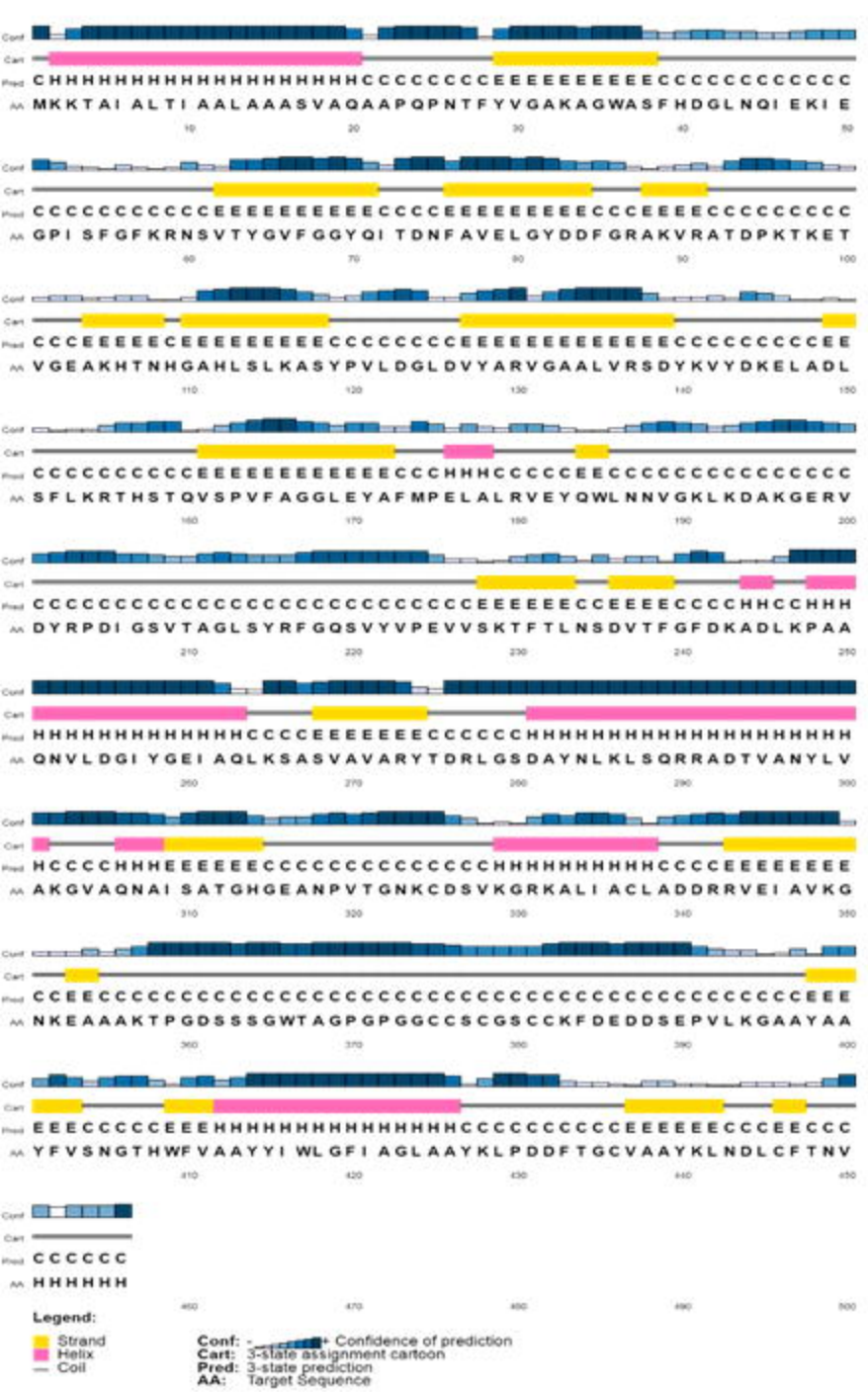

Threshold
Score
Position

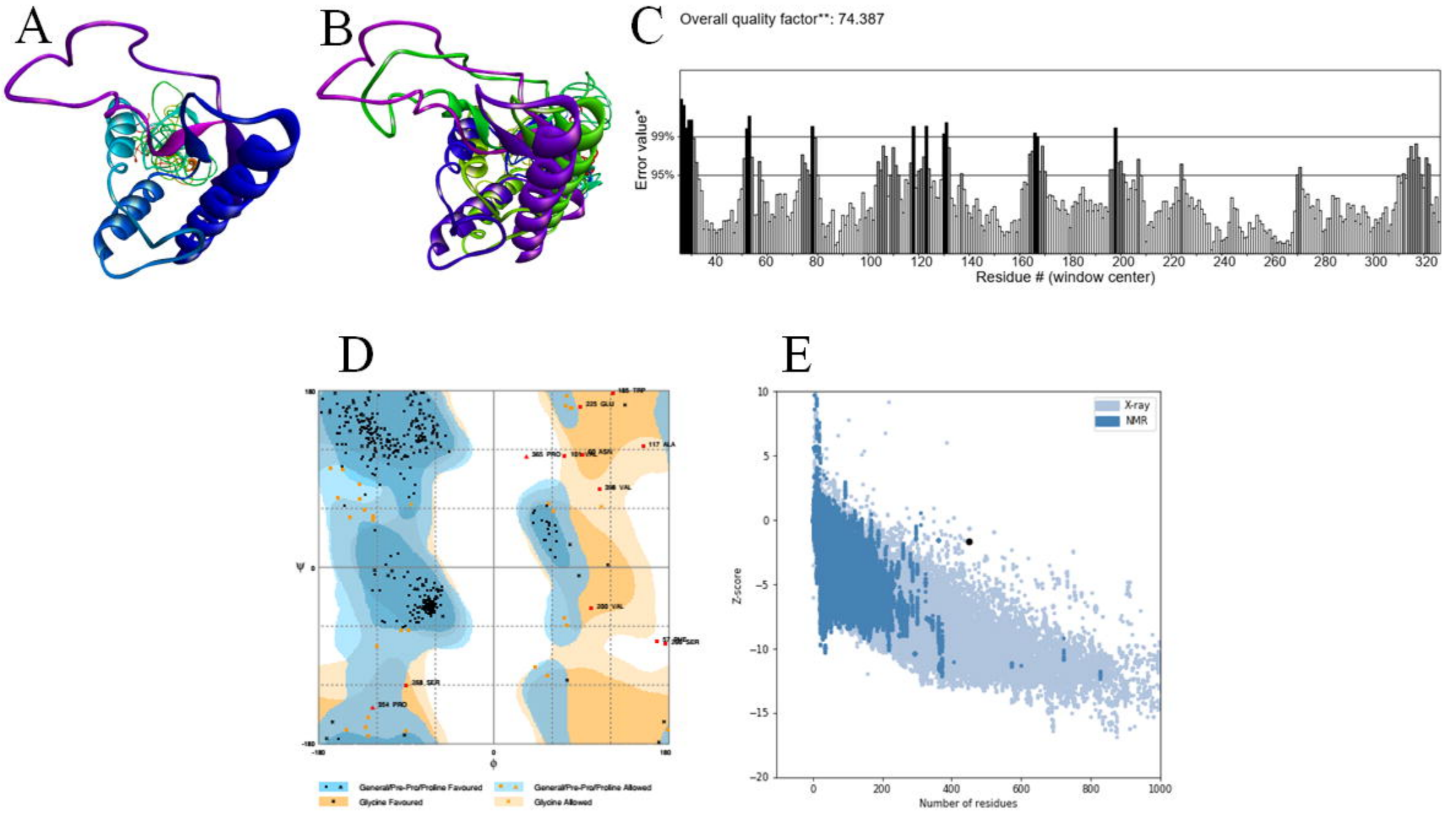

Threshold
Score
Position

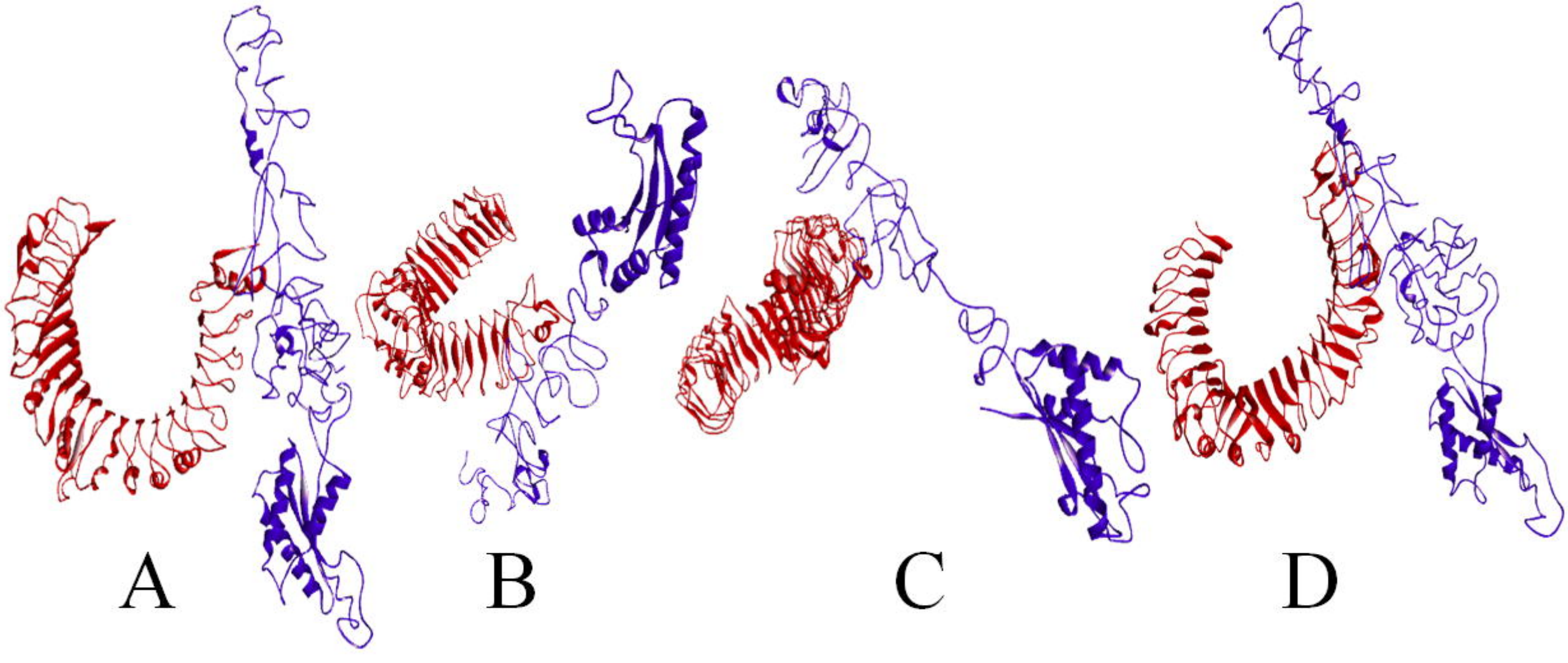

Threshold
Score
Position


Threshold
Score
Position

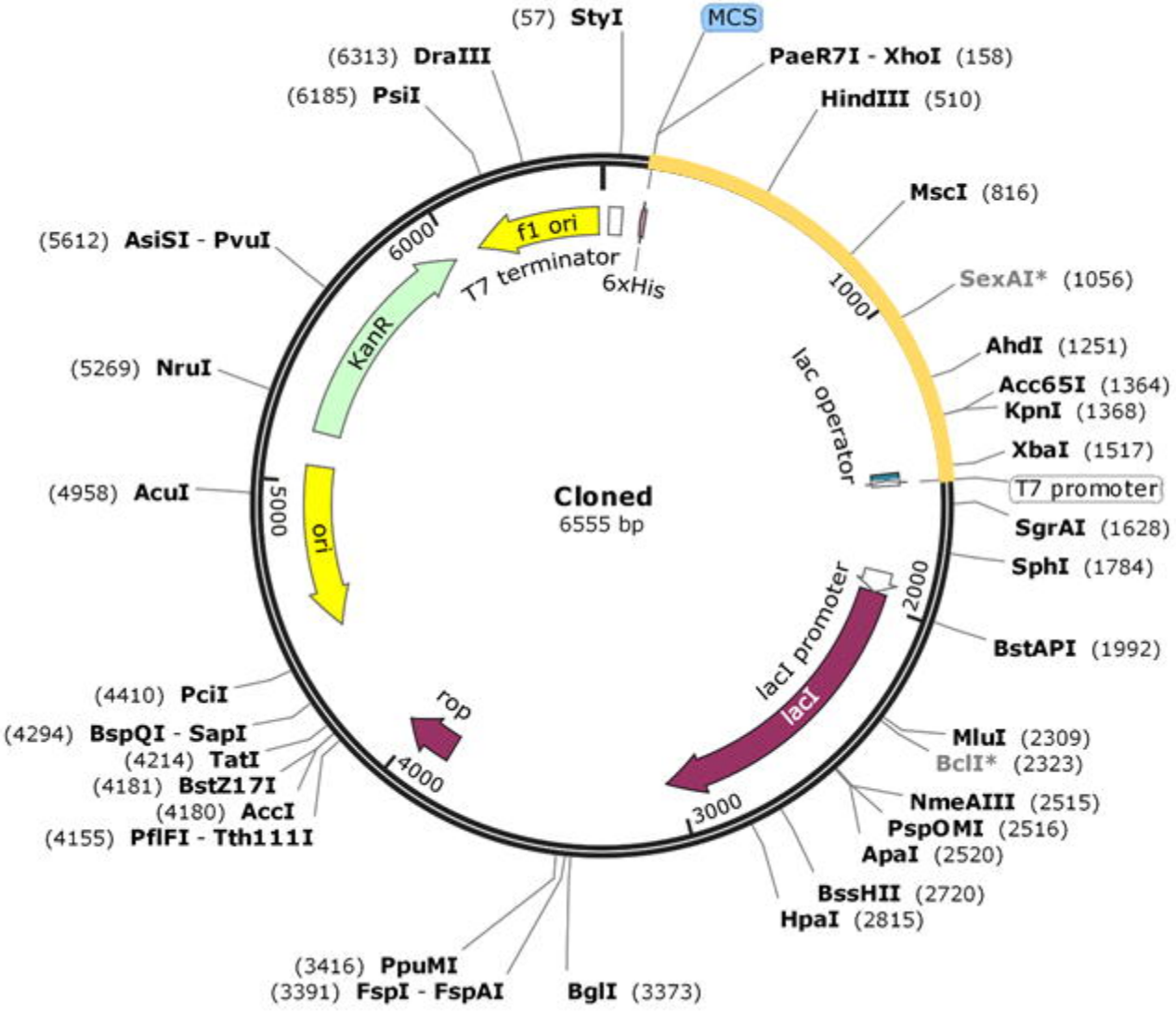

Adjuvant
EAAAK linker
BCR1
GPGPG linker
BCR2
AAY linker
TCR1
TCR2
TCR4
HHHHHH

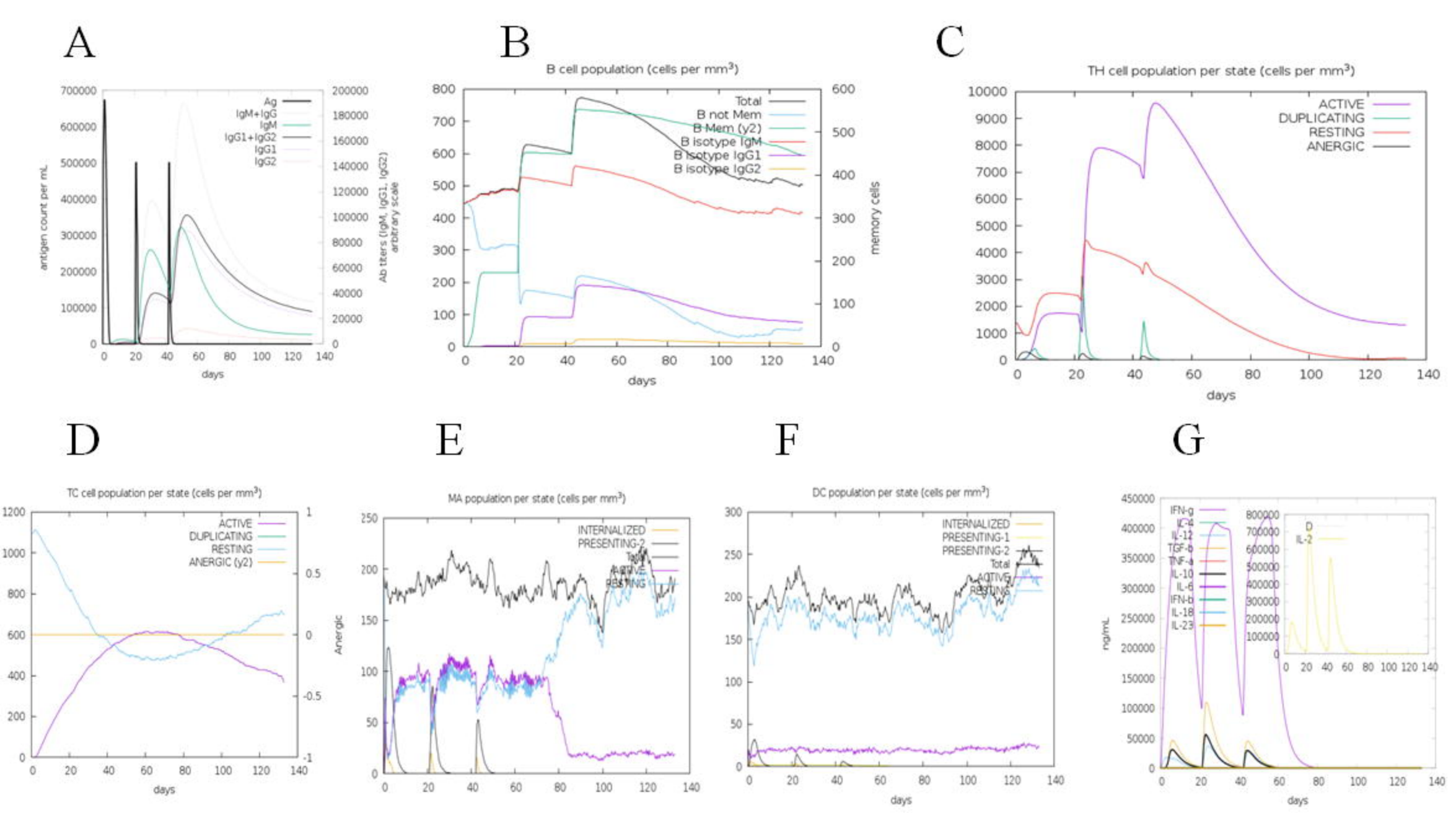

Conf
Cart
Pred
AA
CHHHHHHHHHHHHHHHHHHCCCCCCCCCEEEEEEEEEEEECCCCCCCCCCCC
MKKTAI ALTI AALAAASVAQAAPQPNTFYVGAKAGWASFHDGLNQI EKI E
10 20 30 40 50
CCCCCCCCCCCEEEEEEEEEEEECCCCEEEEEEEEEECCCEEEEECCCCCCCCC
GPI SFGFKRNSVTYGVFGGYQI TDNFAVELGYDDFGRAKVRATDPKTKET
60 70 80 90 100
CCCEEEEEECEEEEEEEEECCCCCCCCCEEEEEEEEEEEEEECCCCCCCCCEE
VGEAKHTNHGAHLSLKASYPVLDGLDVYARVGAALVRSDYKVYDKELADL
110 120 130 140 150
CCCCCCCCCCCEEEEEEEEEEEEECCCHHHCCCCCEECCCCCCCCCCCCCCC
SFLKRTHSTQVSPVFAGGLEYAFMPELALRVEYQWLNNVGKLKDAKGERV
160 170 180 190 200
CCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEECCEEEECCCCHHCCHHH
DYRPDI GSVTAGLSYRFGQSVYVPEVVSKTFTLNSDVTFGFDKADLKPAA
210 220 230 240 250
HHHHHHHHHHHHCCCCEEEEEEECCCCCCHHHHHHHHHHHHHHHHHHHH
QNVLDGI YGEI AQLKSASVAVARYTDRLGSDAYNLKLSQRRADTVANYLV
260 270 280 290 300
HCCCCHHHEEEEEEECCCCCCCCCCCCCCHHHHHHHHHHCCCCEEEEEEEE
AKGVAQNAI SATGHGEANPVTGNKCDSVKGRKALI ACLADDRRVEI AVKG
310 320 330 340 350
CCEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEE
NKEAAAKTPGDSSSGWTAGPGPGGCCSCGSCCKFDEDDSEPVLKGAAYAA
360 370 380 390 400
EEECCCCCEEEHHHHHHHHHHHHHCCCCCCCCCCEEEEEECCCEECCC
YFVSNGTHWFVAAYYI WLGFI AGLAAYKLPDDFTGCVAAYKLNDLCFTNV
410 420 430 440 450
CCCCCC
HHHHHH
460 470 480 490 500
Legend:
Strand
Helix
Coil
Conf: Confidence of prediction
Cart: 3-state assignment cartoon
Pred: 3-state prediction
AA: Target Sequence


A
B
C
Overall quality factor**: 74.387
Error value*
99%
95%
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
Residue # (window center)
D
E
X-ray
NMR
Z-score
Number of residues
General/Pre-Pro/Proline Favoured
General/Pre-Pro/Proline Allowed
Glycine Favoured
Glycine Allowed


A
B
C
D

Created with SnapGene®




A
Ag
IgM+IgG
IgM
IgG1+IgG2
IgG1
IgG2
antigen count per mL
Ab titers (IgM, IgG1, IgG2) arbitrary scale
days
B
B cell population (cells per mm3)
Total
B not Mem
B Mem (y2)
B isotype IgM
B isotype IgG1
B isotype IgG2
memory cells
days
C
TH cell population per state (cells per mm3)
ACTIVE
DUPLICATING
RESTING
ANERGIC
days
D
TC cell population per state (cells per mm3)
ACTIVE
DUPLICATING
RESTING
ANERGIC (y2)
days
E
MA population per state (cells per mm3)
INTERNALIZED
PRESENTING-2
Anergic
F
DC population per state (cells per mm3)
INTERNALIZED
PRESENTING-1
PRESENTING-2
Total
days
G
IFN-g
IL-4
IL-12
TGF-b
TNF-a
IL-10
IL-6
IFN-b
IL-18
IL-23
IL-2
ng/mL
days

A

Epitopes	Sequence	Amino acid number	Net CTLpan	RankPep	IEDB	MHC NP	Net MHC 4.0
TCR1	WYIWLGFIAGLIAI	1214-1227	0.8	42.06	0.32	0.2815	0.3
TCR2	FVSNGTHWFV	1095-1105	0.3	-	0.07	0.344184	0.17
TCR3	KLPDDFTGCV	423-432	0.4	39.12	-0.68	0.591582	0.3
TCR4	KLNDLCFTNV	385-393	0.4	34.15	0.13	0.710803	0.9

B

Epitopes	Sequence	Immunogenicity	Conservancy	Toxicity	Antigenicity	Allergen
TCR1	WYIWLGFIAGLIAI	0.38584	100	Non toxic	0.5798	Probable Non allergen
TCR2	FVSNGTHWFV	0.18796	100	Non toxic	0.5982	Probable Non allergen
TCR3	KLPDDFTGCV	0.1001	100	Non toxic	-0.5308	Probable Non allergen
TCR4	KLNDLCFTNV	0.0961	100	Non toxic	2.6927	Probable Non allergen

A

Method	Average Score	Maximum Score	Minimum Score
BepiPred 2.0	0.470	0.183	0.695
Emini Surface accessible	1.0	0.042	6.051
Chua-Fasman β turn	0.997	0.541	1.484
Karplus-Schulz flexibility	0993	0.876	1.125
Kolaskar–Tongaonkar antigenicity	1.041	0.866	1.261
Parker hydrophilicity	1.238	-7.629	7.743

B

BCR epitopes	Region	Sequence
BCR1	250-260 (11 aa)	TPGDSSSGWTA
BCR2	1246-1267 (22 aa)	GCCSCGSCCKFDEDDSEPVLKG

Characters	score
Antigenicity (AntigenPro and VaxiJen v2.0)	0.765946 and 0.7048 (Probable antigen)
Allergen (AllergenFP and AllerTop)	0.84 and 0.77 (probable non allergen)
Possible Solubility	0.9557226
Number of amino acids	454
Molecular weight	48.49368
pI	7.94
The instability index (II)	24.58
Aliphatic index	79.02
Grand average of hydropathicity (GRAVY)	-0.172