



Research article

Designing spike protein (S-Protein) based multi-epitope peptide vaccine against SARS COVID-19 by immunoinformatics

Hitesh Singh^a, Renu Jakhar^b, Neelam Sehrawat^{b,*}^a Department of Genetics, Maharshi Dayanand University, Rohtak, Haryana India^b Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana India

ARTICLE INFO

Keywords:

Bioinformatics
 Immune response
 Vaccination
 Viral disease
 Proteins
 Peptides
 Molecular docking
 SARS-COVID 19
 Spike protein
 Multi-epitope
 Peptide
 Vaccine
 Immunoinformatics

ABSTRACT

The outbreak of COVID-19 was originated from China, responsible for Several Acute Respiratory Syndrome (SARS). Scientists are forced to develop vaccine and effective drugs to control COVID-19 infection. To develop effective vaccine for SARS – COVID 19, immunoinformatics and computational approaches could help to design successful vaccine against this biggest danger for humanity. Here we used various *in-silico* approaches to design vaccine against COVID-19. To develop vaccine, we target S-protein, expressed on the virus surface plays important role in COVID-19 infection. We identified 12 B-cell, 9 T-helper and 20 Cytotoxic T-cell epitope based on criteria of selection. The predicted epitopes were linked simultaneously with GPGPG & AAY linkers. The β -defensin was used as adjuvant, linked with selected epitope by using EAAAK linker. For vaccine construct justification we analysed its immunogenicity, allergenicity and physicochemical properties. Our study revealed that vaccine was non-toxic, immunogenic and antigenic in nature and covers 98.6% of world population, important for vaccine effectively. *In-silico* cloning was used to analyse its expression in vector. Molecular docking was performed to study the interaction of construct with TLR (TLR3, TLR4, and TLR9) molecules. The immune simulation was conducted and confirmed that our vaccine constructs can induce both acquired and humoral immunity effectively against COVID-19 at very low concentration, but along with bioinformatics study we need to conduct experiment in laboratory to validate its safety and effectiveness.

1. Introduction

In December 2019, an unspecified viral infection was recognized in seafood market of Wuhan city, China (Lu et al., 2020a,b) named as (2019-nCoV). On 30 January 2020 Chinese outbreak, WHO (World Health Organisation) declared it as public health emergency for international concern because of its high transmission rate (WHO, 2020a,b,c). On 16 Aug 2020 the total numbers of cases for COVID-19 are 21,026, 758 and 755,786 deaths globally (WHO, 2020a,b,c). The high mortality and transmission of COVID-19 infection causes huge burden on health organisation and economy of the countries (WHO, 2020a,b,c; Kock et al., 2020) and condition remains hypercritical throughout the world. The drugs used for the treatment of COVID-19 infection are Remdesivir, Chloroquine, Lopinavir, Ritonavir and its combination, but it becomes a question which drug works effectively against it (Agostini et al., 2018; Aguiar et al., 2018; Cvetkovic and Goa, 2003). The COVID-19 infection and its high rate of transmission challenges scientific research and industries to develop effective vaccine and drugs, but since

now there is no effective vaccine or drug to fight against it. Several medical control measures have been taken to control Corona infection, pre-diagnosis, isolation, effective treatment. For individual WHO advises for hygiene and to avoid crowded places. All these control measures are just to control transmission of COVID-19 infection, not permanent solution.

SARS-Cov 2 is a member of *beta corona virus*, causing pneumonia. COVID-19 is an enveloped virus with single stranded RNA, belonging to the *corona viridae* family which can cause infection in mammals, birds and humans (Tortorici et al., 2019; Lu et al., 2020a,b). The whole genome of SARS – CoV 2 was sequenced (Wu et al., 2020), approximately 29.9 kb. The availability of the genome had opened the opportunity to develop vaccine against this devastating disease. The genome of the SARS – CoV 2 encoded for total (6–11) open reading frame (Cui et al., 2019) (orf1ab, S protein, ORF3a, envelope protein, membrane glycoprotein, ORF6, ORF7a, ORF8 protein, nucleocapsid phosphoprotein, ORF10). From all these proteins we target the S-protein which plays an important role in virus infection in humans. It is outer

* Corresponding author.

E-mail address: neelamsehrawat@gmail.com (N. Sehrawat).

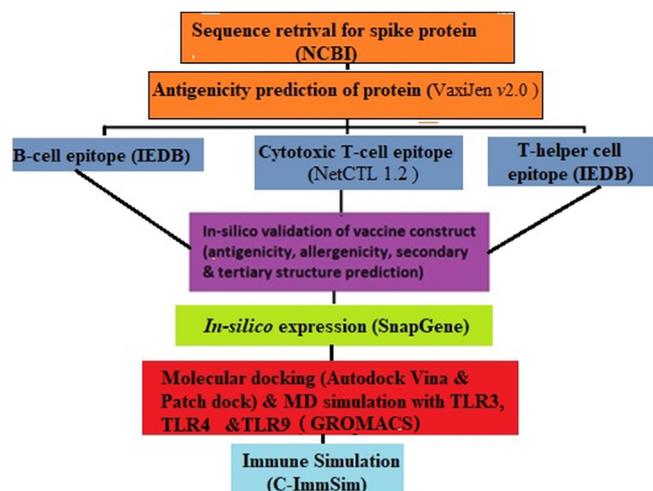


Figure 1. Graphical presentation of scientific research. The methodology for the vaccine formulation against SERS – CoV 2 was represented in this work flow in sequential manner, First two step were for target protein sequence analysis and its antigenicity. Third step for the epitope selection, remaining was for the vaccine construct validation, its interaction with TLR's molecules and last one is immune profiling.

membrane spike glycoprotein which undergoes its glycosylation (Xiong et al., 2018). S protein act as primary interacting protein with host target e. g ACE2, CD26, and other cell receptors) all these play important role in cell adhesion and virulence (Song et al., 2018; Millet et al., 2012). After adhesion the genomic RNA released into cytoplasm and virus enter into the host cell, inside the host cell

genomic RNA translated into two polypeptide and structural protein and start replication (Bergmann et al., 2006).

The spike protein composed of two domains, S1 is receptor binding domain (RBD) supposed that SARS – CoV 2 used angiotensin-converting enzyme 2 (ACE2) receptor to infect human host and another S2 domain responsible for the fusion of viral membrane and host cell membrane. Along with these ACE2 pathway SARS – CoV 2 may use some other pathway for infection because ACE2 expressed in lungs monocytes and macrophages (Bonavia et al., 2003; Li et al., 2003; Yan et al., 2020; Sun et al., 2020). The importance of the S protein during infection and virulence make it an ideal target to develop vaccine for SARS – CoV 2. An effective potential vaccine against COVID-19 has not yet been designed, so here we are trying to designed epitope based vaccine, to activate both innate and acquired immunity. Epitopes are segment of protein which is antigenic in nature and activate immunity against pathogen (Sutton and Boag, 2018). Immediately after the pathogen entry in human host APCs (antigen presenting cell) activates Cytotoxic T-cell to kill infected cell (Khan et al., 2018). Spike protein is antigenic in nature; hence we were detected large number of epitope for both B-cell and T-cell epitope.

Recently due the availability of advance immunoinformatics tool attracts the researchers to develop stable and effective vaccine against pathogenic diseases and it also reduces immunological experimental burden on model organism (Sarkhar et al., 2015). These immunoinformatics tool provide reliable, accurate and rapid pathway to designed potential multiepitope vaccine against diseases e.g enterotoxigenic *Bacteroides fragilis* (Majid and Andle., 2019), *Helicobacter Pylori* (Khan et al., 2019), *Onchocerca volvulus* (Shey et al., 2019) and Cancer (Nezafat et al., 2015), reliable than whole protein and attenuated pathogen which may causes hypersensitivity and other immunological response such local redness, swelling and pain or raise in body temperature after immunization. Based on previous immunoinformatic based multiepitope designed vaccine for other disease we applied the same

Table 1. B-cell epitope for vaccine construct showing its peptide sequence, starting position in S protein and its PI score.

Peptide sequence	Sequence Position	Ellipro score (PI)
IHVSGTNGT	67	0.847
VYFASTEK	89	0.847
TTLDSKTQ	108	0.847
VYYHKNN	143	0.847
MDLEGKQ	177	0.847
SYLTPGDSS	247	0.82
YAWNRKRI	351	0.753
NNLDSKVG	439	0.79
RLFRKSNL	523	0.553
VITPGTNTS	599	0.553
RVYSTGS	634	0.517
QILPDPSKPSKR	804	0.589

Table 2. T–helper cell epitope for MHC II, its percentile rank and immunogenicity score.

MHC II alleles	Sequence Position	Peptide Sequence	Percentile rank	Immunogenicity score
HLA-DRB4*01:01	5	LVLLPLVSSQCVNLT	1.4	0.12
HLA-DRB1*07:01	87	NDGVYFASTESNIIRGWF	0.19	0.2
HLA-DRB5*01:01	187	NLREFVFKNDIGYFKIYS	0.21	0.41
HLA-DRB3*01:01,HLA-DRB1*03:01	205	KHTPINLVRDLPGQFS	0.51	0.126
HLA-DRB4*01:01, HLA-DRB5*01:01	229	PIGINTRFQTLALHRS	1.7	0.4
HLA-DRB5*01:01	342	NATRFASVYAWNRKRIS	0.83	0.27
HLA-DRB3*01:01	395	YADSFVIRGDEVRQIAPGQ	0.49	0.34
HLA-DRB1*07:01	710	IAIPTNFTISVITTEILPVS	0.47	0.35
HLA-DRB3*01:01	1082	HFPEGVFSNGTHWFFVTQR	1.7	0.38

Table 3. Cytotoxic T-cell epitope for MHC I, its immunogenicity score.

Peptide	Sequence Position	HLA supertype	Immunogenicity
VTWFHAIHV	62	A2	0.39
VLPFNDGVY	83	A1, B62	0.18
WTAGAAAYY	258	A1, A26, B58, B62	0.51
LQPRTFLK	270	A3	0.18
VRFPNITNL	327	B27	0.175
FNATRFASV	342	B8	0.15
NYLYRLFRK	450	A24	0.161
YQPYRVVVL	505	A2, A24, B8, B39, B62	0.14
VLSFELLHA	512	A2	0.16
QLTPTWRVY	628	A1, B62	0.31
TPTWRVYST	630	B7	0.224
YECDPIGA	660	B44	0.26
KRSFIEDLL	814	B27	0.30
RSFIEDLLF	815	A1, B58, B62	0.27
QKFNGTLVL	853	B39, B62	0.110
ITSGWTFGA	882	A2	0.35
WTFGAGAAL	886	A26, B62	0.20
AALQIPFAM	892	B7, B58	0.120
FVSNQTHWFV	1095	B58, A26, A1	0.166
YEQYIKWPWTIW	1207	B62, B44, A24, B58	0.40

approaches for SARS-CoV2. Here, S protein antigenic which is antigenic in nature was used as target to detect both B-cell and T-cell epitope to design a potential vaccine construct was formed. *In-silico* cloning was used to study its expression in expression vector; Molecular docking was performed to validate its efficiency to activate immune response and its stability. With the use of immunoinformatics tool validation we required, its affirmation experimentally in animal model and humans.

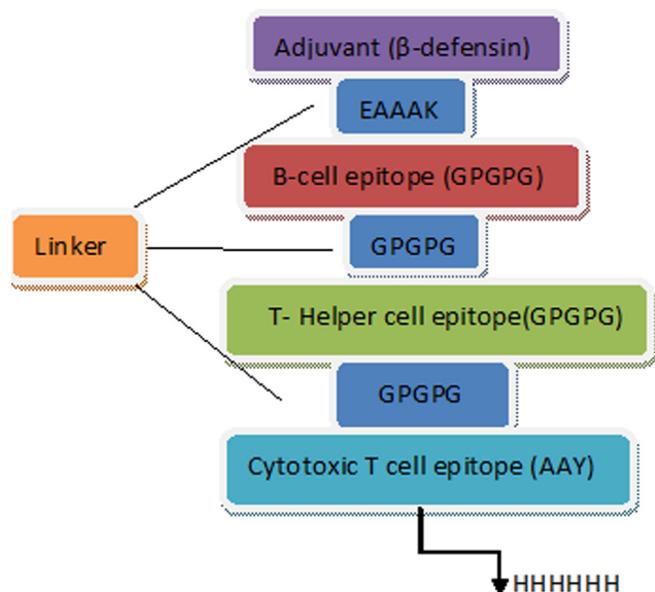


Figure 2. Schematic representations of final vaccine construct. The final vaccine construct contain 678 amino acids, first 64 amino acid are of adjuvant (β -defensin), remaining portion for selected epitope and different linkers. Here the B-cell epitopes are linked together with GPGPG linker and it is useful to connect B-cell and T-helper cell epitope. T cell epitopes are also linked together by GPGPG linker and the same was used to link T-helper and Cytotoxic T-cell epitope. AAY linker was used to link Cytotoxic T-cell epitope and His tag was used at the N-terminal of the vaccine.

2. Methodology

2.1. Sequence retrieval of S-protein and its antigenicity

The sequence of S-protein was retrieved from NCBI. Here, the selection of protein as target for vaccine development is based on its antigenicity and its importance in viral infection. The VaxiJen v2.0 server (Irimi et al., 2007) was used to analyse its antigenicity, important factor to activate immunity against SARS – CoV 2. The graphical work flow for research was shown in Figure 1.

2.2. Conformational and linear B-cell epitope prediction

B-cell lymphocytes plays very significant role in immunity response and produces long lasting immune response against the pathogen. To predict the accurate B-cell linear epitope we used BCPred and IEDB tool (Chou and Fasman, 1978; Parker et al., 1986; Emini et al., 1985; Kolaskar and Tongaonkar, 1990). BCPred (El-Manzalawy et al., 2008; Saha and Raghava, 2004) tool was based on SVM (Support Vector Machine) algorithm to predict B-cell linear epitope. The linear epitopes are selected based on its surface accessibility, flexibility, beta-turn and antigenicity.

To predict Conformational B-cell epitope, Ellipro online server (Ponomarenko et al., 2008) was used. Ellipro selected the epitope based on Residue clustering algorithm along with Thorton's techniques. The standard used for the selection of conformational epitope was PI (Protrusion index: 0.5) score.

2.3. Cytotoxic T-cell epitope (CTL) prediction

The NetCTL 1.2 (Larsen et al., 2007) online sever was used to predict MHC I binding epitope. The prediction of epitope based on C-terminal cleavage and its TAP transport efficiency. These epitopes were selected based on threshold value (0.75). The immunogenicity of CTL epitope were also analysed using IEDB MHC I immunogenicity prediction tool (Calis et al., 2013).

2.4. T-helper cell epitope prediction

IEDB MHCII tool (Wang et al., 2008) was used to predict 15-mer of peptide from S-protein. The epitope were predicted for seven HLA

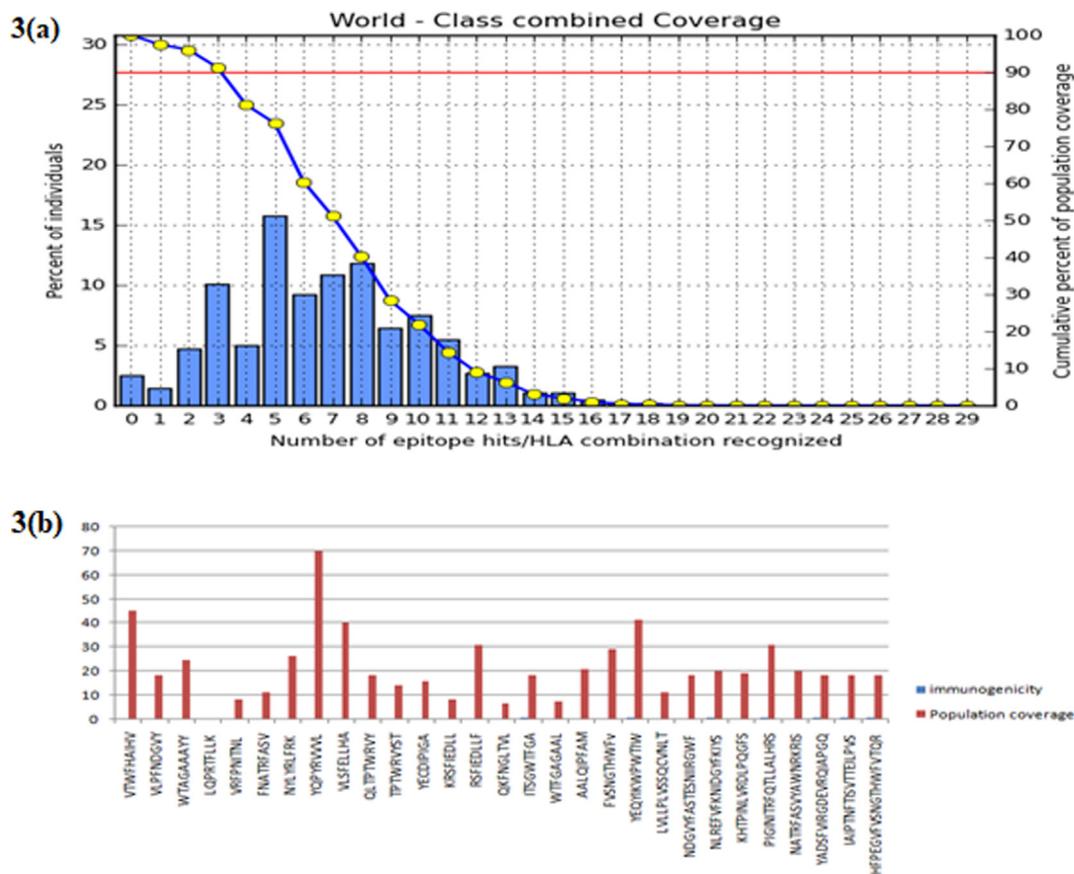


Figure 3. The world population coverage covered with selected MHC I and MHC II epitopes shown in Figure 3(a). In Figure 3(b) showing the immunogenicity and world population coverage covered with the individually selected epitope.

reference set molecules (DRB3 * 01:01, DRB5*01:01, DRB1 * 03:01, DRB1 * 15:01, DRB3 * 02:02, DRB4 * 01:01, -DRB1 * 07:01.). The peptides were predicted based on its percentile rank ≤ 2 and SMM and ANN values ≤ 100 nM. The percentile rank, lower values define the strong binding affinity for MHC II molecules.

2.5. Calculation of population coverage covered for selected epitope

For the vaccine designing it is very important that the vaccine construct covered maximum world population. To calculate the population coverage covered with selected MHC I and MHC II epitopes, we used IEDB population coverage tool (Trott and Olson, 2010).

2.6. Strategic views for development of vaccine construct

B-cell epitope were selected based on its criteria of selection and its allergenicity. The selection of T cell epitope (Cytotoxic T-cell epitope & T-helper based on its binding affinity for HLA molecules and immunogenicity. To construct multiple epitope vaccine, B-cell and T-helper cell epitopes were linked by GPGPG linker and CTL epitopes were linked with AAY linker. To enhance immune response we used β -defensin (Chen et al., 2007) as an adjuvant at C-terminal and linked with EAAAK linker. Polyhistidine-tag was adjoining at C-terminal of the vaccine construct.

2.7. Allergenicity and antigenicity profiling for vaccine construct

To conform whether the vaccine are not eliciting allergy, we use Allertop v 2.0 (Dimitrov et al., 2013), worked on Auto cross covariance of protein sequence.

The antigenicity of the vaccine is important factor to activate immune response. Here we used Vaxijen v 2.0, to predict the antigenicity of vaccine construct. For antigenicity prediction are based on pathogen specific and physiochemical properties of the proteins sequence.

2.8. Analysis of physiochemical properties of vaccine

To understand the vaccine response and its stability, the physiochemical properties of the vaccine construct were analysed with the use of online tool ProtParam (Gasteiger et al., 2005). Some of the protein properties important for its antigenic nature, so we were focused to study some physiochemical properties e.g the amino acid composition, theoretical pI, GRAVY index, its instability index etc.

2.9. Secondary structure predictions for vaccine construct

To predict the secondary structure of the vaccine used SOPMA (Geourjon and Deleage, 1995) online software. The prediction of secondary structure based on the Homology Protein modelling.

The secondary structure analysis was used to predict its solvent accessibility, trans-membrane helix, globular region and β -turn region which are the important factor for protein stability and for efficiency of vaccine.

2.10. 3-D structure prediction

The 3-D structure of the predicted vaccine was modelled by using the online server RaptorX (Kallberg et al., 2012). The modelled structure was refined with the use of Galaxy refine tool online server (Ko et al., 2012).

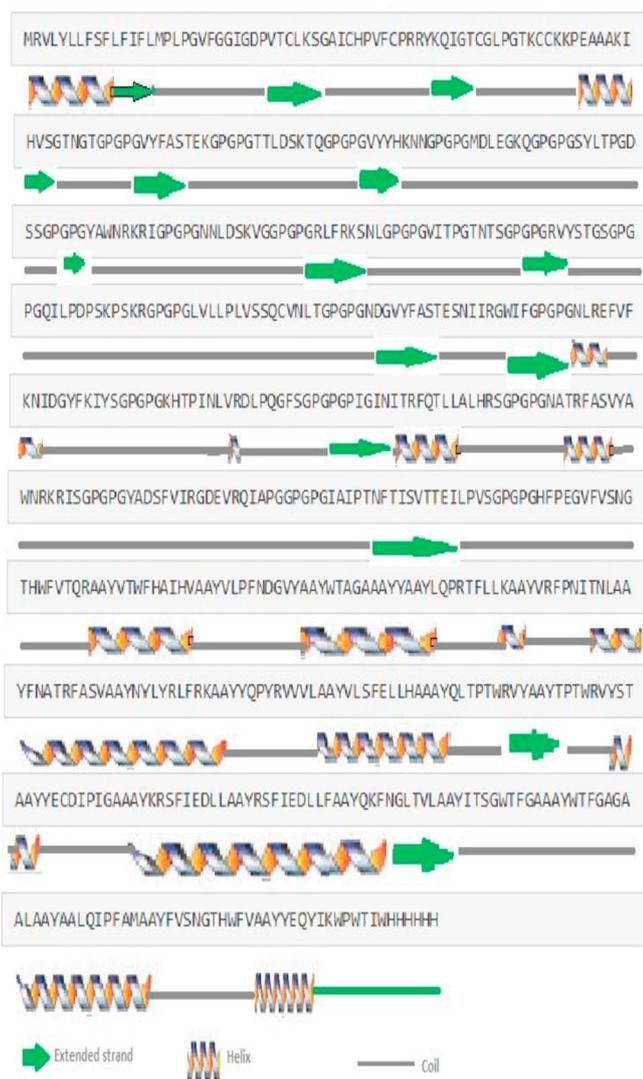


Figure 4. The secondary structure prediction analysis with the SOPMA online server.

2.11. 3-D structure refinement and its validation

For Validation of refine 3-D structure, we used RAMPAGE (Lovell et al., 2003) and PROSA (Wiederstein and Sippl, 2007) web online tool. RAMPAGE is based on the Ramachandran plot (Hoof et al., 1997) to analyse the amino acid in favoured and unfavoured region.

2.12. In silico validation of vaccine construct

2.12.1. Codon optimisation of vaccine

To reverse transcription and codon optimisation, we used Jcat online tool (Grote et al., 2005) to calculate GC content and CAI score for maximum expression in vector. For maximum expression of query sequence GC content should be more than 50% and CAI score near to 1.

2.12.2. In-silico expression of vaccine

For integration of query sequence in E. coli pET-28a(+), SnapGene software was used. The software was used to verify the maximum expression of vaccine in expression vector at specific site.

2.13. Molecular docking of vaccine construct with TLR4, TLR3 & TLR9 receptor

To analyse vaccine construct interaction with TLR's (TLR2, TLR3, TLR4, TLR5, TLR7, TLR8 and TLR9) molecules, we performed docking with multiple TLR'S molecules. The analysis of docking result show that the vaccine construct shows high binding affinity with the TLR3, TLR4 and TLR 9 which are toll like receptor to activate immune response. For molecular docking Autodock vina and Patchdock was used (Forli et al., 2016), and result was visualized in Pymol (<http://www.pymol.Org>) and Discovery studio (Pal et al., 2019) was used. CASTp web server was used to assess the binding pocket of vaccine construct with the receptors. GROMACS was used for MD simulation to check the stability and flexibility of the protein.

2.14. Immune response profiling of vaccine construct

C-immSim (<http://150.146.2.1/C-IMMSIM/index.php>) immune stimulator was used for the analysis of immune response developed by the vaccine construct. It works based on position specific scoring algorithm (PSSM) for the prediction of immunogenic epitope and its interaction with immune system. For the immune response analysis simulation was done with the default parameter and single injection was used.

3. Results

3.1. SARS – CoV 2 spike protein sequence retrieval

The sequence for Spike protein (S-protein) was retrieved from NCBI, to predict of B-cell and T- cell epitope for the construction of multi epitope vaccine to fight against the COVID-19. The selection of S- protein for epitope prediction was based on its importance in viral infection and virulence.

3.2. Antigenicity of S-protein in SARS – CoV 2

The antigenicity of the protein was analysed with Vaxijen v 2.0 online immunoinformatics tool, results its antigenicity score 0.48, indicates it is a probable antigen to activate immunity against SARS – CoV 2.

3.3. Prediction of B-cell epitope, Cytotoxic T-cell & T helper cell epitope for vaccine

BCpred and IEDB online tools were used for the selection of highly efficient B-cell linear epitope. The selection was based on the antigenicity, its position in Beta turn, flexibility and solubility. Ellipro was used for the selection of conformational epitope. Based on the selection criteria there 12 B- cell epitope (Table 1), to produce antibodies against SARS – CoV-2.

There are 9 T-helper (Table 2) and 20 Cytotoxic T-cell (Table 3) were selected for vaccine are non allergenic and immunogenic in nature.

3.4. Formulation of multi-epitope for vaccine construct against SARS – CoV 2

Twelve B-cell epitope and nine T-helper cell epitopes were linked together with GPGPG linker. T-helper and Cytotoxic T-cell epitope are linked with GPGPG and twenty Cytotoxic T-cell epitope are joined together with AAY linker. β - defensin was used as an adjuvant at C-terminal for the protection of vaccine from degradation and enhance immune response. The linker EAAAK was used to connect the adjuvant with

MFVF L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R
 S S V L H S T Q D L F L P F F S N V T W F H A I H V S G T N G T K R F D N P V L P F N D
 G V Y F A S T E K S N I R G W I F G T T L D S K T Q S L L I V N N A T N V V I K V C E F
 Q F C N D P F L G V Y Y H K N N K S W M E S E F R V Y S S A N N C T F E Y V S Q P F L
 M D L E G K Q G N F K N L R E F V F K N I D G Y F K I Y S K H T P I N L V R D L P Q G F
 S A L E P L V D L P I G I N I T R F Q T L L A L H R S Y L T P G D S S S G W T A G A A A Y
 Y V G Y L Q P R T F L L K Y N E N G T I T D A V D C A L D P L S E T K C T L K S F T V E
 K G I Y Q T S N F R V Q P T E S I V R F P N I T N L C P F G E V F N A T R F A S V Y A W N
 R K R I S N C V A D Y S V L Y N S A S F S T F K C Y G V S P T K L N D L C F T N V Y A D
 S F V I R G D E V R Q I A P G Q T G K I A D Y N Y K L P D D F T G C V I A W N S N N L D
 S K V G G N Y N Y L Y R L F R K S N L K P F E R D I S T E I Y Q A G S T P C N G V E G F
 N C Y F P L Q S Y G F Q P T N G V G Y Q P Y R V V V L S F E L L H A P A T V C G P K K
 S T N L V K N K C V N F N F N G L T G T G V L T E S N K K F L P F Q Q F G R D I A D T T
 D A V R D P Q T L E I L D I T P C S F G G V S V I T P G T N T S N Q V A V L Y Q D V N C T
 E V P V A I H A D Q L T P T W R V Y S T G S N V F Q T R A G C L I G A E H V N N S Y E
 C D I P I G A G I C A S Y Q T Q T N S P R R A R S V A S Q S I I A Y T M S L G A E N S V A
 Y S N N S I A I P T N F T I S V T T E I L P V S M T K T S V D C T M Y I C G D S T E C S N L
 L L Q Y G S F C T Q L N R A L T G I A V E Q D K N T Q E V F A Q V K Q I Y K T P P I K D
 F G G F N F S Q I L P D P S K P S K R S F I E D L L F N K V T L A D A G F I K Q Y G D C L
 G D I A A R D L I C A Q K F N G L T V L P P L L T D E M I A Q Y T S A L L A G T I T S G
 W T F G A G A A L Q I P F A M Q M A Y R F N G I G V T Q N V L Y E N Q K L I A N Q F N
 S A I G K I Q D S L S S T A S A L G K L Q D V V N Q N A Q A L N T L V K Q L S S N F G A
 I S S V L N D I L S R L D K V E A E V Q I D R L I T G R L Q S L Q T Y V T Q Q L I R A A E I
 R A S A N L A A T K M S E C V L G Q S K R V D F C G K G Y H L M S F P Q S A P H G V
 V F L H V T Y V P A Q E K N F T T A P A I C H D G K A H F P R E G V F V S N G T H W F
 V T Q R N F Y E P Q I I T D N T F V S G N C D V V I G I V N N T V Y D P L Q P E L D S F
 K E E L D K Y F K N H T S P D V D L G D I S G I N A S V V N I Q K E I D R L N E V A K N
 L N E S L I D L Q E L G K Y E Q Y I K W P W Y I W L G F I A G L I A I V M V T I M L C C
 M T S C C S C L K G C C S C G S C C K F D E D D S E P V L K G V K L H Y T

Figure 5. Here the position of the predicted epitope was represented in target proteins (S-protein). Different colours were used to show epitope; B-cell (green), T-helper (blue) and Cytotoxic T-cell (Red) epitope, the pink colour was used for the overlapping region between to constitutive epitope.

B-cell epitope. *His-tag* was used at N-terminal of the vaccine construct (Figure 2).

3.5. Population coverage for vaccine construct

The world population covered with the vaccine construct was 98.6% (Figure 3a, b), calculating with the IEDB population coverage online tool. The population coverage covered with the vaccine construct conformed that the designed vaccine was effective for most of the population throughout the world.

3.6. Analysis of physio-chemical properties

Several physio-chemical properties were analysed for the vaccine construct, its amino acid construct, molecular weight, pI and its instability index. The antigenicity and allergenicity of the vaccine construct conformed to Vaxijen v2.0 server and Allertop online server, the study confirmed that it the final construct was non-allergenic and antigenic in nature with antigenicity score 0.46. The total length of vaccine construct was 678 amino acid with molecular weight of 72.605 Kda. The isoelectric

point for the vaccine construct was 9.57, indicating it is a basic in nature. The GRAVY index (0.065) and instability index (26.9) indicated that the vaccine construct was stable protein.

3.7. Secondary structure prediction

The secondary structure of the protein was predicted with the use of online server SOPMA. The analysis of secondary structure results that the vaccine composed of Alpha helix (23.5%), Random coil (44%), extended strand (25.2%) and beta-turn (7.2%) and epitope positions in vaccine construct are shown (Figure 4 and 5).

3.8. 3-D structure prediction and structure refinement

RaptorX was used to model the 3-D structure of the vaccine. The PDB templates (6VSB, 1FD3, 2LXO, 6MOJ, 1BNB, 2RUN, 2DD8) were used for modelling its 3-D structure (Figure 6a) using tread modelling method. The structure of the Vaccine construct was refined with Galaxy refine tool.

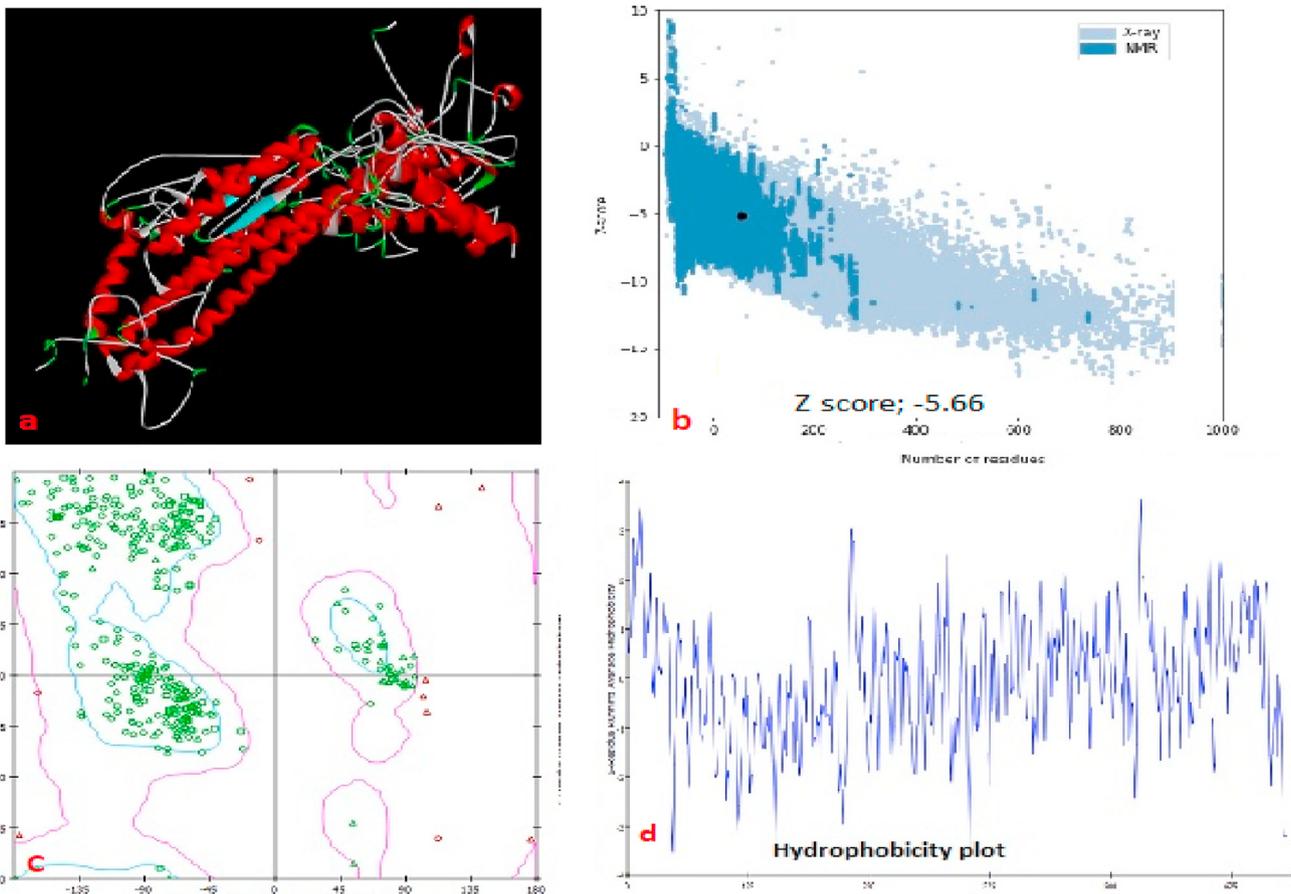


Figure 6. 3-D structure and its validation of vaccine construct (a) 3-D structure of the vaccine construct formed by raptorx (b) Z core plot of refined structure (c) Ramachandran plot (d) Hydrophobicity plot.

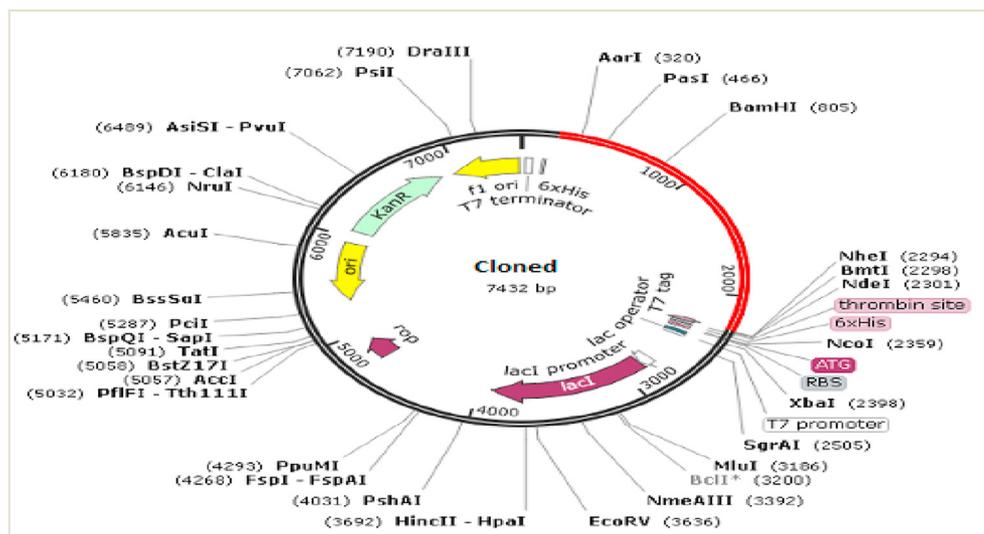


Figure 7. In-silico cloning of vaccine construct in pET 28a (+) expression vector in which red part represent the insert and rest part are vector genome.

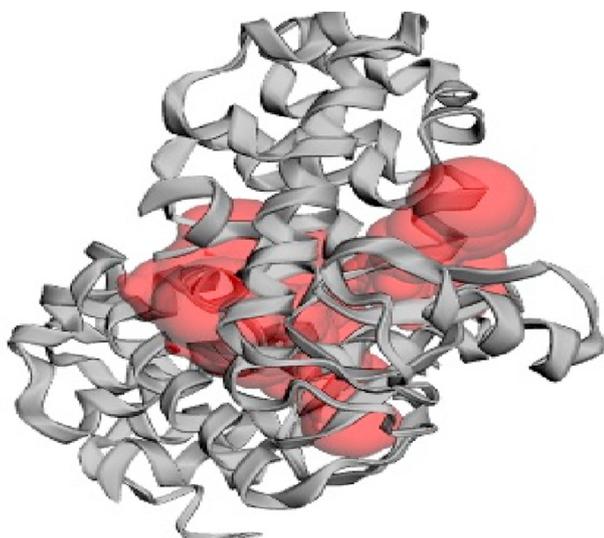


Figure 8. CASTp web server result showing active binding pocket in vaccine construct represented in red colour and grey for surrounding amino acids.

3.9. 3-D structure validation of vaccine construct

The refined structure of the protein was validate with RAMPAGE, it confirmed that 89.8% amino acid are in favourable region, 5.5% are in allowed region and only 4.6% amino acid are in outlier region. PROSA –web (Figure 6b) online server used to analyse the standard of the structure. The Z score value of the refined structure was -5.66 and its Ramachandran plot (Figure 6c) and hydrophobicity plot (Figure 6d) show that most of the amino acid in vaccine are in favourable region.

3.10. Codon optimization & in-silico cloning

In this study Jcat online tool was used to understand the expression level of the vaccine construct in the E-coli. There are about 2130 bp DNA sequence was used to input in vector. Jcat codon optimization results, the Codon optimization index (CI) was 1.0 and its GC contain was 58.08% which indicating that the vaccine construct was expressed at best level of expression in *E. coli* K12. The input sequence was cloned in pET-28a expression vector as shown in figure (Figure 7).

3.11. Molecular docking of vaccine construct with TLR receptors

Here, we analyzed the docking of TLR3, TLR4 and TLR9 receptor which are interact with vaccine construct. The CASTp web server shows binding pocket of the vaccine construct for the receptor, the receptor binding pocket was about 1461.7 Å formed with the amino acid at the position; 8 Arg, 9ILE, 10ASP, 11GLU, 12GLY, 14 ARG, 18,TRY, 19LYS, 20ASP, 21 THR, 22GLU, 24,TYR, 26THR, 30GLY, 32LEU,104PHE, 106MET, 107GLY, 137ARG, 138TRP, 143PRO, 144ASN, 145ARG, 141GLN, 142THR (Figure 8). Autodock vina was used for docking and its results shows that the vaccine construct was

effectively bind with the receptors, showing strong binding affinity for TLR3 (-11.2 kcal/mol), TLR4 (-9.55 kcal/mol) and TLR9 (13.7 kcal/mol). To validate autodock vina result further docking was performed with PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/>) and results for all three TLR's was shown in the Table 4. Patchdock docking Score value and atomic contact energy conformed that our vaccine construct bind efficiently to TLR3, TLR4 & TLR9. Pymol and Discovery studio was used for the receptor and vaccine construct visualization Shown (Figure 9).

MD simulation conform the thermo stability of the vaccine construct from RMSD (Root Mean Square Deviation) and RMSF (Root Mean Square Fluctuation) plot. Throughout processing, during simulation the value of RMSD value changed because of the backbone atomic movement. The high variation in RMSF peak for the residues indicating that vaccine construct is highly flexible (Figure 10) and compactness of the plot show its thermo stability.

3.12. Immune simulation

Immune simulation using C-immSimm with antigen single dose was done for HLA alleles (HLA-A*0101, HLA-A*0201, HLA-B*0702,HLA-B*3901,HLA-DRB1*0101, andHLA-DRB1*0401). The simulation result was shown graphically for total number of lymphocytes, antibodies and cytokine concentration. The simulation result show that the secondary and tertiary immune response is more prevalent than primary response. The primary response mainly produces by B memory cell and IgM (Figure 11A).

The concentration of T helper cell (Figure 11C), Cytokine and interleukins was considerably increased (Figure 11D). The concentration of T-cytotoxic cell varied after administration of vaccine construct.

4. Discussion

The COVID-19 is a viral disease and very difficult to control this life threatening disease because of its evolving nature and high transmission rate. There are number of drugs are used for the treatment of the SARS-CoV 2, but none of them are very effective to control it. The vaccines are single one effective tool to fight against COVID-19. The classical approaches for vaccine development involving with some problem related with its efficiency and its expression. To overcome from all difficulties the new approaches “epitope based vaccine” are more promising than whole protein used a vaccine. The availability of computational tools and genome/proteome of the organism make it as easy to make subunit based vaccine, completely antigenic in nature and capable to activate immune response. The immunoinformatics plays important role to develop stable, effective and safe vaccine for human consumption.

In this study we target the spike protein (S protein) of SARS-CoV 2, important for virus infection in humans and antigenic in nature. There are various tools were used for the selection of the antigenic region of the protein to develop vaccine. For vaccine development, we focused on both T-cell and B-cell epitope selection from S-protein of COVID-19.

The designed vaccine construct covered most of the world population (~98.6%). To provide long term immunity against COVID- 19,

Table 4. Protein-protein docking with patchdock result to validate vaccine construct interaction with TLR3, TLR4 and TLR9 molecules.

TLR molecules	Score	Atomic contact energy	Area
TLR3	19171	- 240	4003
TLR4	15724	-246	2261
TLR9	19080	-143.5	3434

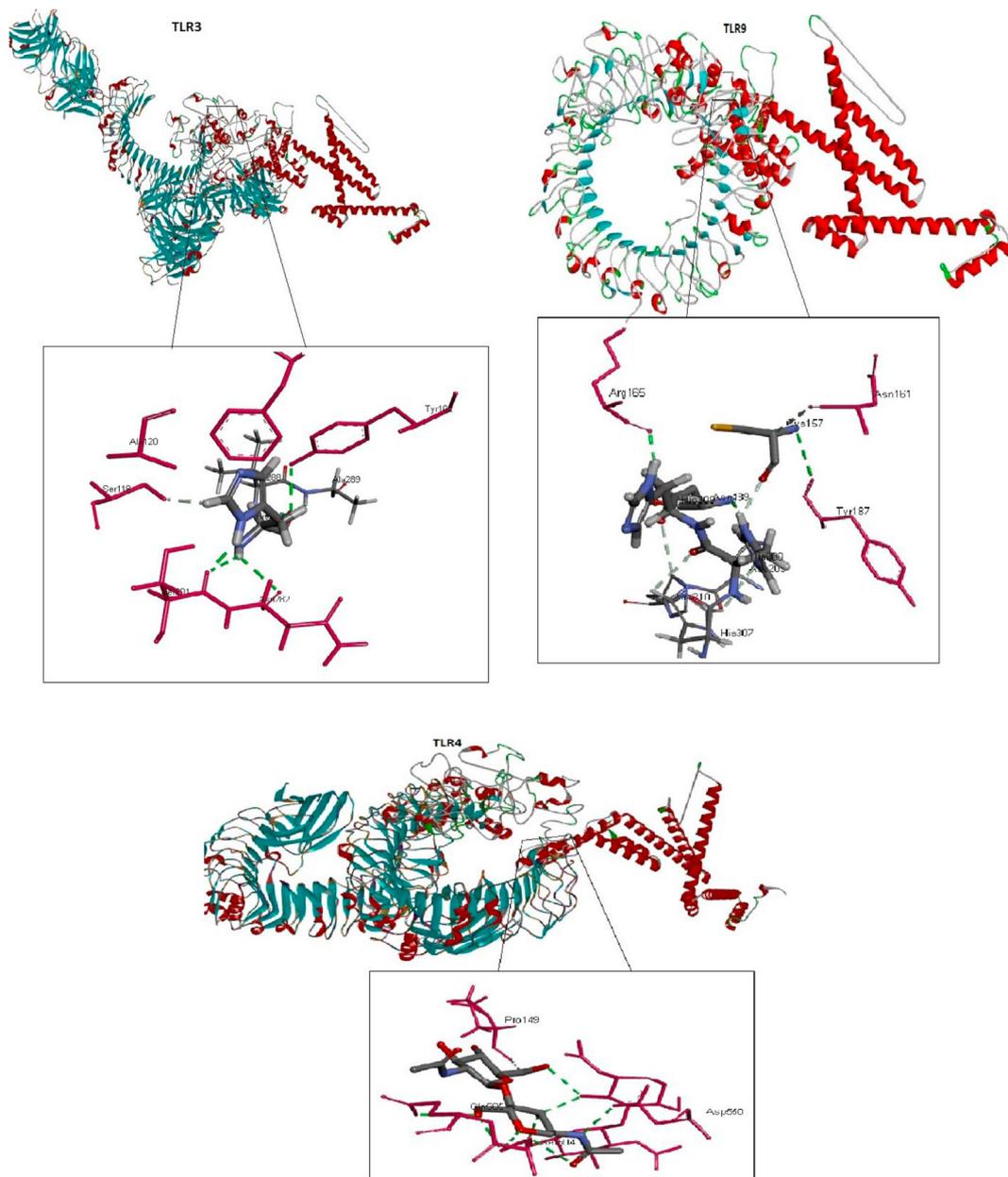


Figure 9. Molecular docking stimulation of vaccine construct with the TLR3, TLR4 and TLR9 receptor was shown. The interaction of receptor and vaccine construct was shown, the green dotted line indicate the hydrogen bonding between vaccine construct (grey) and receptor amino acid residue (pink).

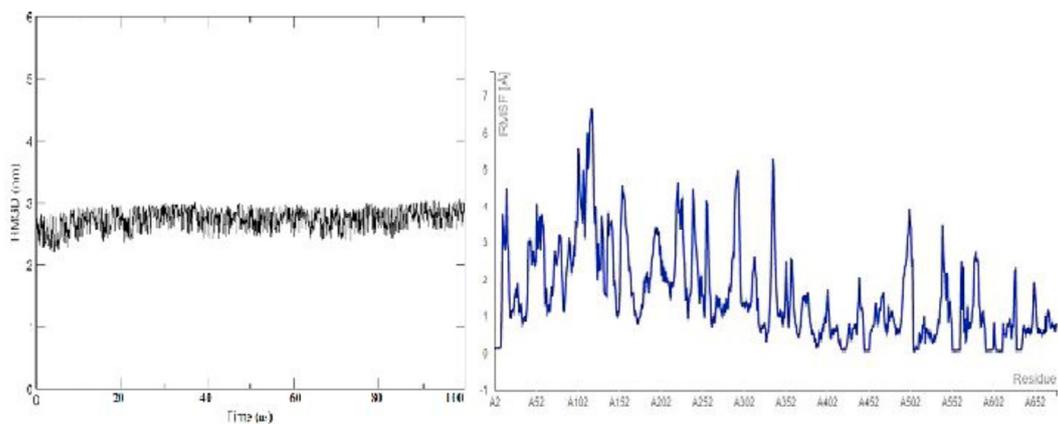


Figure 10. RMSD/RMSF plot value calculated for each protein residue. In figure there is great variation in RMSF (Blue) value indicating the flexibility of the protein. The RMSD plot (Black) conformed stability of the protein.

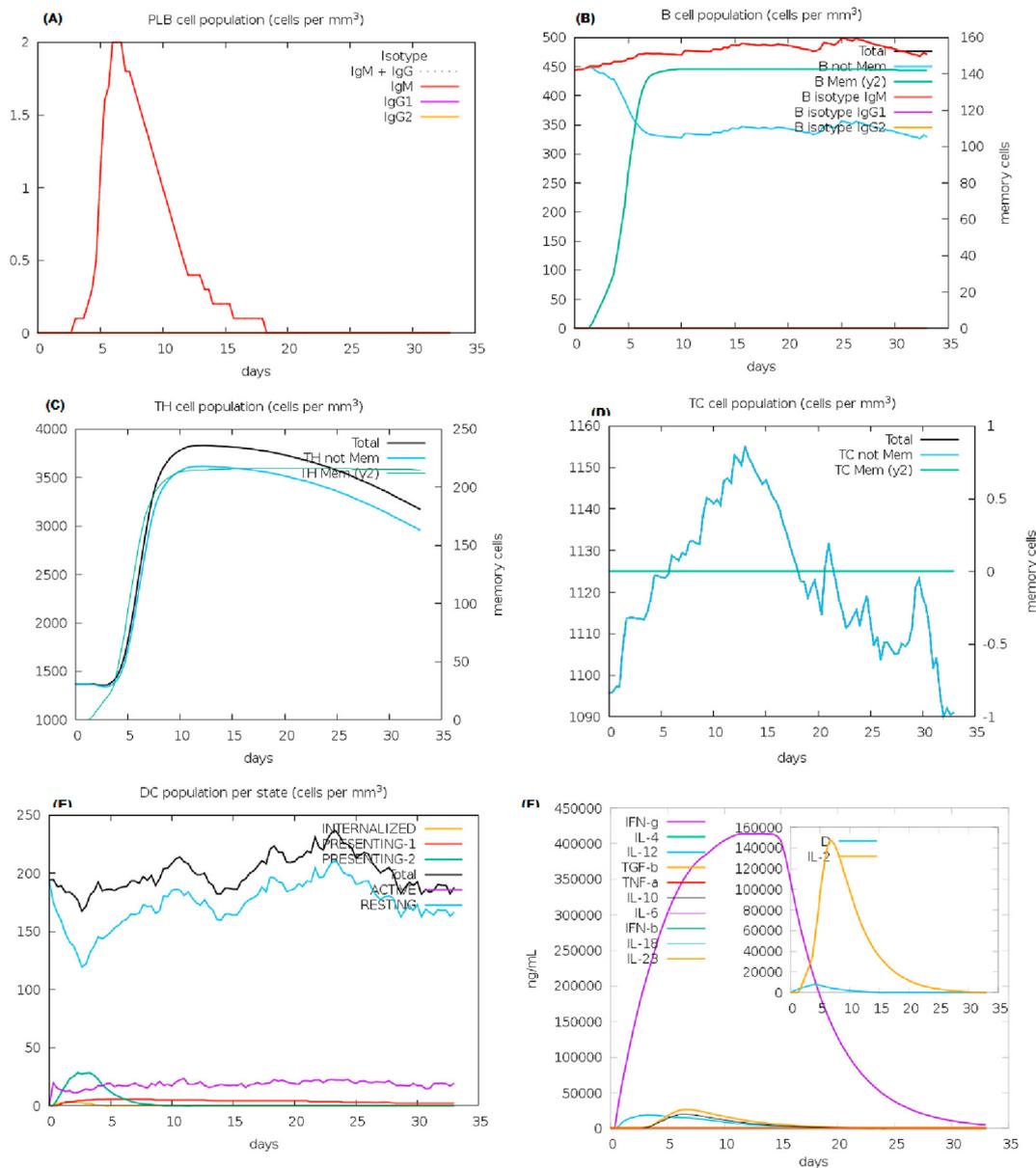


Figure 11. (A) Antigen and immunoglobulins of control (B) antigen and immunoglobulin with vaccine construct showing antibodies subdivided per isotype (C) The graph showing the total and memory T helper cell (D) plot showing the Cytotoxic T-cell count with vaccine construct (E) Dendritic cell total count for vaccine construct (F) Cytokine and interleukin concentration with vaccine construct.

B-cell epitope are selected based on its conformational and physicochemical properties. For the selection of efficient B-cell epitope to generate acquired immunity we target both B-cell linear and conformational epitope. To develop vaccine construct various linker (EAAK, GPGPG, and AAY) are used to adjoin selected epitope and adjuvant was used to enhance the immune response of the vaccine construct. There were 12 B-cell epitope, identified to produce antibodies against SARS – CoV-2. The T-cell epitope selection was based on its high binding affinity for MHC I and MHC II molecules. There are 9 T-helper and 20 Cytotoxic T-cell were selected for vaccine are non allergenic and immunogenic in nature. Twelve B-cell epitope and nine T-helper cell epitopes were linked together with GPGPG linker. T-helper and Cytotoxic T-cell epitope are linked with GPGPG and twenty Cytotoxic T-cell epitope are joined together with AAY linker. β -defensin was used as an adjuvant at C-terminal for the protection of vaccine from degradation and enhance immune response. The linker EAAK was used to connect the adjuvant with B-cell epitope. His-tag was used at N-terminal of the vaccine construct.

To understand the efficiency of the vaccine construct various parameters were analysed to conform its allergenicity and antigenicity. The final construct was non-allergenic and antigenic in nature with antigenicity score 0.46. The total length of vaccine construct was 678 amino acid with molecular weight of 72.605 Kda. The isoelectric point for the vaccine construct was 9.57, indicating it is a basic in nature. The GRAVY index (0.065) and instability index (26.9) indicated that the vaccine construct was stable protein.

The 3-D structure was assessed using RaptorX and refined with Galaxy refine - web server. The RAMPAGE and PROSA web server were used for refined structure validation. Their 89.8% amino acids are in favourable region, 5.5% are in allowed region and only 4.6% amino acids are in outlier region. The Z score value of the refined structure was -5.66 and its Ramachandran plot and hydrophobicity plot shows that most of the amino acid in vaccine are in favourable region.

The expression level of the vaccine construct in *E. coli* K12 analysed with the use of *In-silico* cloning and Jcat online tool revealed its maximum

expression. The molecular docking stimulation of vaccine construct with TLR molecules (TLR3, TLR4, and TLR9) was done to assess its interaction. The vaccine construct was effectively bind with the receptors, showing strong binding affinity for TLR3 (-11.2 kcal/mol), TLR4 (-9.55 kcal/mol) and TLR9 (13.7 kcal/mol). The Immune simulation show that our vaccine construct is able to generate remarkable immune response just after administration at very low concentration of antigen.

The computational and immunoinformatics tool studies conformed that the vaccine construct is highly effective, stable and safe as a target for progressive vaccine against SARS-CoV2 and shows high rate of expression level in *E-coli*. Moreover, accurately the efficiency and validity of the vaccine understand after *in-vitro* experimental studies.

5. Conclusion

Here, our immunoinformatics based vaccine designed to fight against the SARS-CoV2. The potential novel multi-epitope designed vaccine able to generate both humoral and acquired immunity against COVID-19 because of high binding affinity with both HLA molecules. Our vaccine construct can generate remarkable immune response after injection of single dose of antigen. The various tools are used for its structure analysis and validation. The validation of vaccine construct for its safety, stability and its expression level confirmed by the using *in-silico* approaches. The results for the vaccine construct provide the deep view for vaccine designing and provide the base for experimentally verified vaccine to fight against the viral disease.

Declarations

Author contribution statement

Hitesh Singh: Conceived and designed the experiments; Performed the experiments; Wrote the paper.

Renu Jakhar: Analyzed and interpreted the data.

Neelam Sehrawat: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by Radha Krishan Fund, Maharshi Dayanand University, Rohtak. Hitesh Singh was supported by Maharshi Dayanand University (University Research Scholarship).

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

Authors are thankful to Radha Krishan Fund, Maharshi Dayanand University, Rohtak for financial support and Hitesh Singh is thankful for providing University Research Scholarship to MDU, Rohtak.

References

Agostini, M.L., Andres, E.L., Sims, A.C., Graham, R.L., Sheahan, T.P., Lu, X., 2018. Coronavirus susceptibility to the antiviral remdesivir (gs-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *mBio* 9 (2) e00221–18.

Aguiar, A.C., Murce, E., Cortopassi, W.A., Pimentel, A.S., Almeida, M., Barros, D.C.S., 2018. Chloroquine analogs as antimalarial candidates with potent *in vitro* and *in vivo* activity. *Int. J. Parasitol. Drugs Drug Resist.* 8 (3), 459–464.

Bergmann, C.C., Lane, T.E., Stohman, S.A., 2006. Coronavirus infection of the central nervous system: host-virus stand-off. *Nat. Rev. Microbiol.* 4 (2), 121–132.

Bonavia, A., Zelus, B.D., Wentworth, D.E., Talbot, P.J., Holmes, K.V., 2003. Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. *J. Virol.* 77 (4), 2530–2538.

Calis, J.A., Maybeno, M., Greenbaum, J.A., Weiskopf, D., De Silva, A.D., Sette, A., Kesmir, C., Peters, B., 2013. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comp. Biol.* 8 (1), 361.

Chen, X., Niyonsaba, F., Ushio, H., Hara, M., Yokoi, H., Matsumoto, K., Saito, H., Nagaoka, I., Ikeda, S., Okumura, K., Ogawa, H., 2007. Antimicrobial peptides human β -defensin (hBD)- and hBD-4 activate mast cells and increase skin vascular permeability. *Eur. J. Immunol.* 37 (2), 434–444.

Chou, P.Y., Fasman, G.D., 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Area Mol. Biol.* 47, 45–148.

Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17 (3), 181–192.

Cvetkovic, R.S., Goa, K.L., 2003. Lopinavir/ritonavir: a review of its use in the management of HIV infection. *Drugs* 63 (8), 769–802.

Dimitrov, I., Flower, D.R., Doytchinova, I., 2013. April. AllerTOP—a server for *in silico* prediction of allergens. *BMC Bioinform.* 14 (6), S4. BioMed Central.

El-Manzalawy, Y., Dobbs, D., Honavar, V., 2008. Predicting linear B-cell epitopes using string kernels. *J. Mol. Recogn.: Interdiscipl. J.* 21, 243–255.

Emini, E.A., Hughes, J.V., Perlow, D.S., Boger, J., 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* 55, 836–839.

Forli, S., Huey, R., Pique, M.E., Sanner, M.F., Goodsell, D.S., Olson, A.J., 2016. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* 11 (5), 905.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*. Humana press, pp. 571–607.

Geourjon, C., Deleage, G., 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11 (6), 681–684.

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C., Jahn, D., 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33 (suppl_2), W526–W531.

Hooft, R.W., Sander, C., Vriend, G., 1997. Objectively judging the quality of a protein structure from a Ramachandran plot. *Bioinformatics* 13 (4), 425–430.

Irini, A., Doytchinova, Darren, R.F., 2007. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 25, 856–866.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., Xu, J., 2012. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* 7 (8), 1511.

Khan, A., Junaid, M., Kaushik, A.C., Ali, A., Ali, S.S., Mehmood, A., Wei, D.Q., 2018. Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches. *PLoS One* 13 (5).

Khan, M., Khan, S., Ali, A., Akbar, H., Sayaf, A.M., Khan, A., Wei, D.Q., 2019. Immunoinformatics approaches to explore *Helicobacter Pylori* proteome (Virulence Factors) to design B and T cell multi-epitope subunit vaccine. *Sci. Rep.* 9 (1), 1–13.

Ko, J., Park, H., Heo, L., Seok, C., 2012. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* 40, W294–W297.

Kock, R.A., Karesh, W.B., Veas, F., Velavan, T.P., Simons, D., Mboera, L.E., Dar, O., Arruda, L.B., Zumla, A., 2020. 2019-nCoV in context: lessons learned? *The Lancet Planetary Health* 4 (3), e87–e88.

Kolaskar, A.S., Tongaonkar, P.C., 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 276, 172–174.

Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., Nielsen, M., 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinform.* 8 (1), 424.

Li, W., Moore, M.J., Vasilieva, N., 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426 (6965), 450–454.

Lovell, S.C., Davis, I.W., Arendall III, W.B., De Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by C α geometry: ϕ , ψ and β deviation. *Proteins: Structure Funct. Bioinform.* 50 (3), 437–450.

Lu, H., Stratton, C.W., Tang, Y.-W., 2020a. Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle. *J. Med. Virol.* 92 (4), 401–402.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., 2020b. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395 (10224), 565–574.

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

Millet, J.K., Kien, F., Cheung, C.-Y., 2012. Ezrin interacts with the SARS coronavirus spike protein and restrains infection at the entry stage. *PLoS One* 7 (11), e49566.

Nezafat, N., Sadraei, M., Rahbar, M.R., Khoshnoud, M.J., Mohkam, M., Gholami, A., Banihashemi, M., Ghasemi, Y., 2015. Production of a novel multi-epitope peptide vaccine for cancer immunotherapy in TC-1 tumor-bearing mice. *Biologicals* 43 (1), 11–17.

Pal, S., Kumar, V., Kundu, B., Bhattacharya, D., Preethy, N., Reddy, M.P., Talukdar, A., 2019. Ligand-based pharmacophore modeling, virtual screening and molecular docking studies for discovery of potential topoisomerase I inhibitors. *Comput. Struct. Biotechnol. J.* 17, 291–310.

Parker, J.M., Guo, D., Hodges, R.S., 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25, 5425–5432.

- Ponomarenko, J.V., Bui, H., Li, W., Fussedier, N., Bourne, P.E., Sette, A., Peters, B., 2008. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinf.* 9, 514.
- Saha, S., Raghava, G.P.S., 2004. BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: Nicosia, G., Cutello, V., Bentley, P.J., Timis, J. (Eds.), *ICARIS LNCS*, 3239, pp. 197–204.
- Sakharkar, K.R., Sakharkar, M.K., Chandra, R., 2015. Post-genomic Approaches in Drug and Vaccine Development, 5.
- Shey, R.A., Ghogomu, S.M., Esoh, K.K., Nebangwa, N.D., Shintouo, C.M., Nongley, N.F., Asa, B.F., Ngale, F.N., Vanhamme, L., Souopgui, J., 2019. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Sci. Rep.* 9 (1), 1–18.
- Song, W.F., Gui, M., Wang, X., 2018. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* 14 (8), e1007236.
- Sun, Z., Thilakavathy, K., Kumar, S.S., He, G., Liu, S.V., 2020. Potential factors influencing repeated SARS outbreaks in China. *Int. J. Environ. Res. Publ. Health* 17 (5), 1633.
- Sutton, P., Boag, J.M., 2018. Status of Vaccine Research and Development for *Helicobacter pylori*. *Vaccine*.
- Tortorici, M.A., Walls, A.C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., Boons, G.J., Bosch, B.J., Rey, F.A., de Groot, R.J., Veerles, D., 2019. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* 26 (6), 481–489.
- Trott, O., Olson, A.J., 2010. AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31 (2), 455–461.
- Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., Peters, B., 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput. Biol.* 4 (4), e1000048.
- Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35 (suppl_2), W407–W410.
- World Health Organization, 2020a. Coronavirus Disease 2019 (COVID-19) Situation Report-208.
- World Health Organization, 2020b. Novel coronavirus(2019-nCoV). In: Situation Report – 12.
- World Health Organization, 2020c. 2019-NCoV Outbreak Is an Emergency of International Concern. World Health Organization, 31 Jan.
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579 (7798), 265–269.
- Xiong, X.L., Tortorici, M.A., Snijder, J., 2018. Glycan shield and fusion activation of a delta coronavirus spike glycoprotein fine-tuned for enteric infections. *J. Virol.* 92 (4) e01628–17.
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., Zhou, Q., 2020. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science* 367 (6485), 1444–1448.