Immunoinformatics approach for multi-epitope vaccine design against structural proteins and ORF1a polyproteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2)

Khalid Mohamed Adam (dr.aboalbasher@gmail.com)
University of Bisha

Research

Keywords: Immunoinformatics, Multi-epitope, Vaccine, Coronavirus, COVID-19, SARS-COV-2

Posted Date: November 20th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-108652/v1

Version of Record: A version of this preprint was published at Tropical Diseases, Travel Medicine and Vaccines on July 8th, 2021. See the published version at https://doi.org/10.1186/s40794-021-00147-1.

Abstract

Background

The lack of effective treatment and a protective vaccine against the highly infectious SARS-CoV-2 has aggravated the already catastrophic global health issue. Here, in an attempt to design an efficient vaccine, a vigorous immunoinformatics approach was followed to predict the most suitable viral proteins epitopes for building that vaccine.

Methods

The amino acid sequences of four structural proteins (S, M, N, E) along with one potentially antigenic accessory protein (ORF1a) of SARS-CoV-2 were inspected for the most appropriate epitopes to be used for building the vaccine construct. Several immunoinformatics tools were used to assess the antigenicity, immunogenicity, allergenicity, toxigenicity, interferon-gamma inducing capacity, and the physicochemical properties of the product.

Results

The final candidate vaccine construct consisted of 468 amino acids, encompassing 30 epitopes. All the vaccine properties and its ability to trigger the humoral and cell-mediated immune response were validated computationally. Molecular modeling, docking to TLR3, simulation, and molecular dynamics were also carried out. Finally, a molecular clone using pET28: :mAID expression plasmid vector was prepared.

Conclusion

The overall results of the study suggest that the final multi-epitope chimeric construct is a potential candidate for an efficient protective vaccine against SARS-CoV-2.

1. Introduction

In early December 2019, an acute respiratory disease of unknown etiology emerged in Wuhan, China, which subsequently found to be caused by a highly contagious coronavirus. The virus was initially described as 2019-nCoV and later named by the International Committee on Taxonomy of Viruses (ICTV) as Severe Acute Respiratory Syndrome Coronavirus − 2 (SARS-CoV-2), while the World Health Organization (WHO) named the disease Coronavirus disease − 19 (Covid-19) [1–5]. Within the first three months after its discovery, the disease spread to more than 100 countries and caused more than 4,000 death worldwide [6]. On the 11th of March, 2020, the WHO categorized the newly discovered disease as a

ad clinical spectrum, ranging from asymptomatic, to mild or

subclinical to severe respiratory illness that required intubation and intensive care. The disease course and outcome are contingent on a number of factors, such as age and presence of underlying comorbidities [7]. The clinical manifestations include fever, fatigue, nonproductive cough, dyspnea, myalgia. In severe cases, acute respiratory distress syndrome (ARDS), acute cardiac injury, and acute kidney injury and death can also occur [8, 9].

SARS-CoV-2 along with Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) are β-coronaviruses belong to the subfamily *Coronavirinae* of the *Nidovirales coronaviridae.* These are enveloped, non-segmented, single-stranded, positive-sense RNA viruses, with genome ranging from 26 Kb to 32 Kb. The genome size of SARS-CoV-2 varies from 29.8 to 29.9 Kb, with typical genome structure of earlier well-characterized coronaviruses, covering ORF1ab region and genes encoding 4 structural proteins including spike or surface proteins (S), envelope proteins (E), membrane proteins (M), and nucleocapsid proteins (N), in addition to accessory proteins coding genes ORF3a, ORF6, ORF7a, ORF7b and ORF8 [10–13]. The main role of the spike glycoproteins (S) is to mediate binding to the angiotensin-converting enzyme 2 (ACE-2) receptor and promote membrane fusion and virus entry [14]. Both M and E proteins were reported to play important roles in viral entry, replication, and virions assembly [15]. Proteins (N) are important for viral RNA packaging, virions release and interferon inhibition, promoting the virus pathogenicity [16, 17]. In SARS-CoV, the gene for Protein (N) is upregulated, producing large amounts of the highly immunogenic protein (N) [18]. On the other hand, ORF1a encodes nonstructural polyproteins 1a and 1b (PP1a, PP1b), these polyproteins are involved in viral genome replication and transcription [19].

Covid-19 pandemic has affected all walks of life, stretching health-care systems to their maximum and putting a huge economical, psychological, and mental burden on the entire world population. This dire situation is aggravated by the highly contagious nature of the virus, lack of proper understanding of the disease course and the absence of a reliable cure [6]. The disease containment measures used thus far, are contingent on disrupting the transmissibility of the virus through rapid identification and isolation of infected and carrier individuals, this entails the search for reliable vaccines and effective treatments, hence, the aim of this study is to construct a multi-epitope vaccine against SARS-CoV-2 based on the structural proteins along with the nonstructural proteins of ORF1a, using reverse vaccinology approach.

The selection of the nonstructural ORF1a protein alongside the structural viral proteins in this study was entirely driven by suggestions made by a number of studies, that nonstructural proteins induce immunity and are applicable to prophylaxis of viral disease [20–23].

2. Materials And Methods

2.1 Retrieval of target proteins sequences:

The amino acid sequences for S. protein of 1273 amino acid (Accession No. QLI51913.1), M. protein of 222 amino acid (Accession No. QLI52072.1), E. protein of 75 amino acid (Accession No. QLI52071.1), N.

protein of 419 amino acid (Accession No. QIH45060.1) and ORF1a polyprotein of 4405 amino acid (Accession No. QJQ84087.1) were retrieved from NCBI protein database (https://www.ncbi.nlm.nih.gov/protein) in FASTA format.

2.2 Cytotoxic T- cell lymphocyte (CTL) epitopes prediction

Initially, the amino acid sequences of all 5 proteins were screened for antigenicity with VaxiJen 2.0 server (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) with threshold value of 0.4 [24]. The CTL epitopes for all sequences were predicted using artificial neural network algorithm-based NetCTL 1.2 server (http://www.cbs.dtu.dk/services/NetCTL/) [25], which predicts MHC I binding epitopes. The peptides obtained were then checked for antigenicity using VaxiJen 2.0 server. The antigenic peptides were then submitted for virtual scanning for toxic peptides using ToxinPred server (http://crdd.osdd.net/raghava/toxinpred/multi_submit.php) [26]. The immunogenicity of the resultant non-toxin epitopes was determined using class I immunogenicity tool of Immune Epitope Database (IEDB) (http://tools.iedb.org/immunogenicity/) [27].

2.3 Helper T- Lymphocyte (HTL) epitopes prediction

For prediction of HTL epitopes, MHC-II binding tool of IEDB (http://tools.iedb.org/mhcii/) was used [28], selecting 7-allele HLA reference set that includes; HLA-DRB1*03:01, HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB3*01:01, HLA-DRB3*02:02, HLA-DRB4*01:01, HLA-DRB5*01:01. The resultant epitopes with low percentile ranks were then checked for allergenicity with AlgPred serve (https://webs.iiitd.edu.in/raghava/algpred/submission.html), using SVM module based on amino acid composition as prediction approach [29]. The antigenicity and toxicity status of the non-allergenic epitopes was determined using VaxiJen 2.0 server and ToxinPred server, respectively. Finally, interferon-gamma inducing epitopes were predicted with IFNepitope server (http://crdd.osdd.net/raghava/ifnepitope/) following the Motif and SVM hybrid approach [30]. The resultant epitopes were then inspected for overlapping.

2.3 Population coverage

The prediction of worldwide population coverage of the selected epitopes for MHC-I and MHC-II alleles was carried out using population coverage tool of IEDB (http://tools.iedb.org/population/) [31], calculating the coverage for class I and class II separately and combined.

2.4 B-cell epitopes prediction

The linear B-cell epitopes of all proteins under study were predicted with the antibody epitope prediction tool of IEDB (http://tools.iedb.org/bcell/) using Bepipred linear epitope prediction method 2.0 [32], Emini surface accessibility prediction method [33], and Kolaskar and Tongaonkar antigenicity method [34].

2.5 Construction of multiepitope vaccine sequence

To ensure efficient vaccine construction and proper epitope separation, all candidate epitopes were joined together using linkers. The B-cell epitope and CTL epitopes were linked with AAY linker, and HTL epitopes

were linked together and to the CTL epitopes with GPGPG linker. To facilitate future conjugation of the multi-epitope vaccine construct with a carrier protein, a cysteine residue was added at the N-terminal [35]. Furthermore, a four amino acid (EPEA) tag was added at the C-terminal for efficient purification [36]. The vaccine construct was subjected to further analysis to assess its antigenicity with VaxiJen 2.0 server, allergenicity with AlgPred server, physicochemical properties with ProtParam tool (https://web.expasy.org/protparam/) [37].

2.6 Modeling, structure validation and molecular docking

The secondary structure of the novel vaccine construct was determined using PSIPred server (http://bioinf.cs.ucl.ac.uk/psipred/) [38]. Protein modeling was carried out using threading and *ab initio* approaches with IntFold and trRosetta servers (https://yanglab.nankai.edu.cn/trRosetta/) [39], further protein structure analysis and model validation carried out using ProSA-web server (https://prosa.services.came.sbg.ac.at/prosa.php) [40], Ramachandran plot analysis using RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [41] and ERRAT server (https://servicesn.mbi.ucla.edu/ERRAT/) [42]. The vaccine construct was subjected to molecular docking with Toll-like receptor – 3 (TLR-3) using FRODOCK (http://frodock.chaconlab.org) and Vakser lab of GRAMM-X simulation severs and (http://vakser.compbio.ku.edu/resources/gramm/grammx/index_html) [43].

2.7 Immune response simulation

The immune response to the novel multi-epitope vaccine construct was carried out using C-ImmSim server 10.1 (http://www.cbs.dtu.dk/services/C-ImmSim-10.1/) [44]. The simulation parameters used were, random seed: 12345, simulation steps: 100 and simulation volume: 10 micro L.

2.8 *In silico* molecular cloning

The amino acid sequence for the candidate vaccine was then subjected to reverse translation and codon optimization with JAVA codon adaptation tool (Jcat) (http://www.jcat.de) [45]. The DNA sequence was then used for *in silico* molecular cloning with expression plasmid vector pET28::mAID [46] using Snapgene software.

3. Results

3.1 T - cell epitopes prediction

The initial screening of amino acid sequences of all 5 proteins for antigenicity, showed a score greater than the threshold value of 0.4 indicating probable antigens, these sequences were then submitted to NetCTL server to predict possible CTL epitopes, which resulted in 37 possible epitopes for S. protein, out of which 14 showed no toxicity and 8 positive immunogenicity score. Ultimately, the top 4 epitopes were selected for inclusion in the multi-epitope vaccine construct. For M. protein, 10 epitopes were predicted, 5 epitopes were non-toxin and one showed a positive immunogenicity score. For E. protein, 3 epitopes were

predicted, two of which showed an antigenicity score higher than the threshold value and non-toxic, but neither showed a positive immunogenicity score, hence, not included in the vaccine construct.

For N. protein, 9 epitopes were predicted, 6 showed an antigenicity score higher than the threshold value, all 6 predicted epitopes were non-toxin, of which, five showed positive immunogenicity score, and only the top two were selected to be included in the construct. For the nonstructural polyproteins, on the other hand, 170 epitopes were predicted, of which 96 showed antigenicity score higher than the threshold, and the best 12 were selected based on toxigenicity and immunogenicity results, Table 1.

Table 1
CTL predicted epitopes of all five proteins with the antigenicity, toxicity and immunogenicity results

Protein	peptide	length	Score	toxicity	immunogenicity
S.	QLTPTWRVY	9 mer	1.2119	Non-toxic	0.31555
	VLPFNDGVY	9 mer	0.4642	Non-toxic	0.1815
	WTAGAAAYY	9 mer	0.6306	Non-toxic	0.15259
	CNDPFLGVY	9 mer	0.4295	Non-toxic	0.15232
M.	AGDSGFAAY	9 mer	0.9095	Non-toxic	0.03981
N.	LSPRWYFYY	9 mer	1.2832	Non-toxic	0.35734
	DLSPRWYFY	9 mer	1.7645	Non-toxic	0.25933
ORF1a	VSDIDITFL	9 mer	2.2906	Non-toxic	0.38916
	TLRVEAFEY	9 mer	0.4509	Non-toxic	0.34997
	HVGEIPVAY	9 mer	0.6413	Non-toxic	0.28861
	STNVTIATY	9 mer	0.7143	Non-toxic	0.25822
	LVSDIDITF	9 mer	1.7830	Non-toxic	0.2541
	NGDVVAIDY	9 mer	0.6625	Non-toxic	0.25105
	VVDYGARFY	9 mer	0.4908	Non-toxic	0.18539
	GTDPYEDFQ	9 mer	0.5315	Non-toxic	0.17381
	VTNNTFTLK	9 mer	0.7146	Non-toxic	0.16567
	ETSWQTGDF	9 mer	1.3140	Non-toxic	0.13449
	FMGRIRSVY	9 mer	0.5212	Non-toxic	0.1259
	VVVNAANVY	9 mer	0.4078	Non-toxic	0.10048

The HTL epitopes prediction with MHC-II binding tool of IEDB and based on percentile rank less than 10, e non-allergenic, 10 were non-toxic and a single epitope

showed a positive interferon-gamma induction result. For M. protein were the predicted HTL epitopes were 55, out of which 43 were non-allergenic antigenic non-toxic epitopes, and only 3 epitopes showed positive interferon-gamma induction results. None of the predicted HTL epitopes of N. protein showed interferon-gamma positive results, therefore none were in the vaccine construct. Similarly, all HTL epitopes predicted for E. protein failed to pass either the antigenicity, allergenicity, or interferon-gamma induction assessment. Out of 96 HTL predicted epitopes for polyproteins of ORF1a, only 6 epitopes passed the antigenicity, allergenicity, toxigenicity, and interferon-gamma induction assessment, results are shown in Table 2.

Table 2
HTL predicted epitopes of all five proteins, antigenicity score and IFN-gamma score

Protein	peptide	length	Antigenicity	IFN-gamma
S.	TRFASVYAWNRKRIS	15 mer	0.4963	0.7315567
	FQTLLALHRSYLTPG	15 mer	0.5789	0.26071055
	QPYRVVVLSFELLHA	15 mer	0.9109	0.60855322
M.	SRTLSYYKLGASQRV	15 mer	0.5731	0.09462399
	LVGLMWLSYFIASFR	15 mer	0.5535	0.20134879
ORF1a	VSTQEFRYMNSQGLL	15 mer	0.4972	0.97632046
	AALGVLMSNLGMPSY	15 mer	0.8521	0.11429986
	TLNGLWLDDVVYCPR	15 mer	0.4558	0.04915351
	AYESLRPDTRYVLMD	15 mer	0.5553	0.30777655
	SAGIFGADPIHSLRV	15 mer	0.5839	0.24837266
	MFTPLVPFWITIAYI	15 mer	0.6806	0.1415124

3.2 B-cell epitopes prediction

The B-cell epitopes are an important part of the multi-epitope vaccine because recognition of these epitopes by B lymphocytes elicit antibody production, which is a key process in adaptive immunity. For all five proteins, linear B-cell epitopes were predicted using Bepipred Linear Epitope Prediction 2.0 method, Emini surface accessibility prediction method, and Kolaskar & Tongaonkar antigenicity method, these methods were selected because they assess properties that are important for predicting potential epitopes, such as antigenicity, surface accessibility, and flexibility. The resultant plots were then inspected for overlapping regions showing epitopes by the three methods. The only protein to show such an overlapping region was N. protein with a sequence of 10 amino acids from 380–390. The results of all amino acid sequences are shown in Fig. 1.

3.3 Population coverage

The selected epitopes were then analyzed to determine the percentage of the world population coverage for MHC-I and MHC-II alleles. The MHC-I alleles assessed included; HLA-B*15:01, HLA-A*30:02, HLA-A*01:01, HLA-B*40:01, HLA-B*07:02, HLA-B*51:01, HLA-A*68:02, HLA-A*02:01, HLA-A*02:06, HLA-B*08:01, HLA-A*02:03, HLA-A*33:01, HLA-A*24:02, HLA-A*23:01, HLA-B*44:03, HLA-B*44:02, HLA-A*31:01, HLA-B*53:01, HLA-A*11:01, HLA-A*68:01, HLA-A*30:01, HLA-B*57:01, HLA-A*03:01, HLA-A*26:01, HLA-B*58:01, HLA-A*32:01, HLA-B*35:0. The world coverage for these alleles was 98.55%. For MHC-II, the alleles assessed included; HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB3*01:01. The coverage for these alleles was 49.02. The combine allele coverage for both MHC-I and MHC-II was found to be 99.26% which indicates a high population coverage for selected epitopes. Figure 2.

3.4 Multi-epitope vaccine construction

For the construction of the final vaccine construct, the most appropriate predicted epitopes were selected, this included one B-cell linear epitope from N. protein, 4 CTL and 3 HTL epitopes from S. protein, one CTL and 2 HTL epitopes from M. protein, 2 CTL epitopes from N. protein, 12 CTL and 6 HTL epitopes from ORF1a. These epitopes were joined together with two types of linkers, AAY for linear B-cell and CTL epitopes, and GPGPG for HTL epitopes, with cysteine residue at the N-terminal and EPEA tag at C-terminal, this yielded the following 468 amino acid peptide chain:

CQALPQRQKKQQAAYQLTPTWRVYAAYVLPFNDGVYAAYWTAGAAAYYAAYCNDPFLGVYAAYAGDSGFA

AYAAYLSPRWYFYYAAYSPDDQIGYYAAYVSDIDITFLAAYTLRVEAFEYAAYHVGEIPVAYAAYSTNVTIATY

AAYLVSDIDITFAANGDVVAIDYAAYVVDYGARFYAAYGTDPYEDFQAAYVTNNTFTLKAAYETSWQTGDFA

AYFMGRIRSVYAAYVVVNAANVYGPGPGTRFASVYAWNRKRISGPGPGFQTLLALHRSYLTPGGPGPGQPYR

VVVLSFELLHAGPGPGSRTLSYYKLGASQRVGPGPGLVGLMWLSYFIASFRGPGPGVSTQEFRYMNSQGLLG

PGPGAALGVLMSNLGMPSYGPGPGTLNGLWLDDVVYCPRGPGPGAYESLRPDTRYVLMDGPGPGSAGIFGA

DPIHSLRVGPGPMFTPLVPFWITIAYIGPGPGEPEA

3.5 Physiochemical properties of the vaccine construct

The results obtained from the ProtParam server, showed that the novel vaccine construct has a molecular weight of 50.417 Ka which is an optimum molecular weight for an antigenic protein. The theoretical PI for the construct was 5.41 indicating an acidic nature, with a total of 30 negatively charged residues and 25 positively charged residues. The estimated half-life is 1.2 hours in mammalian reticulocytes in vitro, > 20 hours in yeast in vivo, and > 10 hours in E. coli in vivo, indicating a good construct for future cloning. The instability index was computed to be 30.79 suggesting stable protein. The aliphatic index of 75.38, which indicates a thermostable protein. The grand average of hydropathicity (GRAVY) was 0.040, a positive value close to zero means a slightly hydrophobic molecule.

3.6 Vaccine modeling and structure analysis

Based on the amino acid sequence of the vaccine construct, the result of the PSIpred server revealed different secondary structures. This is considered a primary step towards predicting the three-dimensional structure of the protein. Figure 4

The 3D protein model was then predicted with two modeling approaches; threading model with IntFOLD server and Ab initio modeling with the trRosetta server, the resultant models were then analyzed with Ramachandran plot and ProSA-web z-score based on X-ray crystallography and NMR analysis, the best-predicted model showed 98% of the residues in the favorable region in Ramachandran plot, Fig. 4A, and z-score of – 6.01, determined by x-ray crystallography, Fig. 4B.

The statistics of non-bonded interactions between different atom types, and then the error function value was plotted against a position of a-9 residue sliding window, calculated by comparison with statistics from highly refined structures, carried out using ERRAT server, and the calculated error value obtained was 81.928, which falls well below 91% indicating a relatively average overall quality for the selected protein model, this is can be justified by the fact that the modeling process was carried out using ab initio modeling approach. Figure 5A.

3.7 Molecular docking and dynamics

The final vaccine construct was docked with Toll-like receptor 3 (PDB ID: 1ziw) using the FRODOCK server. The docked vaccine-receptor complex was then prepared for simulation using a protein-prep wizard and Pymol software using the default settings, the molecular dynamics simulation was then carried out using the Desmond tool and Superpose1.0 server (http://superpose.wishartlab.com) for calculating the root mean square deviation (RMSD) value of 3.78 which suggests a relatively poor binding pose at the site of the receptor and vaccine binding. Figure 5B.

3.8 Immune response simulation

Measuring the immune response is a pivotal step for vaccine designing, this contingent on a number of algorithms that make use of mathematical models to illustrate the fine details of the immunological process. In the present study, the C-ImmSim server was used to simulate immune response with the candidate vaccine construct. Simulation with this tool focuses on B-cell epitope binding, class I and II HLA epitope binding, and the binding of the T-cell receptor to HLA-peptide complexes, the tool then details the dynamics of immune cells populations and the molecules involved in the immune response [47].

The simulation results showed an increased and sustained level of B- memory and active cells, and a high level of IgM, which represents the primary response against the antigen, this suggests effective humoral response, Fig. 6. A & B. T helper cell population showed very promising results, as the levels of memory helper cells and active T helper cells remained high for the entire period of simulation, suggesting prolonged humoral and cell-mediated immune response, Fig. 6. C & D. The results of the T cytotoxic cell population steady level of the memory cells, while the active cell population showed an increased level throughout the stimulation period, Fig. 6. E & F. The result of different immunoglobulin [illegible] weeks followed by a gradual decline, similar result was shown

by interferon-gamma level, this can be viewed as a positive point, hence, the first two weeks are considered detrimental for the course and outcome of the disease. [48] Fig. 6. G & H.

3.9 *In silico* molecular cloning

The DNA sequence produced by Jcat showed a GC content of 56% and a codon adaptation index of 1.0, which indicate a stable DNA sequence and a high level of protein expression. Figure 7.

4. Discussion

The current Covid-19 pandemic associated with SARS-CoV-2 infection is the third coronavirus outbreak in the last 20 years besides the severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS). SARS-CoV-2 shows relatively higher transmissibility as compared to other emerging viruses such as H7N9 and MERS-CoV [49, 50]. This entails the imperative search for effective vaccine and treatment in addition to the protective and social distancing measures to contain and control the disease. The immunoinformatics approach provides a promising tool for designing and exploring potential vaccines against bacterial, parasitic, and viral diseases [51]. In this study, a multi-epitope vaccine was constructed using all the virus structural proteins and the largest non-structural polyprotein [52]. These proteins were selected based on suggestions from previous studies [53, 54]. Unlike the single subunit vaccine, the multi-epitope vaccine is believed to induce a better and more protective immune response [55]. In the present study antigenic, non-allergenic, and non-toxic epitopes were identified and used for the construction of the final candidate vaccine. All five proteins were studied for potential epitopes, however, none of the peptides from the envelope protein (E) was eligible for the selection in the final vaccine construct, due to either lack of antigenicity or the allergenicity and toxicity of these peptides. For the final vaccine construct, CTL, HTL, line B-cell epitopes were linked together using AAY, and GPGPG linkers which provide proper proteasomes cleavage sites for different immune cells [56] which will ultimately enhance the antigen presentation process by binding TAP transporters [57]. Furthermore, linking of CTL epitopes from different proteins together forms epitopes on a string which is believed to enhance the immunogenicity of CTL epitopes [58]. To the N-terminal of the vaccine construct a cysteine residue was added to facilitate the binding of this vaccine to protein carrier [35], and to the C- terminal, a small peptide of four amino acids EPEA was added to enable downstream purification process [36]. The candidate vaccine construct consists of 486 amino acid, which is an ideal vaccine length, hence, larger proteins are presented by dendritic cells leading to stronger T-cell immune response [57], while extremely short peptides may induce tolerance and anergy by directly binding MHC molecules of non-professional antigen-presenting cells [59]. Determination of the secondary structure of the protein is a pivotal step towards the prediction of its three-dimensional structure, therefore, the secondary structure of the candidate vaccine was determined using PSIpred server, followed by structure refinement, and protein modeling. Two approaches were used for modeling the protein, threading approach, and ab initio approach, the best resultant model was selected based on the Ramachandran plot and z-score analyses. The docking of the vaccine and TLR-3 showed a possible hydrophilic interaction [60], this interaction indicates a possible recognition of the vaccine by APC specific receptor, which in turn promotes the

immune response [61]. The physicochemical properties of the vaccine construct. The results of immune response simulation showed very promising results, with a sustained response for the cells involved in the humoral and cell-mediated immunity against SARS-CoV-2.

The conventional methods of vaccine development are very costly and time-consuming, alternatively, the immunoinformatics approach has attracted the attention as an ideal method for designing less-expensive, rapid, efficient, multi-epitope vaccines. However, experimental validation is of utmost importance to ensure the safety and efficacy of the resultant vaccine.

5. Conclusion

The highly contagious nature of SARS-CoV-2 left the entire world population with no option but to wait for the production of a safe and protective vaccine to break the chain of infection and tackle the spread of this pandemic. It is rather impractical to rely on the conventional methods for producing such a vaccine due to a number of limiting factors. This study is an attempt to design an efficient multi-epitope chimeric subunit vaccine that is capable of mounting a strong immune response by induction of both humoral and cellular mediated immunity, with the help of a large number of immunoinformatics tools. The vaccine construct effectively fulfilled the requirements for characteristics such as antigenicity, allergenicity, immunogenicity, physiochemical properties, eliciting the immune response in a simulation model. It is concluded that this novel construct represents a promising candidate for an efficient protective vaccine against SARS-CoV-2.

Declarations

Acknowledgment

I am indebted to the faculty members of college of applied medical sciences and the deanship of scientific research, University of Bisha.

Funding: Deanship of scientific research at University of Bisha, COVID-19 initiative project, Grant No. (UB-COVID-15-1441

Author information

Department of Medical Laboratory Science, College of Applied Medical Sciences, University of Bisha, Saudi Arabia

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Competing interests: The author declare that he has no competing interests.

Availability of data: Not applicable

References

1. Fagbule OF. 2019 NOVEL CORONAVIRUS. Ann Ib Postgrad Med 2019;17:108-110.
2. Du Toit A. Outbreak of a novel coronavirus. Nat Rev Microbiol 2020;18:123.
3. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet 2020;395:470-473.
4. Xu YH, Dong JH, An WM, Lv XY, Yin XP, Zhang JZ, et al. Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. Journal of Infection 2020;80:394-400.
5. Wu YC, Chen CS, Chan YJ. The outbreak of COVID-19: An overview. J Chin Med Assoc 2020;83:217-220.
6. Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). Clinical and experimental pediatrics 2020;63:119-124
7. Sun Y, Koh V, Marimuthu K, Ng OT, Young B, Vasoo S, et al. Epidemiological and clinical predictors of COVID-19. Clinical Infectious Diseases 2020;71:786-792
8. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. Jama 2020;323:1061-1069.
9. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet 2020;395:497-506.
10. Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. Gene reports 2020;19:100682.
11. Astuti I, Yasrafil. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response. Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2020;14:407-4012.
12. Zhang G, Zhang J, Wang B, Zhu X, Wang Q, Qiu S. Analysis of clinical characteristics and laboratory findings of 95 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a retrospective analysis. Respiratory research 2020;21:s12931.
13. Phan T. Novel coronavirus: From discovery to clinical diagnostics. Infection, Genetics and Evolution 2020;79:104211.
14. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nature communications 2020;11:1-12.
15. Bianchi M, Benvenuto D, Giovanetti M, Angeletti S, Ciccozzi M, Pascarella S. Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics?. BioMed Research

International 2020;2020.

16. Zeng W, Liu G, Ma H, Zhao D, Yang Y, Liu M, Mohammed A, Zhao C, Yang Y, Xie J, Ding C. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochemical and biophysical research communications 2020;527:618-623.
17. Alanagreh LA, Alzoughool F, Atoum M. The human coronavirus disease COVID-19: its origin, characteristics, and insights into potential drugs and its mechanisms. Pathogens 2020;9:331.
18. Cong Y, Ulasli M, Schepers H, Mauthe M, V'kovski P, Kriegenburg F, et al. Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. Journal of virology 2020;94:e01925.
19. Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. Journal of medical virology 2020;92:418-423.
20. Gibson CA, Schlesinger JJ, Barrett AD. Prospects for a virus non-structural protein as a subunit vaccine. Vaccine 1988;6:7-9.
21. Ip PP, Boerma A, Regts J, Meijerhof T, Wilschut J, Nijman HW, Daemen T. Alphavirus-based vaccines encoding nonstructural proteins of hepatitis C virus induce robust and protective T-cell responses. Molecular Therapy 2014;22:881-890.
22. Ludert JE, Reyes-Sandoval A. The dual role of the antibody response against the flavivirus non-structural protein 1 (NS1) in protection and immunopathogenesis. Frontiers in immunology 2019;10:1651.
23. Henriques HR, Rampazo EV, Gonçalves AJ, Vicentin EC, Amorim JH, Panatieri RH, et al. Targeting the non-structural protein 1 from dengue virus to a dendritic cell population confers protective immunity to lethal virus challenge. PLoS Negl Trop Dis 2013;7:e2330.
24. Irini A Doytchinova and Darren R Flower. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007;8:4.
25. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 2007;8:424.
26. Gupta *al.*: *In silico*Approach for Predicting Toxicity of Peptides and Proteins. PLoS ONE 8(9):e73957.
27. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PloS Comp. Biol* 2013;9:e1003266.
28. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 2018;154:394-406.
29. Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research 2006;34: 202-209.
30. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. Biology direct 2013;8:30.
31. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC bioinformatics 2006;7:1-5.

32. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome research 2006;2:1-7.
33. Emini EA, Hughes JV, Perlow D, Boger J. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of virology 1985;55:836-839.
34. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS letters 1990;276:172-174.
35. Bandyopadhyay A, Cambray S, Gao J. Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation. Chemical science 2016;7:4589-93.
36. Jin J, Hjerrild KA, Silk SE, Brown RE, Labbé GM, Marshall JM, et al. Accelerating the clinical development of protein-based vaccines for malaria by efficient purification using a four amino acid C-terminal 'C-tag'. International journal for parasitology 2017;47:435-46.
37. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook. Humana press 2005; 571-607.
38. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology 1999;292:195-202.
39. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. Proceedings of the National Academy of Sciences 2020;117:1496-503.
40. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins: Structure, Function, and Bioinformatics 1993;17:355-62.
41. Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. Proteins 2003;50:437-50.
42. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein science 1993;2:1511-9.
43. Tovchigrechko A, Vakser IA. Development and testing of an automated approach to protein docking. Proteins: Structure, Function, and Bioinformatics 2005;60:296-301.
44. Rapin N, Lund O, Bernaschi M, Castiglione F. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One 2010;5:e9862.
45. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research 2005;33:W526-31.
46. Besmer E, Market E, Papavasiliou FN. The transcription elongation complex directs activation-induced cytidine deaminase-mediated DNA deamination. Molecular and cellular biology 2006;26:4378-85.
47. Rapin N, Lund O, Castiglione F. Immune system simulation online. Bioinformatics 2011;27:2013-4.

48. Bar-On YM, Flamholz A, Phillips R, Milo R. Science Forum: SARS-CoV-2 (COVID-19) by the numbers. Elife 2020;9:e57309.
49. Dhama K, Khan S, Tiwari R, Sircar S, Bhat S, Malik YS, et al. Coronavirus Disease 2019 – COVID-19. Clinical Microbiology Reviews 2020;33:e00028-20
50. Tobaiqy M, Qashqary M, Al-Dahery S, Mujallad A, Hershan AA, Kamal MA, et al. Therapeutic management of patients with COVID-19: a systematic review. Infection Prevention in Practice 2020;2:100061
51. Kaur R, Arora N, Jamakhani MA, Malik S, Kumar P, Anjum F, et al. Development of multi-epitope chimeric vaccine against Taenia solium by exploring its proteome: an in silico approach. Expert Review of Vaccines 2020;19:105-14.
52. Amawi H, Abu Deiab GA, A Aljabali AA, Dua K, Tambuwala MM. COVID-19 pandemic: an overview of epidemiology, pathogenesis, diagnostics and potential vaccines and therapeutics. Therapeutic delivery 2020;11:245-68.
53. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses 2020;12:254.
54. Wang J, Wen J, Li J, Yin J, Zhu Q, Wang H, et al. Assessment of immunoreactive synthetic peptides from the structural proteins of severe acute respiratory syndrome coronavirus. Clinical Chemistry 2003;49:1989-96.
55. Saadi M, Karkhah A, Nouri HR. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. Infection, Genetics and Evolution 2017;51:227-34.
56. Chen X, Zaro JL, Shen WC. Fusion protein linkers: property, design and functionality. Advanced drug delivery reviews 2013;65:1357-69.
57. Nezafat N, Ghasemi Y, Javadi G, Khoshnoud MJ, Omidinia E. A novel multi-epitope peptide vaccine against cancer: an in silico approach. Journal of theoretical biology 2014;349:121-34.
58. Livingston B, Crimi C, Newman M, Higashimoto Y, Appella E, Sidney J, et al. A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. The Journal of Immunology 2002;168:5499-506.
59. Melief CJ, Van Der Burg SH. Immunotherapy of established (pre) malignant disease by synthetic long peptide vaccines. Nature Reviews Cancer 2008;8:351-60.
60. Botos I, Segal DM, Davies DR. The structural biology of Toll-like receptors. Structure 2011;19:447-59.
61. Droppa-Almeida D, Franceschi E, Padilha FF. Immune-informatic analysis and design of peptide vaccine from multi-epitopes against Corynebacterium pseudotuberculosis. Bioinformatics and Biology Insights 2018;12:1177932218755337.

Figures

Figure 1

The resultant plots were then inspected for overlapping regions showing epitopes by the three methods. The only protein to show such an overlapping region was N. protein with a sequence of 10 amino acids from 380-390. The results of all amino acid sequences are shown in Figure 1.

MHC class	Coverage	Average hit	PC90
combined	99.26%	49.67	30.08

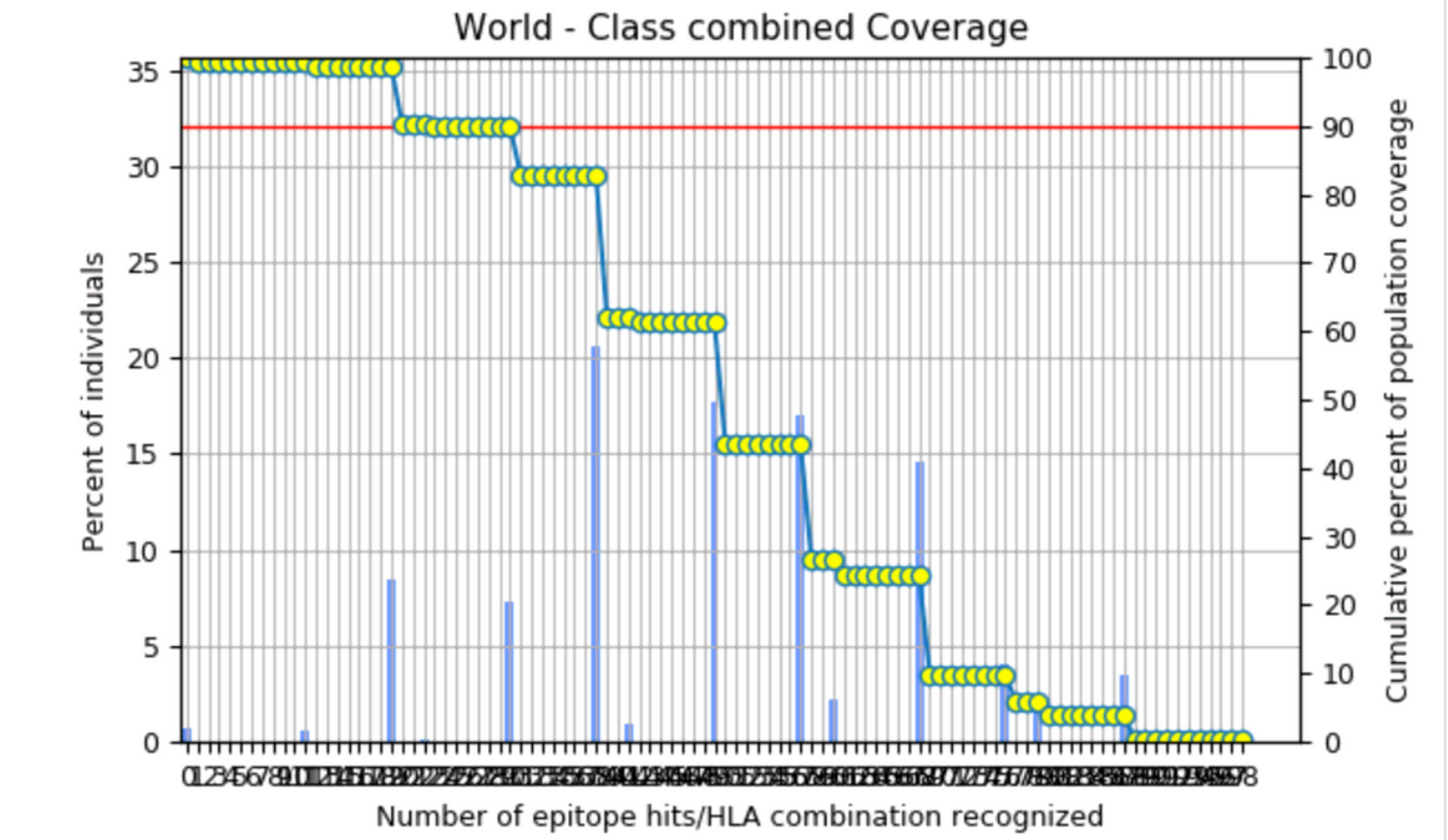

Figure 2

The combine allele coverage for both MHC-I and MHC-II was found to be 99.26% which indicates a high population coverage for selected epitopes. Figure 2.

Figure 3

Based on the amino acid sequence of the vaccine construct, the result of the PSIpred server revealed different secondary structures.

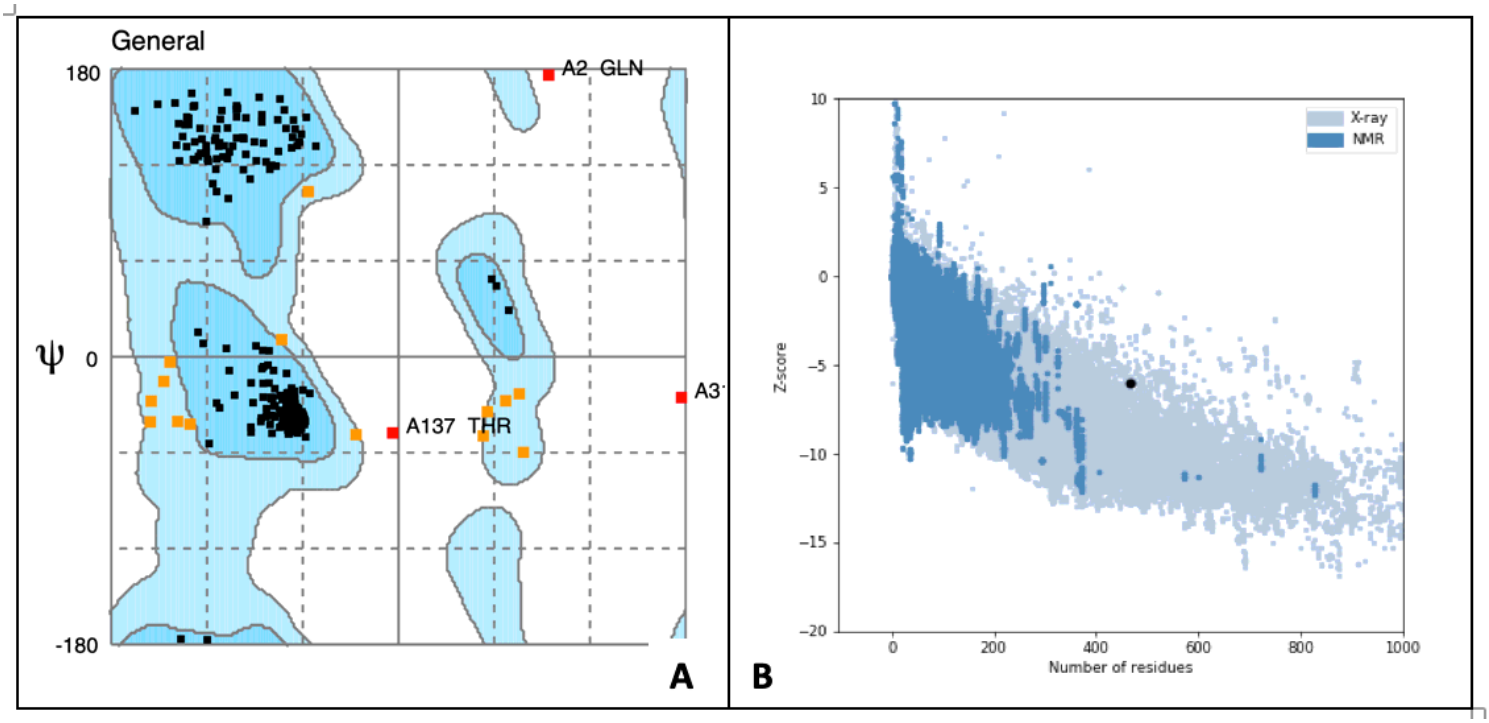

Figure 4

This is considered a primary step towards predicting the three-dimensional structure of the protein. Figure 4

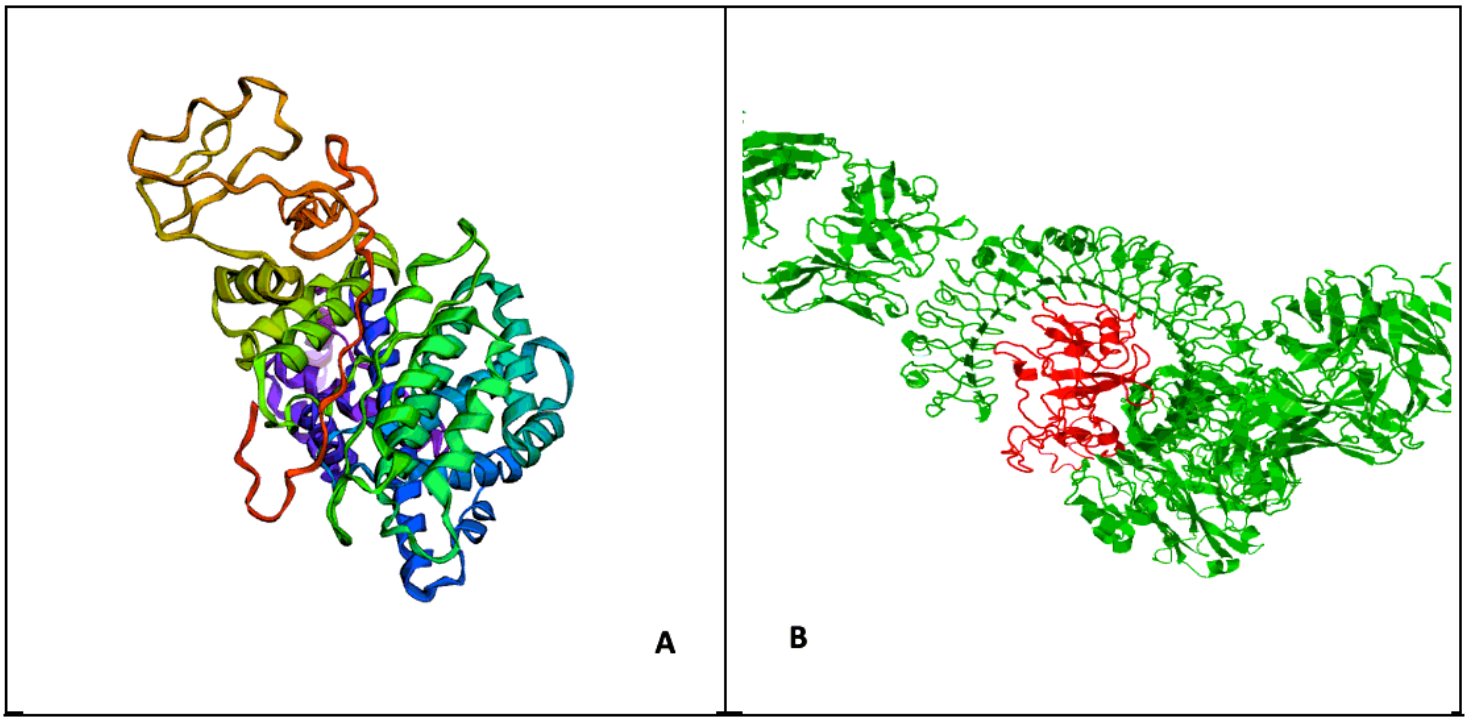

Figure 5

this is can be justified by the fact that the modeling process was carried out using ab initio modeling

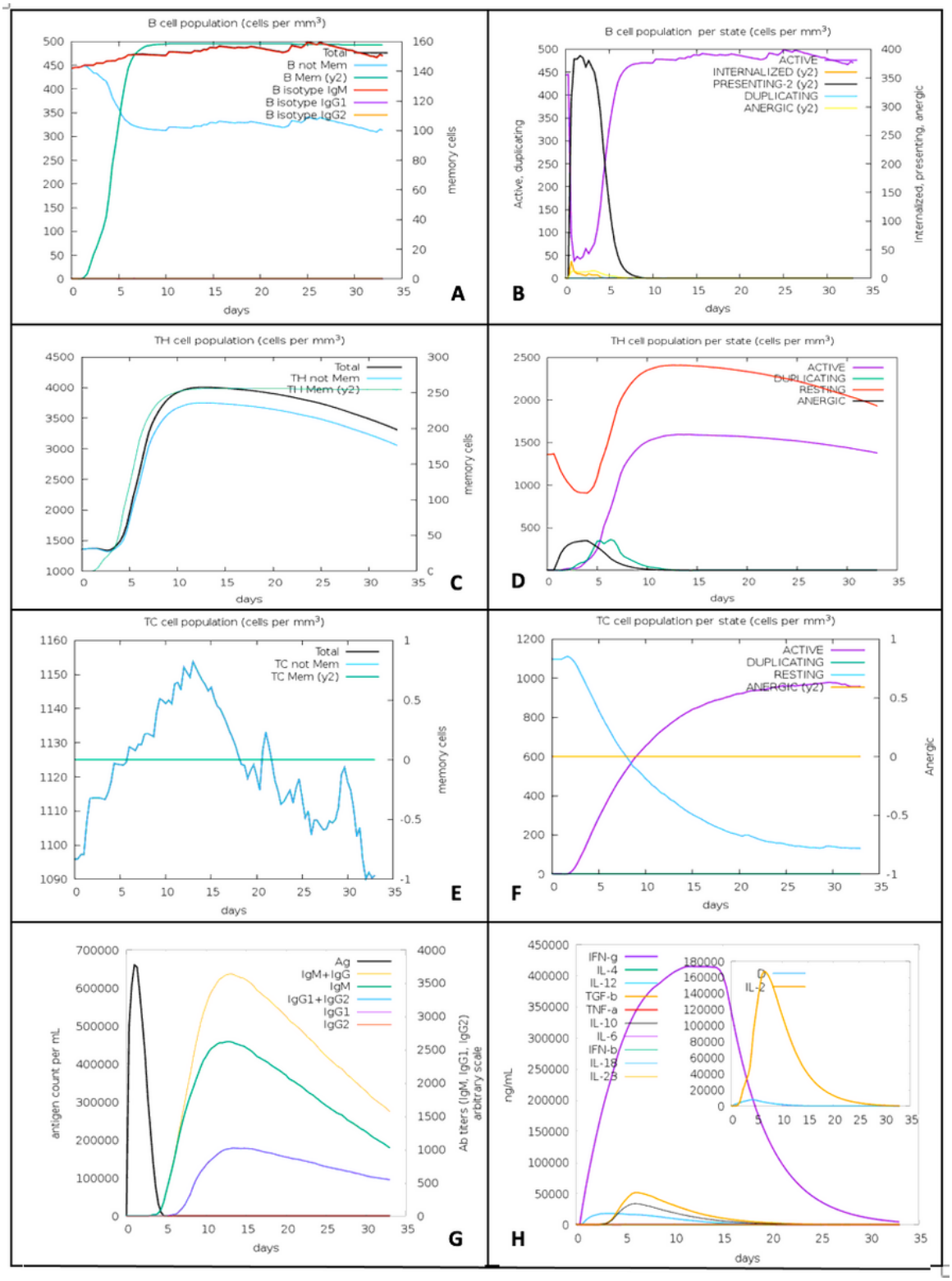

Figure 6

The simulation results showed an increased and sustained level of B- memory and active cells, and a high level of IgM, which represents the primary response against the antigen, this suggests effective humoral response, Figure 6. A & B. T helper cell population showed very promising results, as the levels of memory helper cells and active T helper cells remained high for the entire period of simulation,

iated immune response, Figure 6. C & D. The results of the T

cytotoxic cell population steady level of the memory cells, while the active cell population showed an increased level throughout the stimulation period, Figure 6. E & F. The result of different immunoglobulin isotopes showed high level in the first two weeks followed by a gradual decline, similar result was shown by interferon-gamma level, this can be viewed as a positive point, hence, the first two weeks are considered detrimental for the course and outcome of the disease. [48] Figure 6. G & H.

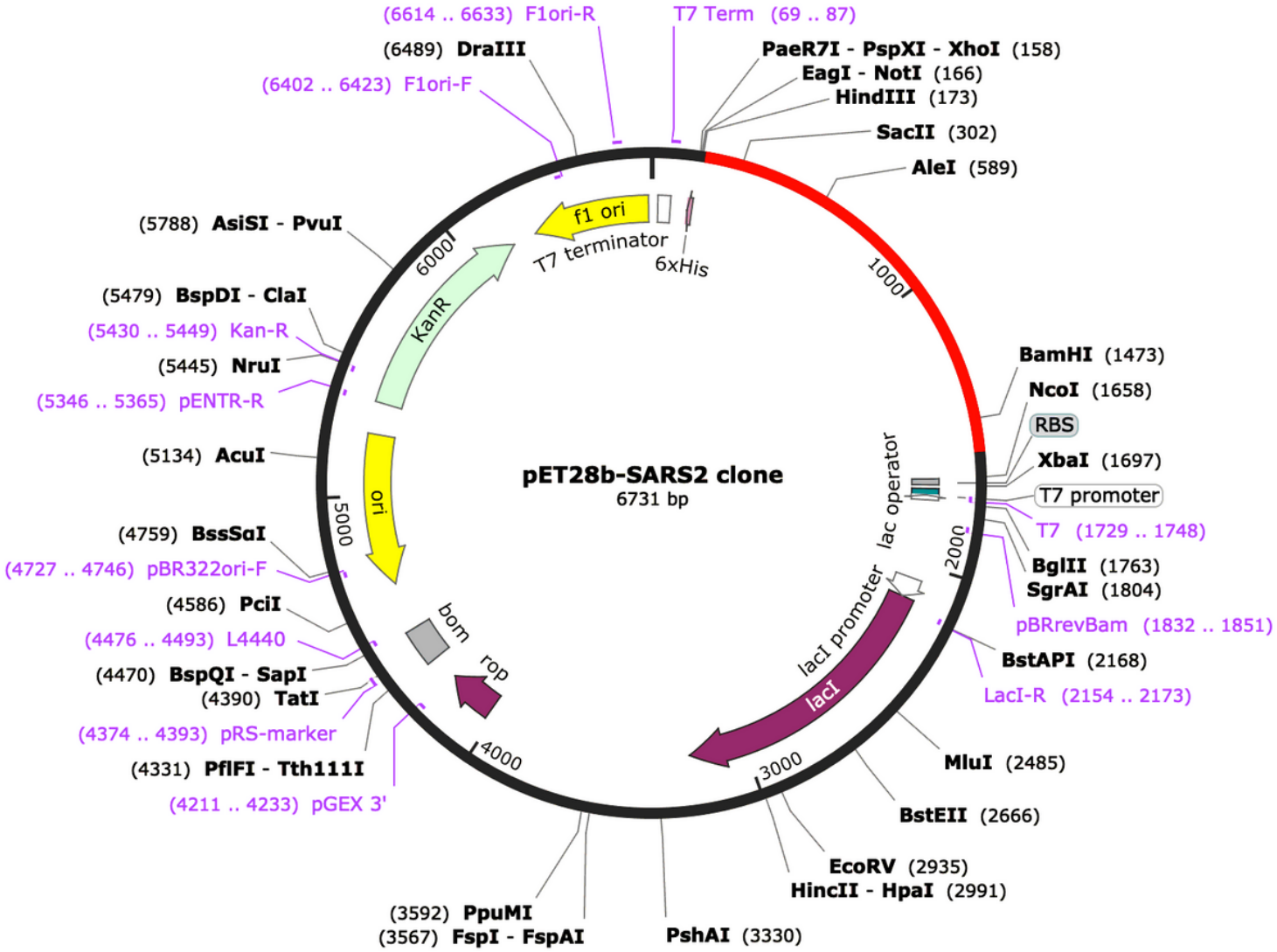

Figure 7

The DNA sequence produced by Jcat showed a GC content of 56% and a codon adaptation index of 1.0, which indicate a stable DNA sequence and a high level of protein expression. Figure 7.