In Silico design and characterization of multi-epitopes vaccine for SARS-CoV2 from its spike proteins

Gunderao H Kathwate*

Department of Biotechnology, Savitribai Phule Pune University, Pune (MS), India

*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 was rich in effective BCR and TCR epitopes screened from the sequence of S-protein of SARS-CoV2. Predicted BCR and TCR epitopes were antigenic in nature non-toxic and probably non-allergen. Modelled and refined tertiary structure was predicted as valid for further use. Protein-Protein interaction prediction of TLR2/4 and designed vaccine indicates promising binding. Designed multiepitope vaccine has induced 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 indicates 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-CoV2, B cell receptor, T cell receptor, 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(1). 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(2). 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(3,4) and highly sophisticated, time consuming diagnostic methods(5–7). As of 13 May 2020, 7 million confirmed cases with more than 400,000 deaths worldwide(8). 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(9). In South Korea, Lopinavir/Ritonavir combination was found significantly effective in lowering of viral load to no detectible or little SAR-CoV2 titer(10). But in another group study same combination was totally ineffective beyond standard care(11). Another drug pair drug, hydroxychloroquine, an antimalarial drug and azithromycin reported to found effectively associated with reduction of viral load(12,13). But QT interval prolongation that may cause life threating arrhythmia(14) and in large scale study both drugs and drug alone compared with neither drugs was not associated with mortality(15,16). In a drug repurposing study, remdesivir, lopinavir, emetine, and homoharringtonine have significantly inhibited replication of SAR-CoV2(17). But this was *in vitro* study, clinical trial reports 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 COVID19 epidemic (18–20). Various epidemiological surveys provide the evidences that naturally acquired immunity can eliminate the SARS-CoV2 titer(21,22). More than 80% patients develop mild and 14% develop severe symptoms of COVID19 with zero case fatality rate(23).

Presently there is no vaccine available against COVID19 disease. Efforts are being taken for the development of effective vaccine all over the globe(24–27). All these vaccines are under

development(28) and may require at least 1to 1.5 year to hit the market(29). 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(30).

SARS-CoV2 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(31). This is a non-segmented positive strand RNA virus with envelop. 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(32). Envelope (E) and membrane (M) proteins and instrumental in virus assembly and virulence promotion(33,34). E and M proteins are effective in induction of immune response(24). Spike (S), a structural protein is responsible for binding to receptor on host cell, Angiotensin converting enzyme 2(35). Thereby proteolytic cleavage by TMPRSS2 allows subsequent entry into cell through endocytosis(36). S protein also found to be involved in activation of T cell response(37,38). Entire S protein expressed in chimpanzee adeno 38 (ChAd)-vector showed protection against SARS-CoV2 in mice and *rhesus* macaques(39). 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-epitopes 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 (40).

Analysis of immunological properties

Immunogenicity, Toxicity, Allergen, Conservancy, and Antigenicity 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(40) 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(41). Online server AllergenFP v.1.0 (http://ddgpharmfac.net/AllergenFP/) was used to predict peptides as allergens(42). 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 (43). 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 (http://www.cbs.dtu.dk/services/BepiPred/) uses a hidden Markov model and

propensity scale method. Similarly, other properties are also being consider to predict good B cell epitopes(44). Those properties are calculated by different methods at IEDB server (http://tools.iedb.org/bcell/)like Kolaskar–Tongaonkar antigenicity scale provide physiology of the amino acid residues(45), Emini Surface accessible score for accessible surface of the epitope(46), Secondary structure of epitopes also has role in antigenicity. Karplus-Schulz flexibility score(47) and Chou-Fasman β turns methods(48) were used for flexibility and β turns prediction respectively. Parker hydrophilicity prediction method (49)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(43,50). 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(51). AllergenFP, is descriptor-based fingerprint, alignment free tool for allergenicity prediction. The tool use four step algorithm. In first step, proteins are described based on 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 (42).

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(52). 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(53). This server utilizes DeepCNF (Deep Convolutional Neural Fields), a new machine learning model that predicts secondary structure and solvent accessibility and disorder regions simultaneously(54). It accomplishes 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(55). It is top ranked 3D protein structure web server in community wide CASP experiments(56).

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(57). 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(58).

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(59). ProSA web server calculate the quality score as Z score that should fall in a characteristic range(60). 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(61). Ramachandran plot displays energetically allowed and disallowed dihedral 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(62).

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 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 residue 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(63,64).

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 designed vaccine and TLR4. HADDOCK (High Ambiguity Driven protein-protein DOCKing) is information derived flexible docking server(65). 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 by PRODIGY (PROtein binDIng enerGY prediction) webserver (https://nestor.science.uu.nl/prodigy/) was utilized for prediction of binding affinities(66).

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 machine learning techniques for immune interaction(67).

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 adaptation Index (CAI), percent GC content essential for to assess the protein expression in host(68). 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(69).

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-CoV2. 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 (Fig1). 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 methods 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 regions 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 were 454 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(70). The predicted GRAVY score is -0.175 (Table 3). Negative GRAVY score indicates that protein is hydrophilic in nature and will interact with water molecules(71).

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)(72).

For refinement purpose tertiary structure predicted by I-TASSER is initially submitted to ModRefiner 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 include 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 TLR4/vaccine complex has 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(73). 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 (Fig 6E). Another cell from cell mediated immunity, dendritic cells were also found increased (Fig 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 COVID19, 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(22,73–76). There are several reports suggesting the importance of Spike protein in activation of cells of immunity and vaccine development(77–79). HLA-A2 restricted epitopes from S protein of SARS-CoV2, elicit T cell specific immune response in SARS-CoV2 recovered patients(27). Similarly serum samples of SARS-CoV2 recovered patients were sufficient to neutralize epitope rich region on the spike S2 protein from SARS-CoV2 (80). Programing of live attenuated form of pathogen is commonly used strategy for the vaccine development. Despite efficacy of such vaccines, reversion of virulence remains a challenge and not utilized in weak immune patients(81,82). 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 based on common sequence and high score three TCR and two BCR epitopes were selected as part of COVID19 vaccine. All the five TCR and BCR epitopes were linked by GPGPG and AAY linker peptides (Fig 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(83). Interaction of S protein with TLR2 increase the production of IL8 in human monocyte macrophages(84). TLR 2 also involved in eliciting innate immune response upon recognition of several other virus components(85,86). 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(87). In molecular docking analysis, TLR2 and TLR4 showed stable protein-protein interaction with designed vaccine compared to adjuvant protein. TLR 2 showed more stable interaction than TLR4, which is consistent with previous report(87). 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(88). 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(89).

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. 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.

Reference

1. Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- A n update on the status. Vol. 7, Military Medical Research. BioMed Central Ltd.; 2020.

2. Sohrabi C, Alsafi Z, O'Neill N, Khan M, … AK-IJ of, 2020 undefined. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S1743919120301977

3. Gandhi M, Yokoe DS, Havlir D V. Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. N Engl J Med. 2020 Apr 24;

4. Bai Y, Yao L, Wei T, Tian F, Jin D, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. jamanetwork.com [Internet]. [cited 2020 Jun 10]; Available from: https://jamanetwork.com/journals/jama/article-abstract/2762028

5. Chan J, Yip C, To K, … TT-J of C, 2020 undefined. 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. Am Soc Microbiol [Internet]. [cited 2020 Jun 10]; Available from: https://jcm.asm.org/content/58/5/e00310-20.abstract

6. Yelin I, Aharony N, Shaer-Tamar E, … AA-A rights reserved. N reuse, 2020 undefined. Evaluation of COVID-19 RT-qPCR test in multi-sample pools. medRxiv.

7. Li Y, Xia Li LY. Role of Chest CT in Diagnosis and Management. AJR [Internet]. 2020 Jun 1 [cited 2020 Jun 10];214:1280–6. Available from: www.ajronline.org

8. Organization WH, others. Coronavirus disease 2019 (? COVID-19)?: situation report, 88. 2020;

9. Mitjà O, Health BC-TLG, 2020 undefined. Use of antiviral drugs to reduce COVID-19 transmission. thelancet.com [Internet]. [cited 2020 Jun 10]; Available from:

https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(20)30114-5/fulltext

10. The Efficacy of Lopinavir Plus Ritonavir and Arbidol Against Novel Coronavirus Infection - Full Text View - ClinicalTrials.gov [Internet]. [cited 2020 Jun 10]. Available from: https://clinicaltrials.gov/ct2/show/study/NCT04252885

11. Cao B, Wang Y, Wen D, Liu W, Wang J, Fan G, et al. A trial of lopinavir-ritonavir in adults hospitalized with severe covid-19. N Engl J Med. 2020 May 7;382(19):1787–99.

12. Gautret P, Lagier J, Parola P, … LM-I journal of, 2020 undefined. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S0924857920300996

13. Gautret P, Lagier J, Parola P, … LM-T medicine and, 2020 undefined. 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. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S1477893920301319

14. Chorin E, Wadhwani L, Magnani S, Dai M, rhythm ES-H, 2020 undefined. QT Interval Prolongation and Torsade De Pointes in Patients with COVID-19 treated with Hydroxychloroquine/Azithromycin. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S1547527120304355

15. Rosenberg E, Dufort E, Udo T, Jama LW-, 2020 undefined. Association of treatment with hydroxychloroquine or azithromycin with in-hospital mortality in patients with COVID-19 in New York state. jamanetwork.com [Internet]. [cited 2020 Jun 10]; Available from: https://jamanetwork.com/journals/jama/article-abstract/2766117

16. Molina JM, Delaugerre C, Le Goff J, Mela-Lima B, Ponscarme D, Goldwirt L, et al. Journal Pre-proof No Evidence of Rapid Antiviral Clearance or Clinical Benefit with the Combination of Hydroxychloroquine and Azithromycin in Patients with Severe COVID-19 Infection. icpcovid.com [Internet]. 2020 [cited 2020 Jun 10]; Available from: https://doi.org/10.1016/j.medmal.2020.03.006

17. Choy K, Wong A, Kaewpreedee P, Sia S, research DC-A, 2020 undefined. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S016635422030200X

18. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. bmj.com [Internet]. [cited 2020 Jun 10]; Available from: https://www.bmj.com/content/368/bmj.m1091

19. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S1473309920302875

20. Prem K, Liu Y, Russell T, … AK-TLP, 2020 undefined. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S2468266720300736

21. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X. COVID-19 infection: the perspectives on immune responses. 2020 [cited 2020 Jun 10]; Available from: https://www.nature.com/articles/s41418-020-0530-3?fbclid=IwAR2ZxeCwG6GQEZTC1GQbizfu0d5i1wqIREmK5IKRgeR2TsUc1dg8ta1GYyA

22. Prompetchara E, Ketloy C, Palaga T. Allergy and Immunology Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. apjai-journal.org [Internet]. [cited 2020 Jun 10]; Available from: http://apjai-journal.org/wp-content/uploads/2020/03/1.pdf?fbclid=IwAR1kVlFnoc6mobqEy7FDz_93fX__AaY8ZgE4TUwtjzhh8oVqgs82rgf3QoU

23. Spychalski P, … AB-S-TLI, 2020 undefined. Estimating case fatality rates of COVID-19. thelancet.com [Internet]. [cited 2020 Jun 10]; Available from: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)302462/fulltext

24. Faraz Ahmed S, Quadeer AA, McKay MR. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. mdpi.com [Internet]. [cited 2020 Jun 10]; Available from: http://www.allelefrequencies.net/

25. Kim E, Erdos G, Huang S, Kenniston T, … SB-PAP, 2020 undefined. Microneedle array delivered recombinant coronavirus vaccines: Immunogenicity and rapid translational development, EBioMedicine. 2020.

26. Tai W, He L, Zhang X, Pu J, … DV-C& molecular, 2020 undefined. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor. nature.com [Internet]. [cited 2020 Jun 10]; Available from: https://www.nature.com/articles/s41423-020-0400-4?report=reader

27. Wang B, Chen H, Jiang X, Zhang M, Wan T, Li N, et al. Identification of an HLA-A*0201-restricted CD8 T-cell epitope SSp-1 of SARS-CoV spike protein. Blood [Internet]. 2004 [cited 2020 Jun 10];104:200–6. Available from: http://bimas.dcrt.nih.gov/molbio/

28. World WHO-, 2020 undefined. DRAFT landscape of COVID-19 candidate vaccines.

29. Amanat F, Immunity FK-, 2020 undefined. SARS-CoV-2 vaccines: status report. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S1074761320301205

30. Shey RA, Ghogomu SM, Esoh KK, Nebangwa ND, Shintouo CM, Nongley NF, et al. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Sci Rep. 2019;9(1):1–18.

31. Osman EEA, Toogood PL, Neamati N. COVID-19: Living through Another Pandemic. ACS Infect Dis. 2020 May 10;

32. Snijder E, Decroly E, research JZ-A in virus, 2016 undefined. The nonstructural proteins directing coronavirus RNA synthesis and processing. Elsevier [Internet]. [cited 2020 Jun 10]; Available from:

https://www.sciencedirect.com/science/article/pii/S0065352716300471

33. Ruch TR, Machamer CE. The Coronavirus E Protein: Assembly and Beyond. Viruses [Internet]. 2012 [cited 2020 Jun 10];4:363–82. Available from: www.mdpi.com/journal/viruses

34. Alsaadi EA, Jones IM. Membrane binding proteins of coronaviruses. Futur Med [Internet]. 2019 Apr 1 [cited 2020 Jun 10];14(4):275–86. Available from: www.futuremedicine.com

35. Hoffmann M, Kleine-Weber H, Schroeder S, Cell NK-, 2020 undefined. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S0092867420302294

36. Glowacka I, Bertram S, Müller MA, Allen P, Soilleux E, Pfefferle S, et al. Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. J Virol [Internet]. 2011 [cited 2020 Jun 10];85(9):4122–34. Available from: http://jvi.asm.org/

37. Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, et al. Coronavirus infections and immune responses. Vol. 92, Journal of Medical Virology. John Wiley and Sons Inc.; 2020. p. 424–32.

38. Grifoni A, Weiskopf D, Ramirez SI, Smith DM, Crotty S, Sette A. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell [Internet]. 2020 [cited 2020 Jun 10]; Available from: https://doi.org/10.1016/j.cell.2020.05.015

39. Doremalen N van, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, Port JR, et al. ChAdOx1 nCoV-19 vaccination prevents SARS-CoV-2 pneumonia in rhesus macaques. bioRxiv. 2020;2020.05.13.093195.

40. Calis J, Maybeno M, … JG-Pl computational, 2013 undefined. Properties of MHC class I presented peptides that enhance immunogenicity. ncbi.nlm.nih.gov [Internet]. [cited 2020

Jun 10]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3808449/

41. In silico approach for predicting toxicity of peptides and proteins. ncbi.nlm.nih.gov [Internet]. [cited 2020 Jun 10]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3772798/

42. Dimitrov I, Naneva L, Doytchinova I, Bioinformatics IB-, 2014 undefined. AllergenFP: allergenicity prediction by descriptor fingerprints. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/bioinformatics/article-abstract/30/6/846/286438

43. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007 Jan 5;8.

44. Jespersen M, Peters B, … MN-N acids, 2017 undefined. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/nar/article-abstract/45/W1/W24/3787843

45. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990 Dec 10;276(1–2):172–4.

46. Emini EA, Hughes,' J V, Perlow DS, Boger2 J. Induction of Hepatitis A Virus-Neutralizing Antibody by a Virus-Specific Synthetic Peptide Downloaded from [Internet]. JOURNAL OF VIROLOGY. 1985 [cited 2020 Jun 10]. Available from: http://jvi.asm.org/

47. Karplus PA, Schulz GE. Prediction of chain flexibility in proteins - A tool for the selection of peptide antigens. Naturwissenschaften. 1985 Apr;72(4):212–3.

48. of PC-A in enzymology and related areas, 1978 undefined. Prediction of the secondary structure of proteins from their amino acid sequence. ci.nii.ac.jp [Internet]. [cited 2020 Jun 10]; Available from: https://ci.nii.ac.jp/naid/80013857431/

49. Parker JMR, Guo D, Hodges RS. 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. 1986;25(19):5425–32.

50. Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine [Internet]. 2007 [cited 2020 Jun 10];25:856–66. Available from: http://www.elsevier.com/locate/permissionusematerial

51. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2 - A server for in silico prediction of allergens. J Mol Model. 2014;20(6).

52. Smialowski P, Doose G, Torkler P, Kaufmann S, Frishman D. PROSO II - A new method for protein solubility prediction. FEBS J. 2012 Jun;279(12):2192–200.

53. Wang S, Li W, Liu S, research JX-N acids, 2016 undefined. RaptorX-Property: a web server for protein structure property prediction. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/nar/article-abstract/44/W1/W430/2499323

54. Wang S, Peng J, Ma J, reports JX-S, 2016 undefined. Protein secondary structure prediction using deep convolutional neural fields. nature.com [Internet]. [cited 2020 Jun 10]; Available from: https://www.nature.com/articles/srep18962/

55. Roy A, Kucukural A, protocols YZ-N, 2010 undefined. I-TASSER: a unified platform for automated protein structure and function prediction. nature.com [Internet]. [cited 2020 Jun 10]; Available from: https://www.nature.com/nprot/journal/v5/n4/abs/nprot.2010.5.html

56. Zheng J, Lin X, Wang X, Zheng L, Lan S, Jin S, et al. In silico analysis of epitope-based vaccine candidates against hepatitis B virus polymerase protein. Viruses. 2017;9(5).

57. Xu D, journal YZ-B, 2011 undefined. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S0006349511012458

58. Ko J, Park H, Heo L, research CS-N acids, 2012 undefined. GalaxyWEB server for protein structure prediction and refinement. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/nar/article-abstract/40/W1/W294/1078340

59. Wang W, Xia M, Chen J, Deng F, Yuan R, brief XZ-D in, et al. Data set for phylogenetic

tree and RAMPAGE Ramachandran plot analysis of SODs in Gossypium raimondii and G. arboreum. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S2352340916303134

60. Wiederstein M, research MS-N acids, 2007 undefined. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/nar/article-abstract/35/suppl_2/W407/2920938

61. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci. 1993;2(9):1511–9.

62. Prisant M, Richardson J, Proteins DR-, 2003 undefined. Structure validation by Calpha geometry: Phi, psi and Cbeta deviation.

63. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 2008 Dec 2;9.

64. Ponomarenko J V., Bourne PE. Antibody-protein interactions: Benchmark datasets and prediction tools evaluation. BMC Struct Biol. 2007;7.

65. Zundert G Van, Rodrigues J, … MT-J of molecular, 2016 undefined. The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S0022283615005379

66. Xue L, Rodrigues J, Kastritis P, … AB-, 2016 undefined. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/bioinformatics/article-abstract/32/23/3676/2525629

67. Rapin N, Lund O, Bernaschi M, One FC-Pl, 2010 undefined. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. ncbi.nlm.nih.gov [Internet]. [cited 2020 Jun 10]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2855701/

68. Grote A, Hiller K, Scheer M, … RM-N acids, 2005 undefined. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/nar/article-abstract/33/suppl_2/W526/2505472

69. Goldberg M, Roeske E, Ward L, Immunity TP-, 2018 undefined. Salmonella persist in activated macrophages in T cell-sparse granulomas but are contained by surrounding CXCR3 ligand-positioned Th1 cells. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S1074761318304746

70. Biochemistry AI-TJ of, 1980 undefined. Thermostability and aliphatic index of globular proteins. academic.oup.com [Internet]. [cited 2020 Jun 10]; Available from: https://academic.oup.com/jb/article-abstract/88/6/1895/773432

71. Kyte J, Doolittle RF. A Simple Method for Displaying the Hydropathic Character of a Protein [Internet]. Vol. 157, J. Mol. Biol. 1982 [cited 2020 Jun 10]. Available from: https://www.sciencedirect.com/science/article/pii/0022283682905150

72. Zheng W, Zhang C, Bell E, Computer YZ-FG, 2019 undefined. I-TASSER gateway: A protein structure and function prediction server powered by XSEDE. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S0167739X18314705

73. Gao Q, Bao L, Mao H, Wang L, Xu K, Yang M, et al. Development of an inactivated vaccine candidate for SARS-CoV-2. Science (80-). 2020;(May):eabc1932.

74. Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, Port JR, Avanzato V, et al. ChAdOx1 nCoV-19 vaccination prevents SARS-CoV-2 pneumonia in rhesus macaques Neeltje van Doremalen. biorxiv.org [Internet]. [cited 2020 Jun 10]; Available from: https://doi.org/10.1101/2020.05.13.093195

75. Zhu F-C, Li Y-H, Guan X-H, Hou L-H, Wang W-J, Li J-X, et al. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. Lancet [Internet]. 2020;6736(20):1–10. Available from:

https://linkinghub.elsevier.com/retrieve/pii/S0140673620312083

76. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, et al. Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Springer [Internet]. 2020 [cited 2020 Jun 10];63(3):457–60. Available from: https://doi.org/10.1007/s11427-020-1637-5

77. Wang N, Shang J, Jiang S, Du L. Subunit Vaccines Against Emerging Pathogenic Human Coronaviruses. Vol. 11, Frontiers in Microbiology. Frontiers Media S.A.; 2020.

78. Jiang S, He Y, diseases SL-E infectious, 2005 undefined. SARS vaccine development. ncbi.nlm.nih.gov [Internet]. [cited 2020 Jun 10]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3371787/

79. Tsao Y, Lin J, Jan J, Leng C, … CC-B and, 2006 undefined. HLA-A∗ 0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S0006291X06006905

80. Zhong X, Yang H, Guo Z-F, Fion W-Y, Chen W, Xu J, et al. 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. J Virol [Internet]. 2005 [cited 2020 Jun 10];79(6):3401–8. Available from: http://jvi.asm.org/

81. Plitnick L, Herzyk D. Nonclinical development of novel biologics, biosimilars, vaccines and specialty biologics [Internet]. 2013 [cited 2020 Jun 10]. Available from: https://books.google.com/books?hl=en&lr=&id=U0wEcrLkozYC&oi=fnd&pg=PP1&dq=Global+regulatory+guidelines+for+vaccines.+In+Nonclinical+Development+of+Novel+Biologics,+Biosimilars,+Vaccines+and+Specialty+Biologics+&ots=uupokzuPyo&sig=Esm9wpqT23KubtsZRVU5LKi0ZBQ

82. Biologics LP-D of N, Biosimilars undefined, Vaccines undefined, 2013 undefined. Global regulatory guidelines for vaccines. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/B9780123948106000095

83. Xagorari A, journal KC-T open microbiology, 2008 undefined. Toll-like receptors and viruses: induction of innate antiviral immune responses. ncbi.nlm.nih.gov [Internet]. [cited 2020 Jun 10]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2593046/

84. Dosch S, Mahajan S, research AC-V, 2009 undefined. SARS coronavirus spike protein-induced innate immune response occurs via activation of the NF-κB pathway in human monocyte macrophages in vitro. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S0168170209000124

85. Wang J, Kurt-Jones E, microbiology RF-C, 2007 undefined. Innate immunity to respiratory viruses. Wiley Online Libr [Internet]. [cited 2020 Jun 10]; Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-5822.2007.00961.x

86. Finberg R, infection EK-J-M and, 2004 undefined. Viruses and Toll-like receptors. Elsevier [Internet]. [cited 2020 Jun 10]; Available from: https://www.sciencedirect.com/science/article/pii/S1286457904002953

87. 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. J Med Virol. 2020 May 8;

88. Englmeier L. A theory on SARS-CoV-2 susceptibility: Reduced tlr7-activity as a mechanistic link between men, obese and elderly. 2020 [cited 2020 Jun 10]; Available from: https://osf.io/preprints/2r5fc/

89. Majid M, Andleeb S. Designing a multi-epitopic vaccine against the enterotoxigenic Bacteroides fragilis based on immunoinformatics approach. Sci Rep. 2019;9(1):1–15.

Table and figure legends

Table 1: A. Common peptides predicted by at least four TCR epitopes 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.

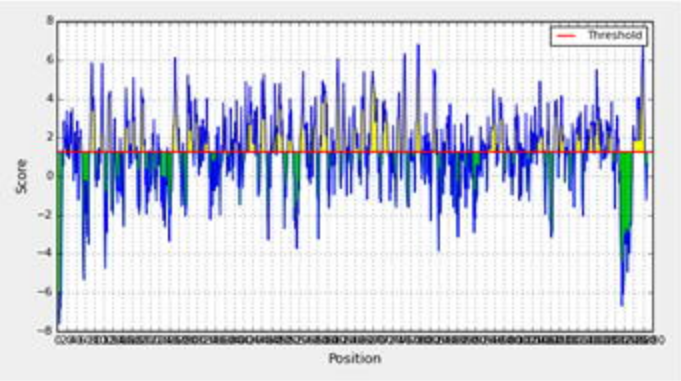
Figure 3: Graphical representation of secondary structure characters; 42% beta sheet region, 40% coil region and 16% alpha helix region

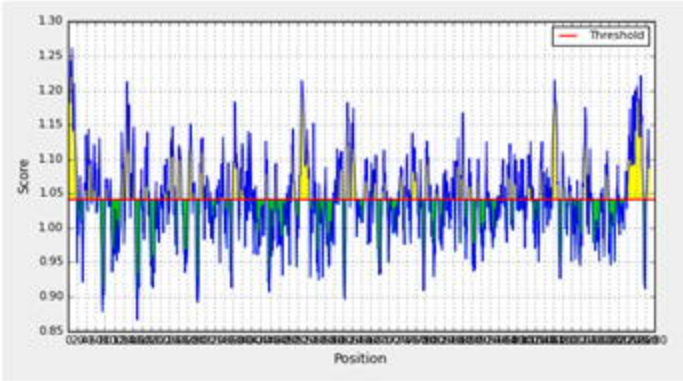
Figure 4: Vaccine tertiary structure modelling, refinement and validation A: tertiary structure of ITASSER modelled multi-epitopes Vaccine along with OmpA as adjuvant; B: Refined 3D model (green) after GalaxyRefine superimposed on ITASSER crude model (violet) C: Errata score for refined model with graphical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

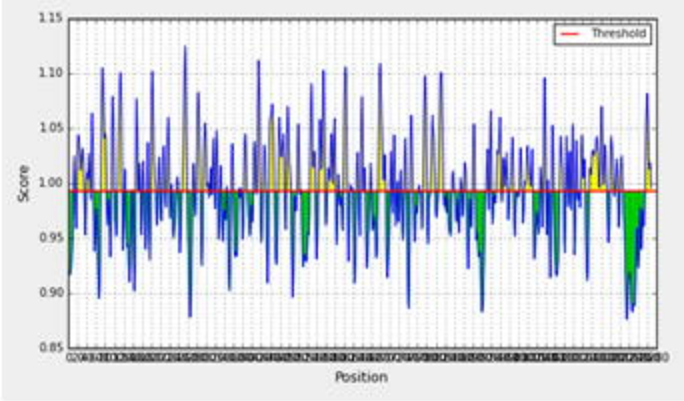
Figure 5: Haddock docking cartoon models showing interaction A: TLR2-Vaccine; TLR2 in red and Vaccine in blue; B:TLR2-Adjuvant; TLR2 in red and Adjuvant in blue; C: TLR4-Vaccine; TLR4 in red and Vaccine in red; D: TLR4-Adjuvant; TLR4 in red and Adjuvant in blue

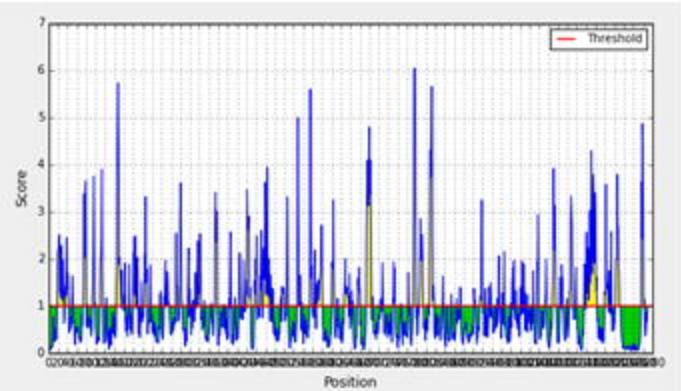
Figure 6: *In silico* cloning of designed vaccine construct in pET30a+ expression vector where yellow arc segment represents cloned construct and remaining segment represent vector backbone.

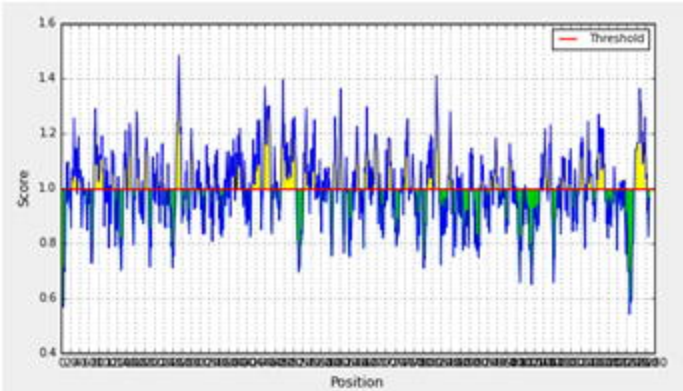
Figure 7: Immune response simulation by designed vaccine, A: Immune response upon antigen exposure; B: B cell population per state(cells/mm^3); C: TH cell population state (cells/mm^3); D: TC cell population state (cells/mm^3); E: activity of macrophage population in three subsequent immune responses; 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

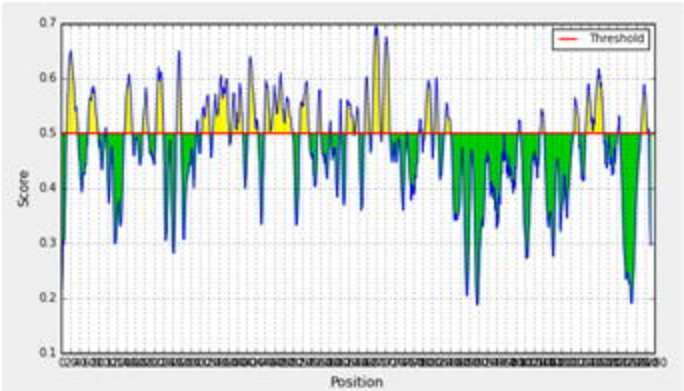

Threshold
Score
Position

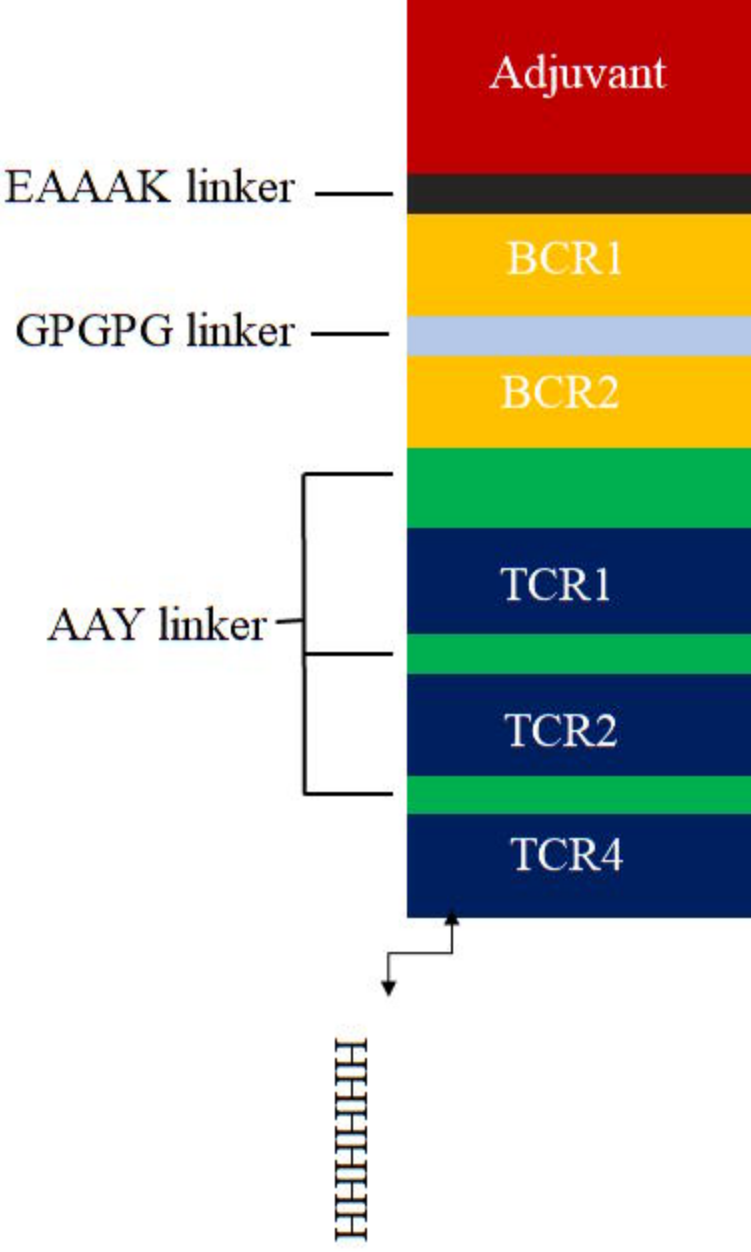

Threshold
Score
Position

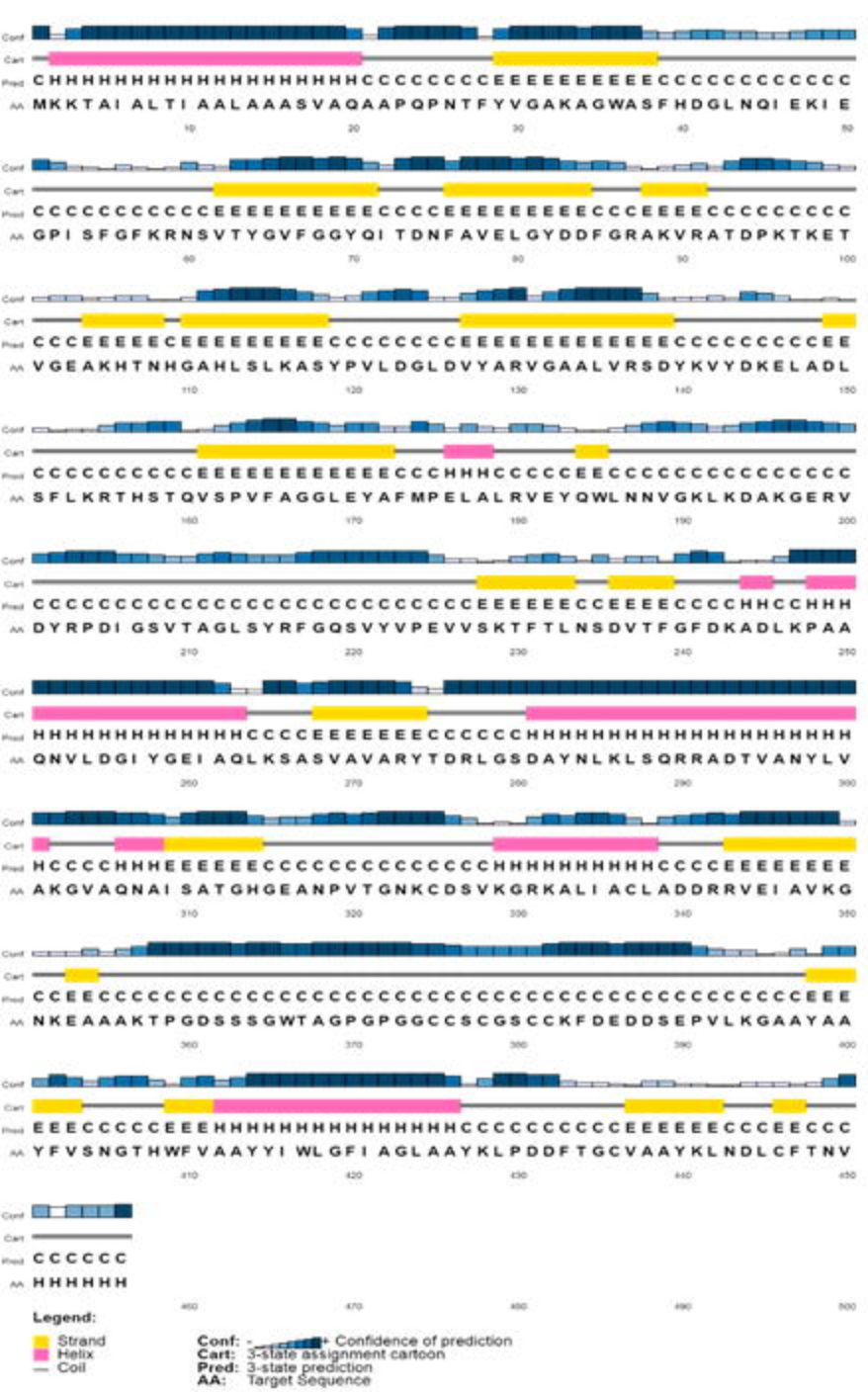

Threshold
Score
Position

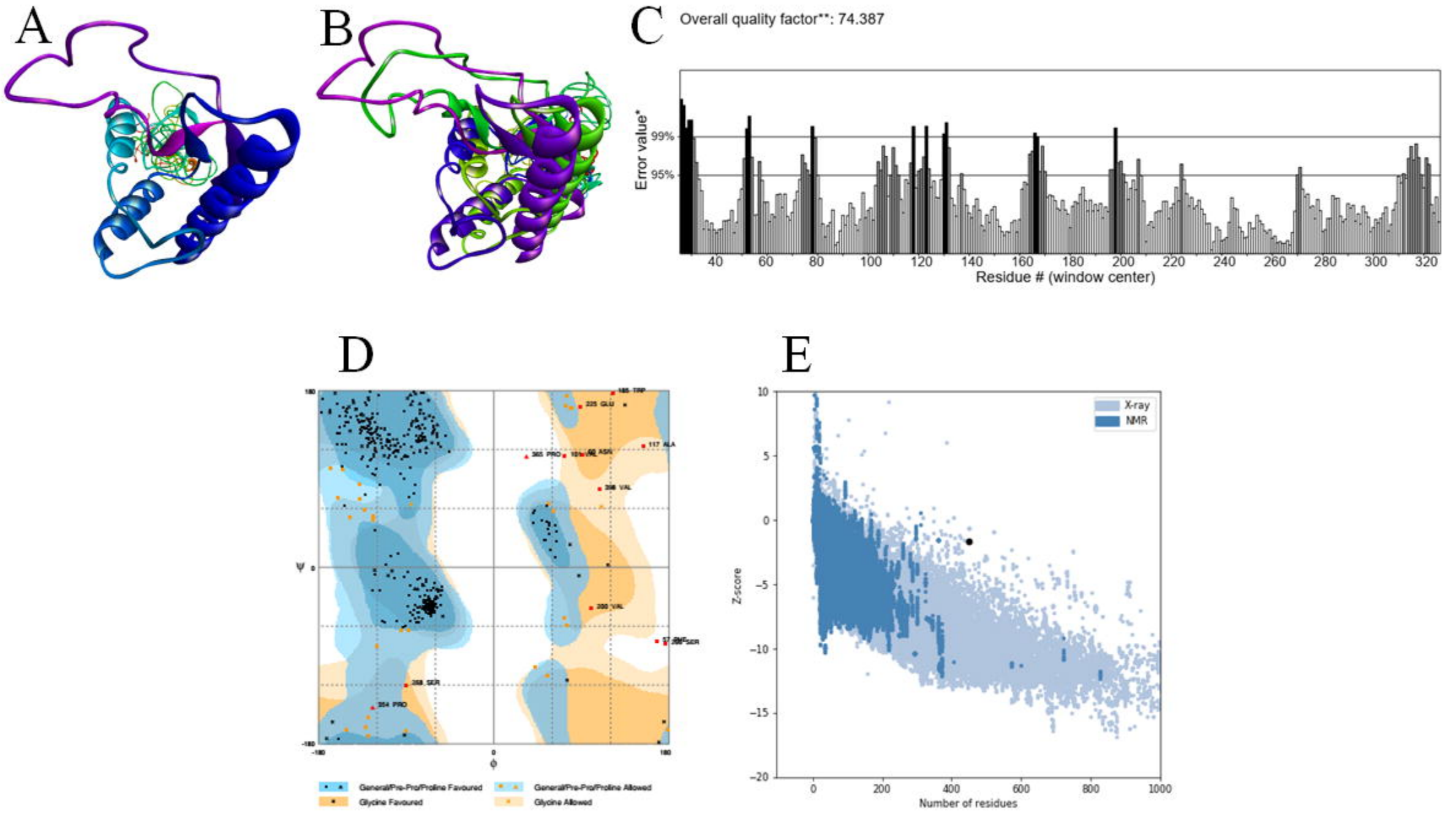

Threshold
Score
Position

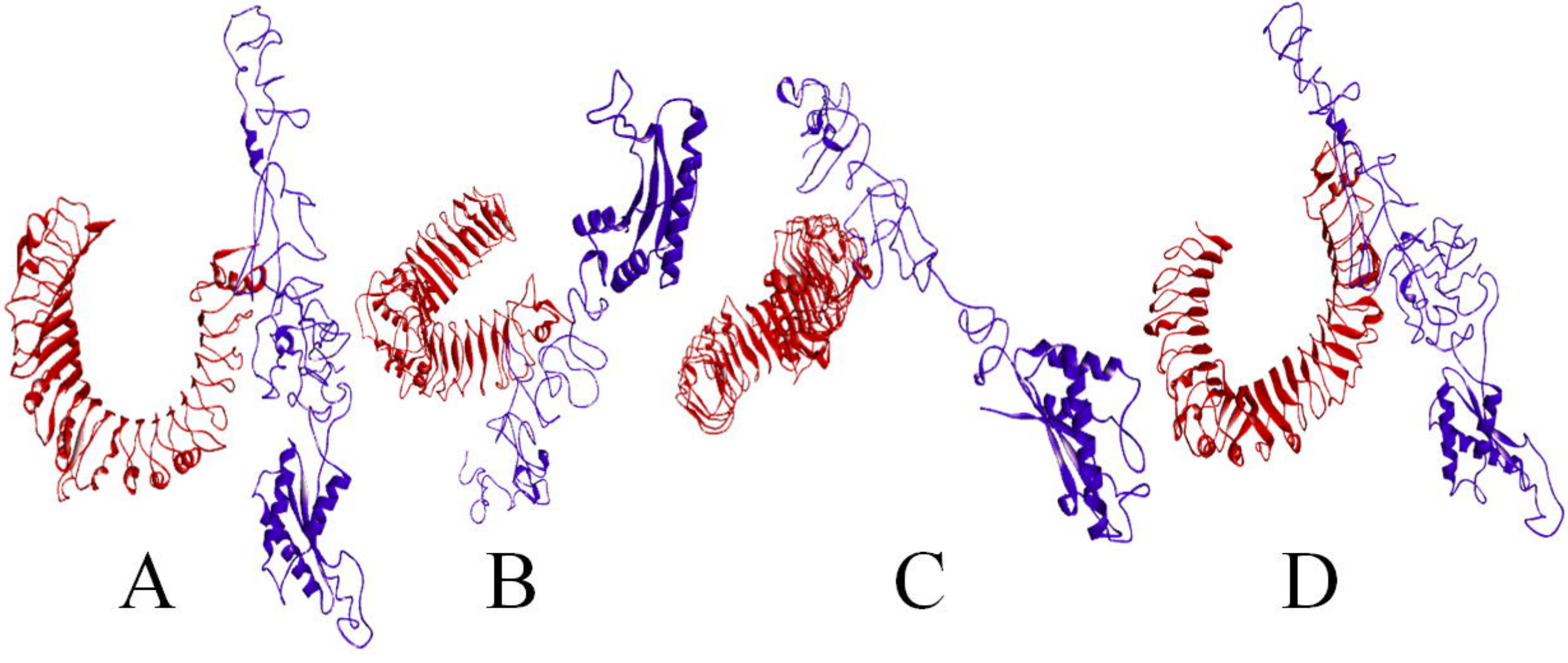

Threshold
Score
Position


Threshold
Score
Position

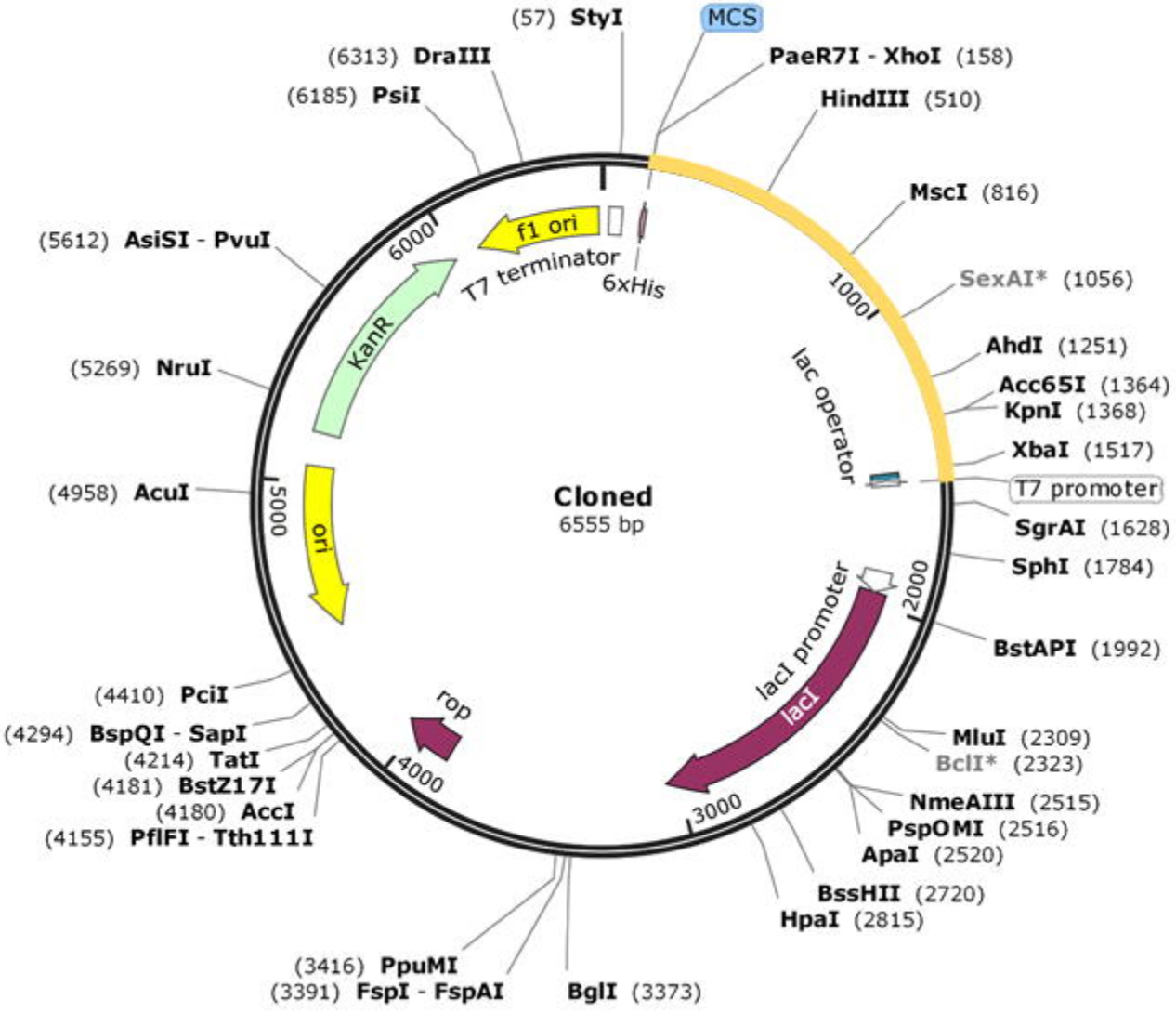

Adjuvant
EAAAK linker
BCR1
GPGPG linker
BCR2
AAY linker
TCR1
TCR2
TCR4
HHHHHH

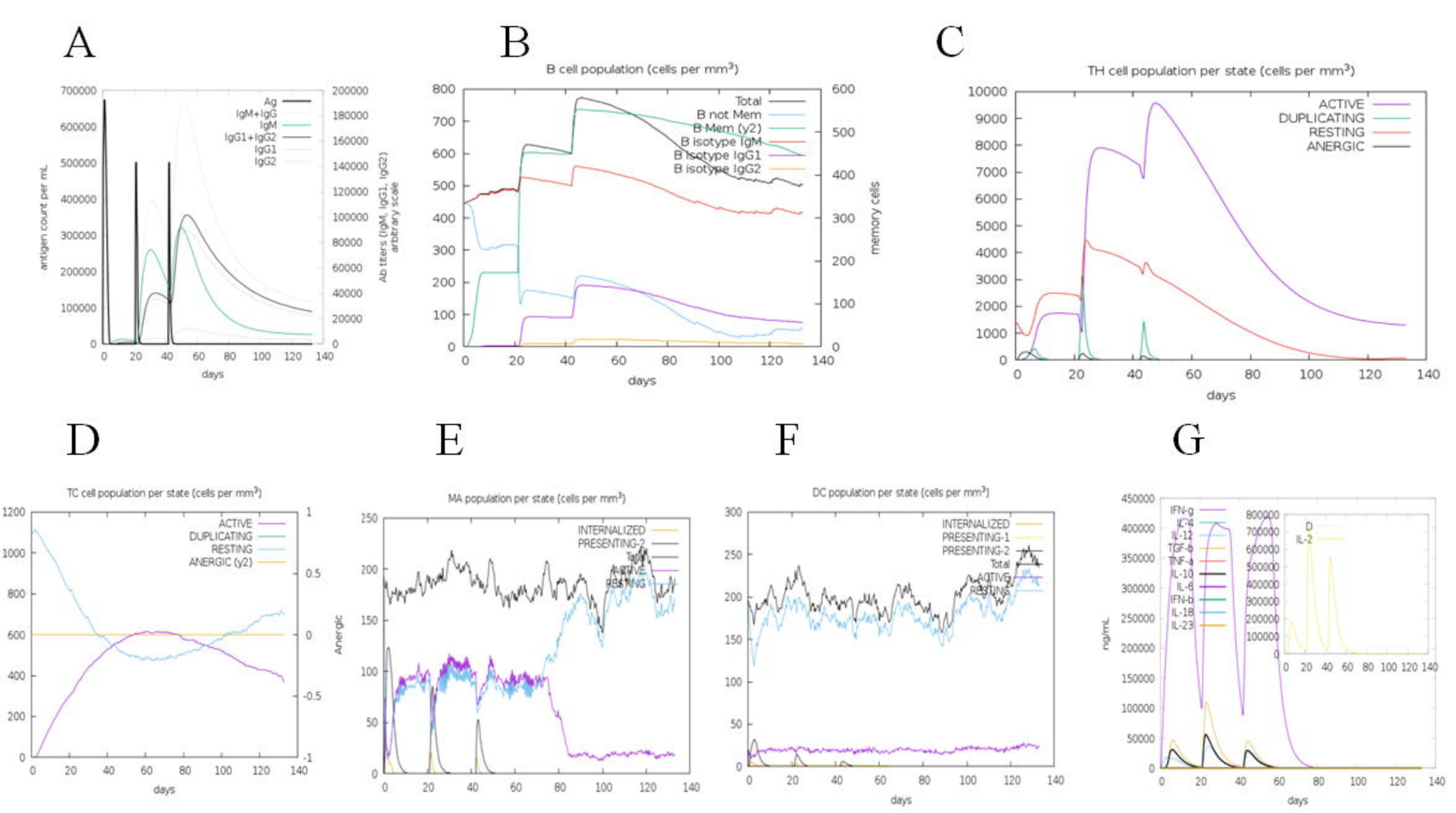

Conf
Cart
Pred
AA
CHHHHHHHHHHHHHHHHHHCCCCCCCCCEEEEEEEEEEEECCCCCCCCCCCC
MKKTAI ALTI AALAAASVAQAAPQPNTFYVGAKAGWASFHDGLNQI EKI E
10 20 30 40 50
CCCCCCCCCCCEEEEEEEEEEEECCCCEEEEEEEEEECCCEEEEECCCCCCCCC
GPI SFGFKRNSVTYGVFGGYQI TDNFAVELGYDDFGRAKVRATDPKTKET
60 70 80 90 100
CCCEEEEEECEEEEEEEEECCCCCCCCCEEEEEEEEEEEEEECCCCCCCCCEE
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