Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Towards a novel peptide vaccine for Middle East respiratory syndrome coronavirus and its possible use against pandemic COVID-19

Salman Khan a,1, Bilal Shaker b,1, Sajjad Ahmad c,*,1, Sumra Wajid Abbasi d,*, Muhammad Arshad e, Abdul Haleem c, Saba Ismail c, Anita Zaib c, Wasim Sajjad b,*

a School of Life Sciences, Lanzhou University, Lanzhou, Gansu 73000, PR China
b School of Integrative Engineering, Chung ANG University, Seoul, South Korea
c National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan
d Department of Biological Sciences, National University of Medical Sciences, the Mall, Rawalpindi 46000, Pakistan
e School of Life Sciences, Lanzhou University, Lanzhou, Gansu 73000, PR China

A R T I C L E   I N F O

Article history:
Received 12 July 2020
Received in revised form 5 September 2020
Accepted 2 November 2020
Available online 6 November 2020

Keywords:
Middle East respiratory syndrome coronavirus
Middle East respiratory syndrome
Multi-epitope peptide
Toll-like receptor 3

A B S T R A C T

Middle East respiratory syndrome coronavirus (MERS-CoV) is an emerging health concern due to its high mortality rate of 35%. At present, no vaccine is available to protect against MERS-CoV infections. Therefore, an in silico search for potential antigenic epitopes in the non-redundant proteome of MERS-CoV was performed herein. First, a subtractive proteome-based approach was employed to look for the surface exposed and host non-homologous proteins. Following, immunoinformatics analysis was performed to predict antigenic B and T cell epitopes that were used in the design of a multi-epitopes peptide. Molecular docking study was carried out to predict vaccine construct affinity of binding to Toll-like receptor 3 (TLR3) and understand its binding conformation to extract ideas about its processing by the host immune system. We identified membrane protein, envelope small membrane protein, non-structural protein ORF3, non-structural protein ORF5, and spike glycoprotein as potential candidates for subunit vaccine designing. The designed multi-epitope peptide then linked to β-defensin adjuvant is showing high antigenicity. Further, the sequence of the designed vaccine construct is optimized for maximum expression in the Escherichia coli expression system. A rich pattern of hydrogen and hydrophobic interactions of the construct was observed with the TLR3 allowing stable binding of the construct at the docked site as predicted by the molecular dynamics simulation and MM-PBSA binding energies. We expect that the panel of subunit vaccine candidates and the designed vaccine construct could be highly effective in immunizing populations from infections caused by MERS-CoV and could possibly applied on the current pandemic COVID-19.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Middle East respiratory syndrome caused by a novel coronavirus, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), is a highly fatal viral respiratory tract infection in humans and is a global public health threat [1,2]. MERS-CoV was first described on June 13, 2012 in a patient from Dr. Soliman Fakeeh Hospital in Jeddah, Kingdom of Saudi Arabia [3,4]. The patient died 11 days after the virus detection. Since then, the virus was being reported frequently from across the Arabian Peninsula [4,5]. Furthermore, the spread of MERS-CoV to the countries of Middle East and other continents has brought international attention to the severe implications of the MERS pandemic [1]. Till to date, a total of 2279 laboratory-confirmed cases of MERS are reported worldwide according to the World Health Organization (WHO). This include 806 deaths with fatality rate of 35.3%. The bulk of these case occurred in Saudi Arabia: 1901 cases with death troll climbed to 732 at fatality rate of 38.5% (https://www.who.int/emergencies/mers-cov/en/). Dromedary camels (Camelus dromedaries) are implicated as a primary reservoir for the spillover of infection to the humans [3]. Person-to-person transmission of the virus have been shown to occur frequently especially in the hospital settings [6]. Direct contact with the camel or consumption of its products appear to be the main route for camel-to-human transmission, while prolonged contact among human hosts leads to person-to-person transmission [7]. Both camel-to-human and person-to-person transmission have been well documented in Saudi Arabia [5]. In South Korea, transmission of the diseases is entirely human-to-human contact [8]. In between 2012 and 2016, small outbreaks and sporadic episodes of the diseases have been reported across the Arabian Peninsula and were detected in humans from Bahrain, Kuwait, Oman, Qatar, the most notorious outbreak was seen in 2014 with more than 500 confirmed cases in the province of

https://doi.org/10.1016/j.molliq.2020.114706
0167-7322 © 2020 Elsevier B.V. All rights reserved.
Riyadh and Makkah [9]. Till to date, cases of MERS-CoV are continuously reported from Saudi Arabia every month and that’s why this country is known as critical global epicenter for MERS-CoV [10].

A range of therapeutics are investigated for MERS-CoV but still no approved treatment is available [11]. Development of a vaccine is a cost effective prophylactic measure and remains the most scalable [12]. In this regard, several candidates have been evaluated using variety of different approaches: however, still vaccine for MERS-CoV is not available [13]. Initially, vaccine development studies were hampered due to the absence of an appropriate small animal model of the MERS-CoV disease [14]. Barriers like these have been overcome in the recent past by developing in vivo approaches that allow and facilitate the testing vaccine candidates for MERS-CoV in small animal models [14]. Subunit vaccines that contains one or more pathogen immunogenic component are immunologically focused and have gained popularity because of reduced risk in vivo and relative ease of production [15]. The virus proteome comprises several vital vaccine targets that contribute in infection and pathogenesis of MERS-CoV [16]. The receptor-binding domain of spike (S) glycoprotein in particular mediate virus interaction with the host cell [15]. The envelope E protein plays a significant role in host cell recognition [17]. The interferon (IFN)-antagonizing properties of membrane protein aid in reducing IFN levels in infected cells [18]. Similarly, proteins from the open reading frame (ORF) group like ORF3, ORF5 and ORF4a are important from viral infection and pathogenesis point of view and mutational studies revealed that mutation in these genes leads to abrupt inflammation, host cell processes disruption and dysregulation IFN pathway activation [6]. The use of these single subunit vaccine in vaccine formulation is also not an appropriate choice to have as they harbor several different antigenic determinants, all of which are not desired and some may have detrimental effects on protective immunity induction [19]. This rationale has led to create an interest in designing a peptide vaccine for MERS-CoV that comprise of epitopes required for desired cellular and humoral immune response [20]. Peptide vaccines are considered sufficient enough to induce appropriate T and B-cell immunity, while at the same time excluding the risk of reactogenic and allergenic responses [21]. Moreover, peptide vaccine can be implicated for induction of broad-spectrum immunity against multiple variant of the pathogen [22]. On the contrary, because of the small size peptide vaccine is weakly immunogenic and require a carrier molecule for adjuvanting and chemical stability [19]. Several peptide vaccines are under process of development for different viral diseases, such as hepatitis C virus (HCV) [23], human papilloma virus (HPV) [24], human immunodeficiency virus (HIV) [25], and influenza [26] etc. A multi-epitope vaccine contains a series of overlapping epitopes that can elicit effective cytotoxic T cells, T helper cells and B cells responses against viral pathogen [27]. Unlike, subunit or single epitope-based vaccines, multi-epitope vaccines have a distinctive concept design with additional properties. These properties include: (i) the chances of adverse effects or pathological responses are less because of reduced unwanted components, the introduction of adjuvant capacity enhances chances of long-lasting and immunogenicity responses, (ii) the final formation has multiple epitope the spectra of targeted viruses is expanded, (iv) they are capable of producing strong cellular and humoral immunity simultaneously, (v) the epitopes in multi-epitope peptide can be recognized by T cell receptors (TCRs) of various T-cell subsets [19].

Keeping in view the broad applicability of multi-epitope vaccine, in this work we explored the complete proteome of MERS-CoV exoproteome and secretome for cytotoxic T lymphocyte (CTL) and helper T lymphocyte (HTL) epitopes containing B-cell epitopes. Additionally, molecular docking, molecular dynamics simulation and MM-PBSA binding free energies were performed to underpin multi-epitope peptide binding pose and interactions with TLR3 receptor proteins.

2. Materials and methods

2.1. Retrieval of MERS-CoV proteome and subtractive proteomics

The complete reference proteome of MERS-CoV (Proteome ID: UP000139997) was retrieved from Swiss-Prot reviewed universal protein knowledgebase database (Uniprot) [28] and subjected to subtractive proteomics pipeline for identification of potential vaccine candidates [29]. First in the process, proteins subcellular localization was predicted using an online tools: CELLO2GO [30] tool with E-value set at 0.001 and Virus-PLoc server (http://www.csbio.sjtu.edu.cn/bioinf/virus/). Proteins having localization in extracellular matrix and membrane were considered and analyzed further along the framework. All the screened proteins were evaluated for redundancy in Cluster Database at High Identity with Tolerance (CD-Hit) with sequence identity threshold set at 50%. [31] Next, comparative homology analysis of the filtered proteins with the human proteome (H. sapiens, taxid: 9606) was performed with parameters of E-value: 0.005, bit score: ≥100 and sequence identity: ≤50% in online BLASTp tool of the National Center for Biotechnological Information (NCBI) [32,33]. To avoid any homology of the protein sequence with the beneficial bacteria of human gut, a subsequent BLASTp was run against Lactobacillus rhamnosus (taxid: 47715) taking the identity sequence cut-off ≤30%, bit score of 100 and E-value of 0.005 [34,35].

2.2. Prediction of epitopes

The CTL epitopes for the proteins were predicted using NetCTL 1.2 server (http://www.cbs.dtu.dk/services/NetCTL/). NetCTL 1.2 server combines prediction of C-terminal cleavage, binding affinity for reference set of MHC class I which is restricted to 12 MHC class I super type, and TAP transport efficiency. The prediction of MHC peptide binding, proteasome cleavage event, TAP transport efficiency is done using neural networks, NetChop neural networks, and weight matrix based methods, respectively. Predicted CTL epitopes with overall combine score of 1 were selected. CTL epitopes from the pathogen proteins will elicit cell mediated immunity for inhibiting the growth and development of the pathogen and produce memory cells for remembering the pathogen in future encounters [36,37]. The predicted CTL epitopes were then subjected to Vaxijen to evaluate antigenicity of the epitopes [38]. Those with values greater >0.4 were considered antigenic and considered only. The response from Helper T-lymphocyte (HTL) epitopes constitute the major part of cellular immunity and aid in producing various cytokines and immune cell to clear the pathogen [39]. The HTL epitopes induce both CTL and HTL immune responses by secreting different lymphokines and could be crucial part of prophylactic and immunotherapeutic vaccine [40]. HTL epitopes interacting with a reference set of HTL alleles were predicted using IEDB MHC-II epitope prediction module keeping the available parameters as default [41]. All predicted epitopes further subjected to allergenicity, toxicity and IFN-gamma inducing peptides prediction by using online AllergenFP v.1.0 [42], ToxinPred [43], and IFNepitope server [44], respectively. The predicted epitopes were ranked based percentile rank score: lower score describe higher affinity binding of the epitope for HTL binding. The predicted HTL epitopes for its affinity to activate T1 type immune response were cross-validated using IFN epitope server [44].

2.3. Multi-epitope peptide vaccine construction

The shortlisted conserved epitopes were then linked to each other through a flexible linker AAY, β-defensin was used as an adjuvant to recruit naïve T-cells and immature dendritic cells to the infection site [45]. The CTL-epitopes peptide is linked to β-defensin at the N-terminal using flexible linker AAY.
EAAAK linker while the CTL and HTL epitopes were linked to each other AAY and GPAGG linker, respectively. The structure of the construct was predicted using Raptor X [46] and ab Ignition I-tasser [47]. The physicochemical parameters of the vaccine construct were evaluated using ProtParam server [48]. The generated structures of the multi-epitope peptide were then utilized in PDBsum to create Ramachandran plot for the structures [49]. The antigenicity and allergenicity were predicted by employing Vaxijen [38] and AlgPred servers [50], respectively.

2.4. B-cell epitopes assessment

Identification of B-cell epitopes is vital in vaccine designing considering the fact that it leads to production of B lymphocytes that differentiate into antibody producing plasma cells and memory cells [51]. B-cell epitopes of 20 amino acids long in the input protein sequences were identified using BCPRED [52]. The epitope specificity threshold was allowed as default by 75%.

2.5. Complementary DNA analysis for cloning and expression

Reverse translational analysis was carried out using Codon Usage Wrangler Tool (https://www.mrc-lmb.cam.ac.uk/ms/methods/codon.html) to generate optimized cDNA. The cDNA sequence was then used in GeneScript to compute GC content and codon adaptation index (CAI) score. Ideally, the GC content of a sequence should be in between 30% to 70%, while the CAI score of 1 is considered good.

2.6. Molecular docking with TLR3

Multi-epitope peptide vaccine sequence was then used in molecular docking studies in PatchDock [53]. TLR3 (PDB ID, 1ZIW) was used as receptor proteins in the process. The best predate complexes were subsequently subjected to FireDock for refinement [54]. The affinity of the vaccine construct for TLR3 was also validated by ClusPro 2.0 docking software [55].

2.7. In silico immune simulation of the vaccine construct

The vaccine construct immunogenic potential was elucidated using C-immune server [56]. The simulation parameters were treated as: 10 number of antigen dose, random seed (12345), volume of simulation 100, number of simulation steps,1000, HLA alleles used (MHC class I A0101 allele, B MHC class I B0702, DR MHC class II DRB1_0101 allele), and time step of injection employed is 1.

2.8. Molecular dynamics simulation

MD simulation was run for 120 ns on the most suitable docking solution of TLR3-MEPVC. AMBER16 was used to execute this MD simulation [57]. The ff14SB force field was used to parameterize the complex [58]. The complex was solvated into a TIP3P water box. Na + ions were added to the system to counterbalance its charge. Hydrogen atoms of the system were minimized for 500 steps, solvation box and carbon alpha atoms for 1000 steps, and all non-heavy atoms for 300 steps. Later on, in an attempt to manage the system temperature, it was heated to 300 K (NVT) for 20 ps through langevin dynamics [59] where a 2 femtosecond time step long restraint of 5 kcal/mol-A2 on carbon alpha atoms was permitted. System was then unrestrained for 100 ps while being equilibrated. NPT ensemble was put in practice for 50 ps to manage the pressure on the system. Finally, simulation run of 120 ns was achieved at a time step of 2 femtoseconds. AMBER CPTRAJ [60] was then used to examine the trajectories for to evaluate complex dynamics and stability.

2.9. Calculation of binding free energies

MMPSBA [55].py package was used to predict binding free energies of the system [61]. A total of 100 frames were derived from simulation trajectories to be employed in MMPSBA (Molecular Mechanics Generalized Born Surface Area) and MMPSB (Molecular Mechanics Poisson–Boltzmann Surface Area) [62]. The examination of the free energy difference between solvated phase and gas phase of the system is the ultimate goal of this assay. The determination of net binding free energy was carried out as,

\[ \Delta G_{\text{bind}} = \Delta G_{\text{bind, vacuum}} + \Delta G_{\text{solv, vacuum}} - \Delta G_{\text{bind, solv}} - \Delta G_{\text{solv, solv}} + \Delta G_{\text{TLR3}} \]

The binding free energy calculation was done on 100 frames extracted from complete length of MD simulation as well as from last 20-ns to thoroughly validate intermolecular affinity and stability. Entropy of the system was estimated using a bash script following the protocol provided by Duan et al., 2016 [63].

3. Results and discussion

3.1. Subtractive proteomics

The complete MERS-CoV proteome contains 13 proteins in total, namely: Replicate polypeptide 1ab, Replicas polypeptide 1a, Membrane protein, Envelope small membrane protein, Non-structural protein ORF4b, Non-structural protein ORF3, Non-structural protein ORF5, Non-structural protein ORF4a, Nucleoprotein, and Spike glycoprotein. These proteins were first investigated in subcellular localization of subtractive proteomics to screen only those candidates that are exposed for direct interaction with the host cells. Such proteins identification represent an important goal as they are involved in virulence and pathogenesis and their vital importance from stimulating immune responses [64]. Only proteins that are localized to either extracellular (exoprotein) or outer membrane (secretome) were opted in subcellular localization filter. Five proteins: Membrane protein, Envelope small membrane protein, Non-structural protein ORF3, Non-structural protein ORF5, Spike glycoprotein were shortlisted in this analysis as highlighted in bold in Table 1. Next in the framework, these proteins were subjected non-redundancy check. In this, non-redundant dataset was only considered. All the five proteins mentioned above were non-redundant which means that each protein represented in the dataset is only singly repeated [35]. The comparative homology check with the humans revealed no hit, indicating highly non-similarity between these virus proteins and the human proteome. This is quite vital in avoiding the auto-immune responses and designing a safe vaccine [65,66]. Lastly, a comparative homology analysis between the MERS-CoV shortlisted proteins and the beneficial gut bacterium, Lactobacillus rhamnosus, was performed to eliminate the risk of accidental inhibition of good bacteria in the host. Bacteria of such type play profound role in improving functions of the host immune system, aid in digestion of food of good bacteria in the host. Bacteria of such type play profound role in for extracting vital nutrients, and prevent the growth of hostile infection causing bacteria etc. [34]. All the five proteins were noticed to show no homology to the said representative gut bacterium proteome allowing the safe use of the filtered proteins epitopes.

3.2. Prediction of cytotoxic T-lymphocyte epitopes

Cytotoxic T cell lymphocytes (also known as Killer T cells) are important component of immune system for tumor surveillance and against intracellular pathogens including bacteria, protozoan and viruses [36]. When these cells recognized its antigen, it become activated and follow three step mechanisms to kill the infected or malignant cells [67]. First, it secrete cytokines primarily IFN–γ and TNF–α which have antiviral and antitumor effects [68]. Secondly, it produces and release cytokiotic
granules. These granules have perforin that make the target cell membrane porous and granymes which being serine proteases hydrolyze intracellular proteins, block viral proteins synthesis and eventually leads to apoptosis [67]. All the five proteins shortlisted in the previous

Table 2

| Protein ID | Protein | CELLDECO (Prediction score) | Virus-PLoc |
|------------|---------|-----------------------------|------------|
| sp[K9N7C7]| Replicate polyprotein 1ab | Plasma membrane (2.35) | Plasma membrane |
| >sp[K9N638]| Replicate polyprotein 1a | Plasma membrane (2.54) | Nucleus |
| >sp[K9N7A1]| Membrane protein | Plasma membrane (4.95) | Plasma membrane |
| >sp[K9N83]| Envelope small membrane protein | Plasma membrane (3.33) | Plasma membrane |
| >sp[K9N643]| Non-structural protein ORF4b | Mitochondrial (1.63) | Plasma membrane |
| >sp[K9N796]| Non-structural protein ORF5 | Extracellular (3.65) | Plasma membrane |
| >sp[K9N720]| Non-structural protein ORF5 | Plasma membrane (4.94) | Plasma membrane |
| >sp[K9N405]| Non-structural protein ORF4a | Extracellular (2.21) | Cytoplasm |
| >sp[K9N477]| Nucleoprotein | Nucleus (3.07) | Nucleus |
| >sp[K9N563]| Spike glycoprotein | Plasma membrane (3.31) | Plasma membrane |

Table 1

Proteins that constitute the proteome of the MERS-CoV. Bold highlighted proteins are the one selected in subcellular localization analysis.

Table 2

CTL epitopes for shortlisted 5 proteins with antigenicity scores.

| Protein | Peptide | VaxiJen Score | Result | Allergenicity | Toxicity |
|---------|---------|---------------|--------|---------------|----------|
| Membrane protein | LLTIVLQY | 0.8719 | Probable ANTIGEN | Probable non-allergen | −1.09 (Non-Toxin) |
| | ITLYQLYC | 1.2147 | Probable ANTIGEN | Probable non-allergen | −0.85 (Non-Toxin) |
| | ALISIGAVY | 0.1615 | Probable NON-ANTIGEN | Probable non-allergen | −1.17 (Non-Toxin) |
| | VSAMIMISY | 0.6059 | Probable NON-ANTIGEN | Probable non-allergen | −0.79 (Non-Toxin) |
| | GTNGSVAIY | 0.4620 | Probable ANTIGEN | Probable non-allergen | −1.16 (Non-Toxin) |
| | TYITIVQLY | 0.2772 | Probable NON-ANTIGEN | Probable non-allergen | −1.23 (Non-Toxin) |
| | LVQPALLY | 0.0512 | Probable NON-ANTIGEN | Probable non-allergen | −0.87 (Non-Toxin) |
| Envelope small membrane protein | MTGFMNLLV | 0.0861 | Probable NON-ANTIGEN | Probable non-allergen | −1.41 (Non-Toxin) |
| | TLLVQALNY | 0.2681 | Probable NON-ANTIGEN | Probable non-allergen | −1.22 (Non-Toxin) |
| Non-structural protein ORF3 | TSTSVTVY | 0.3229 | Probable NON-ANTIGEN | Probable non-allergen | −0.91 (Non-Toxin) |
| | FYIPIASCY | 0.1568 | Probable NON-ANTIGEN | Probable non-allergen | −0.89 (Non-Toxin) |
| | MVLLOYLVL | 0.1428 | Probable NON-ANTIGEN | Probable non-allergen | −0.81 (Non-Toxin) |
| | SITYCETY | 0.8016 | Probable ANTIGEN | Probable non-allergen | −0.30 (Non-Toxin) |
| Spike glycoprotein | FLITICTSY | 0.2812 | Probable NON-ANTIGEN | Probable non-allergen | −1.38 (Non-Toxin) |
| | EVDIQTTFF | 0.8992 | Probable NON-ANTIGEN | Probable non-allergen | −1.29 (Non-Toxin) |
| | VSKADDITY | 0.1659 | Probable NON-ANTIGEN | Probable non-allergen | −1.19 (Non-Toxin) |
| | TYSNITYT | 1.5697 | Probable ANTIGEN | Probable non-allergen | −1.10 (Non-Toxin) |
| | ITYQQLFPY | 0.9146 | Probable ANTIGEN | Probable non-allergen | −0.89 (Non-Toxin) |
| | YQDGVHCDMY | 0.2627 | Probable NON-ANTIGEN | Probable non-allergen | −0.73 (Non-Toxin) |
| | TSATLRKCV | 0.1760 | Probable NON-ANTIGEN | Probable non-allergen | −1.32 (Non-Toxin) |
| | SFDGKVGRF | 1.0567 | Probable NON-ANTIGEN | Probable non-allergen | −0.24 (Non-Toxin) |
| | NSYTSFATY | 0.2171 | Probable NON-ANTIGEN | Probable non-allergen | −0.96 (Non-Toxin) |
| | ATDCDDNY | 0.7838 | Probable ANTIGEN | Probable non-allergen | −0.19 (Non-Toxin) |
| | ASLNSFKEY | −0.133 | Probable NON-ANTIGEN | Probable non-allergen | −0.50 (Non-Toxin) |
| | ITDSCWILE | 0.1074 | Probable NON-ANTIGEN | Probable non-allergen | −1.21 (Non-Toxin) |
| | QTGQGHVHI | 0.3715 | Probable NON-ANTIGEN | Probable non-allergen | −0.98 (Non-Toxin) |
| | GVVHFSSRY | 0.1530 | Probable NON-ANTIGEN | Probable non-allergen | −0.54 (Non-Toxin) |
| | FSRRYVVDL | 1.0488 | Probable ANTIGEN | Probable non-allergen | −0.44 (Non-Toxin) |
| | YVDSLQCMNM | 0.2930 | Probable ANTIGEN | Probable non-allergen | −0.81 (Non-Toxin) |
| | QSDQKAQWA | 0.2840 | Probable NON-ANTIGEN | Probable non-allergen | −0.87 (Non-Toxin) |
| | LLDVFSDCG | 0.0436 | Probable NON-ANTIGEN | Probable non-allergen | −0.95 (Non-Toxin) |
| | DLQSJCHSY | 0.9444 | Probable ANTIGEN | Probable non-allergen | −0.83 (Non-Toxin) |
| | SDVQEGSVY | 0.9858 | Probable ANTIGEN | Probable non-allergen | −1.33 (Non-Toxin) |
| | LGTTIPQVY | 0.2591 | Probable NON-ANTIGEN | Probable non-allergen | −0.96 (Non-Toxin) |
| | PAAASNCY | 0.5141 | Probable NON-ANTIGEN | Probable non-allergen | −0.61 (Non-Toxin) |
| | CYSSLILDY | 0.4038 | Probable ANTIGEN | Probable non-allergen | −0.65 (Non-Toxin) |
| | YSSULDDY | 0.0327 | Probable NON-ANTIGEN | Probable non-allergen | −1.03 (Non-Toxin) |
| | TITTKPLY | 0.1377 | Probable NON-ANTIGEN | Probable non-allergen | −1.48 (Non-Toxin) |
| | ITKPKGDFY | 0.6931 | Probable NON-ANTIGEN | Probable non-allergen | −1.77 (Non-Toxin) |
| | NTYGQDFGM | −0.0613 | Probable NON-ANTIGEN | Probable non-allergen | −0.73 (Non-Toxin) |
| | MTQEQMQMF | 1.2508 | Probable ANTIGEN | Probable non-allergen | −1.40 (Non-Toxin) |
| | YYSTDGNYY | −0.0896 | Probable NON-ANTIGEN | Probable non-allergen | −0.68 (Non-Toxin) |
| | YSDCDGNYC | 0.1491 | Probable NON-ANTIGEN | Probable non-allergen | −0.52 (Non-Toxin) |
| | VSNPPSVFY | 0.0738 | Probable NON-ANTIGEN | Probable non-allergen | −1.15 (Non-Toxin) |
| | HISTSMQY | 0.8394 | Probable ANTIGEN | Probable non-allergen | −0.92 (Non-Toxin) |
| | QVDQLNSSY | 0.4655 | Probable ANTIGEN | Probable non-allergen | −0.68 (Non-Toxin) |
| | FSFGVQTYE | 1.6924 | Probable ANTIGEN | Probable non-allergen | −1.30 (Non-Toxin) |
| | RSAALDLF | 0.1712 | Probable NON-ANTIGEN | Probable non-allergen | −1.01 (Non-Toxin) |
| | VNAPNGLY | 0.0211 | Probable NON-ANTIGEN | Probable non-allergen | −0.68 (Non-Toxin) |
| | WSSYGCFSY | 1.0625 | Probable NON-ANTIGEN | Probable non-allergen | −1.27 (Non-Toxin) |
| | NTTILDDY | 1.5316 | Probable ANTIGEN | Probable non-allergen | −1.40 (Non-Toxin) |
### Table 3
HTL epitopes for shortlisted 5 proteins with antigenicity scores.

| Protein                          | Start amino acid | End amino acid | Peptide                          | Percentile score | Vaxijen antigenicity | Allergenicity | Toxicity |
|----------------------------------|------------------|----------------|----------------------------------|------------------|----------------------|---------------|----------|
| Membrane protein                 | 16-30            | 30             | SQIISGIVAAVSAMMM                | 3.02             | 0.1060 (Probable non-antigen) | Probable allergen | −1.13 (Non-Toxin) |
|                                  | 17-31            | 31             | QIISGIVAAVSAMMW                 | 3.02             | 0.1324 (Probable non-antigen), non-allergen | Probable | −1.07 (Non-Toxin) |
|                                  | 18-32            | 32             | ISSGIVAAVSAMMWI                | 4.57             | 0.1871 (Probable non-antigen), non-allergen | Probable allergen | −0.73 (Non-Toxin) |
| Envelope small membrane protein  | 43-44            | 44             | MTGFNTLQVQAPYLY                | 0.32             | 0.2107 (Probable non-antigen) | Probable | −1.52 (Non-Toxin) |
|                                  | 43-45            | 45             | CMICGFNTLLQVAPLY                | 0.71             | 0.1828 (Probable non-antigen) | Probable | −1.20 (Non-Toxin) |
|                                  | 43-45            | 45             | TGFNTLQVQAPYLY                  | 0.88             | 0.1509 (Probable non-antigen) | Probable | −1.15 (Non-Toxin) |
|                                  | 43-45            | 45             | QCMICGFNTLLQVAPY                | 1.24             | 0.0771 (Probable non-antigen) | Probable | −1.28 (Non-Toxin) |
|                                  | 43-45            | 45             | GFNTLQVQAPYLY                  | 1.71             | 0.1781 (Probable non-antigen) | Probable | −1.26 (Non-Toxin) |
|                                  | 46-50            | 50             | GFNTLQVQAPYLY                  | 1.71             | 0.1781 (Probable non-antigen) | Probable | −1.26 (Non-Toxin) |
|                                  | 46-50            | 50             | GFNTLQVQAPYLY                  | 1.71             | 0.1781 (Probable non-antigen) | Probable | −1.26 (Non-Toxin) |
| Non-structural protein ORF3      | 5-9              | 9              | RPPTLLLVFSLSLLV                 | 1.36             | 0.4001 (Probable non-antigen) | Probable | −1.12 (Non-Toxin) |
|                                  | 5-9              | 9              | RPPTLLLVFSLSLLV                 | 1.36             | 0.4001 (Probable non-antigen) | Probable | −1.12 (Non-Toxin) |
|                                  | 5-9              | 9              | RPPTLLLVFSLSLLV                 | 1.36             | 0.4001 (Probable non-antigen) | Probable | −1.12 (Non-Toxin) |
|                                  | 5-9              | 9              | RPPTLLLVFSLSLLV                 | 1.36             | 0.4001 (Probable non-antigen) | Probable | −1.12 (Non-Toxin) |
|                                  | 5-9              | 9              | RPPTLLLVFSLSLLV                 | 1.36             | 0.4001 (Probable non-antigen) | Probable | −1.12 (Non-Toxin) |
| Non-structural protein ORF5      | 11-15             | 15             | LVFLVLYNLFAVLV                  | 0.04             | 0.4775 (Probable antigen) | Probable | −1.11 (Non-Toxin) |
|                                  | 11-15             | 15             | LVFLVLYNLFAVLV                  | 0.04             | 0.4775 (Probable antigen) | Probable | −1.11 (Non-Toxin) |
|                                  | 11-15             | 15             | LVFLVLYNLFAVLV                  | 0.04             | 0.4775 (Probable antigen) | Probable | −1.11 (Non-Toxin) |
|                                  | 11-15             | 15             | LVFLVLYNLFAVLV                  | 0.04             | 0.4775 (Probable antigen) | Probable | −1.11 (Non-Toxin) |
|                                  | 11-15             | 15             | LVFLVLYNLFAVLV                  | 0.04             | 0.4775 (Probable antigen) | Probable | −1.11 (Non-Toxin) |

(continued on next page)
| Protein | Start amino acid | End amino acid | Peptide | Percentile score | Vaxijen antigenicity | Allergenicity | Toxicity |
|---------|-----------------|----------------|---------|------------------|---------------------|---------------|----------|
|         |                 |                |         |                  | Probable non-antigen.| Probable allergen|          |
| 27      | 41              |                | DSVFTYIPASCYVA | 2.27 | 0.1703 | Probable allergen | −1.06 (Non-Toxin) |
| 9       | 23              |                | RSHIRVSTVSSHGM | 2.51 | 0.1042 | Probable non-allergen | −1.11 (Non-Toxin) |
| 10      | 24              |                | SHIRVSTVSSHGMV | 2.51 | 0.1973 | Probable non-allergen | −1.43 (Non-Toxin) |
| 8       | 22              |                | MRSHRVSTVSSHGM | 2.74 | 0.0550 | Probable non-allergen | −1.26 (Non-Toxin) |
| 7       | 21              |                | DMRSHRVSTVSSH | 3.36 | 0.3701 | Probable non-allergen | −1.16 (Non-Toxin) |
| 31      | 45              |                | FYIPASCGYVAAALAV | 3.49 | 0.2843 | Probable non-allergen | −1.16 (Non-Toxin) |
| 29      | 43              |                | RDSVFHLIAPSLI | 3.88 | 0.2191 | Probable non-allergen | −1.07 (Non-Toxin) |
| 30      | 44              |                | VFTYIPASCGYVAAALA | 3.95 | −0.0681 | Probable non-allergen | −1.13 (Non-Toxin) |
| 26      | 40              |                | TDSVFTHIPASGVY | 4.16 | 0.0533 | Probable allergen | −0.97 (Non-Toxin) |
| 25      | 39              |                | STDSTVFTHIPASGVY | 4.72 | 0.0749 | Probable non-allergen | −0.89 (Non-Toxin) |
| 14      | 28              |                | QSIYRVLNGVGITQQ | 0.39 | 0.7573 | Probable non-allergen | −2.15 (Non-Toxin) |
| 15      | 29              |                | SIYRVLNGVGITQQV | 0.62 | 0.7288 | Probable non-allergen | −2.30 (Non-Toxin) |
| 4       | 18              |                | QQLRVSESAALSQ | 0.62 | 0.3902 | Probable non-allergen | −1.16 (Non-Toxin) |
| 43      | 57              |                | IVSFVVNAPNGLYFM | 0.71 | 0.1671 | Probable non-allergen | −0.39 (Non-Toxin) |
| 38      | 52              |                | YIWLGFIAGLVALAL | 0.77 | 0.7583 | Probable non-allergen | −0.79 (Non-Toxin) |
| 40      | 54              |                | WLFIAGLVALALCV | 0.77 | 0.4240 | Probable non-allergen | −0.74 (Non-Toxin) |
| 41      | 55              |                | LGFAIWLVALALCF | 0.77 | 0.4118 | Probable non-allergen | −0.84 (Non-Toxin) |
| 13      | 27              |                | AQSIYRVLNGVGITQ | 0.79 | 0.6283 | Probable non-allergen | −1.97 (Non-Toxin) |
| 46      | 60              |                | VYTLIKYYSIIPSI | 0.88 | 0.2157 | Probable non-allergen | −1.09 (Non-Toxin) |
| 12      | 26              |                | FASQSIYRVLNGVGIT | 0.96 | 0.4940 | Probable non-allergen | −1.77 (Non-Toxin) |
| 39      | 53              |                | WLFGIAGLVALALCV | 0.96 | 0.4564 | Probable non-allergen | −0.77 (Non-Toxin) |
| 16      | 30              |                | IFYRLNGVGITQQVSL | 1.15 | 0.6444 | Probable non-allergen | −2.29 (Non-Toxin) |
| 42      | 56              |                | GFIAGLVALALCVFF | 1.15 | 0.2054 | Probable non-allergen | −0.59 (Non-Toxin) |
| 5       | 19              |                | QVRVSESAALSQAQLA | 1.33 | 0.5139 | Probable non-allergen | −1.22 (Non-Toxin) |
| 11      | 25              |                | PFAQSIYRVLNGVGI | 1.36 | 0.4443 | Probable non-allergen | −1.66 (Non-Toxin) |
| 44      | 58              |                | VSFVNVAPNGLYFMH | 1.71 | 0.4899 | Probable non-allergen | −0.43 (Non-Toxin) |
| 3       | 17              |                | AQQLRVSESAALSQAQLA | 1.81 | 0.4991 | Probable non-allergen | −0.91 (Non-Toxin) |
| 6       | 20              |                | LVRSEASAALSQAALK | 1.9 | 0.4473 | Probable non-allergen | −1.02 (Non-Toxin) |
| 42      | 56              |                | HVSFVNVAPNGLYFY | 2.18 | 0.1797 | Probable non-allergen | −0.67 (Non-Toxin) |
| 37      | 51              |                | WYILGFIAGLVALALA | 2.27 | 0.6569 | Probable non-allergen | −0.72 (Non-Toxin) |
| 17      | 31              |                | FYRLNGVGITQQVSL | 2.55 | 0.6337 | Probable non-allergen | −2.25 (Non-Toxin) |
| 28      | 42              |                | DFNLTLIEPSISTG | 2.91 | 1.5583 | Probable non-allergen | −1.73 (Non-Toxin) |
| 2       | 16              |                | VAQQVRVSESAALS | 3.09 | 0.3358 | Probable non-allergen | −1.11 (Non-Toxin) |
| 21      | 35              |                | HGDMYVYSAGHATGTT | 3.49 | −0.1383 | Probable non-allergen | −1.27 (Non-Toxin) |
| 22      | 36              |                | GDMMYVYSAGHATGTT | 3.49 | 0.0241 | Probable non-allergen | −0.99 (Non-Toxin) |
| 45      | 59              |                | SFVNVAPNGLYFMHV | 3.63 | 0.3855 | Probable non-allergen | −0.48 (Non-Toxin) |
| 23      | 37              |                | DMYYVYSAGHATGTP | 3.72 | 0.1080 | Probable non-allergen | −1.03 (Non-Toxin) |
step was analyzed for 9mer cytotoxic T-cell receptor epitopes. Briefly, 5 CTL epitopes were obtained for membrane protein, 3 for envelop small membrane protein, 3 for non-structural protein ORF3, 4 for non-structural protein ORF7 and 38 for the spike protein. Among these predicted epitopes, we found only 4 antigenic epitopes for membrane protein, 0 for envelop small membrane protein, 1 for non-structural protein ORF3 and ORF5, and 17 for spike protein. Only antigenic epitopes were considered while non-antigenic were discarded. The selected epitopes are also non-allergic, non-toxic and are IFN-gamma inducing peptide. Predicted CTL epitopes for the screened 5 subunit vaccine proteins are shown in Table 2.

### 3.3. Prediction of helper T-lymphocyte epitopes

T helper lymphocytes or CD4+ cells are important player of the immune system by instigating and shaping adaptive immune system. T helper lymphocytes recognize peptides displayed on the MHC class II molecules. They release cytokines that aid in activating several other immune cells. They are essential for B cell antibody class switching, enhancing bactericidal activity of macrophages, and growth and activation of cytotoxic T cells [51,69]. The five proteins were subjected to MHC-II epitope prediction of 15mer length as shown in Table 3. Only epitopes of lowest percentile score were selected and as such epitopes with percentile score < 3 were opted only. Using this criterion, we predicted 3 HTL epitopes for membrane protein, 9 for envelope small membrane protein and non-structural protein ORF3, 25 for non-structural protein ORF5 and 31 for the spike protein. Antigenicity evaluation revealed 0 epitopes for membrane protein, envelope small membrane protein and non-structural protein ORF5 4 and 3 respectively, 6 for non-structural protein ORF3 and 17 for spike proteins. The selected epitopes besides being antigenic are also non-allergic, non-toxic and are IFN-gamma inducing peptide.

### 3.4. Multi-epitope peptide construction

A multi-epitope peptide of 874 residues containing sequence conserved epitopes of both T cell immunity (CTL and HTL) and B cell immunity was designed. This design was based on the fact that individual peptide epitope is weakly immunogenic, therefore, fusing the screened epitopes to make a multi-epitope peptide could potentially enhance their immunogenicity capacity. Further, linking this peptide to an adjuvant can further increase the chance of good immune presentation [19].

---

**Table 3 (continued)**

| Protein | Start amino acid | End amino acid | Peptide | Percentile score | Vaxijen antigenicity | Allergenicity | Toxicity |
|---------|-----------------|----------------|---------|------------------|---------------------|--------------|---------|
| 9       | 23              | 23             | EGGWLVAGSTVAM  | 3.95             | 0.1129 (Probable non-antigen). | Probable non-allergen | −1.03 (Non-Toxin) |
| 10      | 24              | 24             | GGGWLVAGSTVAMT | 3.95             | 0.2673 (Probable non-antigen). | Probable non-allergen | −1.12 (Non-Toxin) |
| 43      | 57              | 57             | FIALGLVALALCVFFI | 4.16           | 0.2672 (Probable non-antigen). | Probable non-allergen | −0.60 (Non-Toxin) |
| 39      | 53              | 53             | KIFKIFQALGAMQTGFT | 4.77           | 0.3893 (Probable non-antigen). | Probable allergen | −0.79 (Non-Toxin) |

---

**Fig. 1.** Schematic presentation of the multi-epitope vaccine construct design.
In total, 53 epitopes were predicted for cytotoxic T cell immunity among which 19 epitopes were found antigenic and only considered for multi-epitope vaccine construct designing. Similarly, 77 HTL epitopes were predicted among which 30 were antigenic and considered. The CTL epitopes were joined together using AAY linkers. These linkers possess the cleavage sites for proteasome in the mammalian cells thus aiding in the production of natural epitopes and stopping the generation of junctional epitopes and by this way enhancing the presentation of the epitopes presentation to the host immune system [70]. The HTL epitopes were interconnected to each other through GPGPG linker for adequate presentation and stimulation of T cell responses. The EAAAK linker was employed to link the peptide of CTL epitopes to \( \beta \)-defensin adjuvant at the N-terminal and HTL epitopes at the C-terminal. \( \beta \)-defensin is a potent immune adjuvant and functions by stimulating the production of lymphokines which subsequently cellular immunity and production of antigen specific immunoglobulin [45]. The multi-epitope vaccine construct is schematically presented in Fig. 1.

The full sequence of the construct contains large portion of B cell epitopes indicating the construct high potential of activating both arms of the host immunity. The predicted B cell epitope region in the vaccine construct is shown in S-Fig. 1. The average value for the construct is 0.602 with maximum value of 0.713 and minimum value of 0.177. Humoral immunity is a key component of the host immune system in clearing the microbial pathogen in extracellular fluid and block the spread of intracellular infection.

### 3.5. Allergenicity and antigenicity prediction of multi-epitope peptide construct

To ensure the non-allergic responses of the vaccine construct upon administration, the allergenicity of the construct was evaluated [71]. The prediction was based on SVM that securitized the amino acids composition of the input sequence for allergic sequences. The construct sequence was declared as non-allergen with score of \(-0.80559987\)

---

**Table 4**

| Structure assessment of multi-epitope peptide vaccine construct. |
|---------------------------------------------------------------|
| **Ramachandran plot statistics**                              | **Number of residues unrefined/refined** | **Percentage unrefined/refined** |
| ProCheck analysis                                             |                                            |                                  |
| Most favored regions (A,B,L)                                  | 527/587                                   | 80.5%/ 90.3%                     |
| Additional allowed regions \([a,b,l,p]\)                       | 122/64                                    | 18.6%/ 9.8%                      |
| Generously allowed regions \([-a,-b,-l,-p]\)                  | 4/ 3                                      | 0.6%/ 0.5%                       |
| Disallowed regions \([XX]\)                                   | 2/1                                       | 0.3%/ 0.1%                       |
| Non-glycine and non-proline residues                         | 655/655                                   | 100.0%/100.0%                    |
| End residues (Excluding glycine and proline)                 | 1/1                                       |                                    |
| Glycine residues                                              | 123/123                                   |                                    |
| Proline residues                                              | 70/70                                     |                                    |
| Total number of residues                                      | 849/849                                   |                                    |
| G-factors                                                     |                                            | Average score                     |
| Phi-psi distribution                                          | \(-0.43/\)                                | \(-0.64/0.23\)                   |
| Chi1-chi2 distribution                                        | \(-0.32/\)                                |                                    |
| Chi1 only                                                     | 0.01/0.28                                 |                                    |
| Chi3 & chi4                                                   | 0.90/0.47                                 |                                    |
| Omega                                                        | 1.32/0.98                                 |                                    |
| Main-chain bond lengths                                       | 0.47/\(-0.11\)                           |                                    |
| Main-chain bond angles                                        | \(-0.13/\)                                |                                    |
| Overall average                                               |                                          |                                    |
(cut-off = −0.4). The percent of positive predictive value deduced was 0% in contrast to negative predictive value that was equal to 0%. Similarly, the antigenicity of the construct was re-evaluated to confirm its immune system evoking potential. The overall antigen prediction score for the vaccine construct is 0.4744 that indicates the probable antigenic behavior of the designed chimeric peptide.

3.6. Physiochemical properties of the multi-epitope peptide vaccine construct

The physiochemical properties of the designed construct were predicted to guide the experimentalists in synthesis and understanding the formulation of the vaccine. The molecular weight estimated for the construct is 140 kDa that explain two things: (i) first the medium size of the construct will allow its convenient purification and secondly, small and medium size chimeric proteins harbor appropriate number of epitopes for specific and targeted immune responses against the pathogen thus limiting the chances of allergic and adverse reactions [72]. The vaccines composed of whole organism or a single large protein have large number of antigenic determinants all of which are probably not required as they could lead to non-specific immunity [19]. The theoretical pl computed for the vaccine construct is 8.05 indicating its basic nature. The instability index of the protein is estimated as 27.07 and classify it as stable. This instability index tells an estimate of protein stability [73]. It has been determined that stable nature of proteins is due to the occurrence of certain dipeptides that differ significantly from non-stable. The protein instability index calculated by Protparam assigned a weightage of instability to different 400 dipeptides. Proteins with instability index of higher than 40 are categorized as unstable. The aliphatic index represents relative volume occupied by amino acid residues: leucine, isoleucine, alanine and valine [74]. It is regarded as a key factor confering thermostability to globular proteins. The aliphatic index for the designed multi-epitope vaccine is 106.10 presenting the high thermostability of the construct. The estimated half-life of the construct in mammalian cell is 140 kDa that explain two things: (i) the formulation of the vaccine. The molecular weight estimated for the multi-epitope peptide vaccine construct is 0.4744 that indicate the probable antigenic behavior of the designed protein.

3.7. Prediction of multi-epitope peptide secondary and tertiary structure

The secondary structure of the designed multi-epitope vaccine construct consists of 23 (2.7%), 338 (39.8%), 63–10 helix (0.7%), and 482 (56.8%) among other. More specifically, the remaining are composed of 3 β-sheets, 3 β-hairpins, 2 beta bulges, 6 strands, 20 helices, 22 helix-helix interactions, 51 β-turns, and 46 γ-turns. All the 874 residues of the input construct were modelled. The best structure was modelled using the A chain of the 4KNH template by Raptor-X with p-value of 3.18e-04. The 3D model of the predicted structure is shown in Fig. 2.

3.8. Ramachandran plot analysis of the multi-epitope peptide construct

Ramachandran diagram or Ramachandran plot is a method of visualizing Phi or ψ (the bonds between nitrogen and carbon alpha) and Psi or ϕ (the bond between carbon and carbon alpha) of a given polypeptide [75]. The Phi and Psi are also known as Ramachandran or dihedral angles and hold control local structural importance vital for protein folding. Ramachandran plot is a vital assessment factor for evaluating three dimensional structure of a protein. This plot can be divided into four Quadrants (from I to IV) that give important information about secondary structure of a protein. Quadrant I represent conformations that
Fig. 3. Disulfide engineering of the multi-epitope vaccine construct.

Fig. 4. Cloned vaccine construct shown in red into the expression vector.
...were found in the disallowed regions having conformation covers 122 residues (18.6%) and 4 (0.6%), respectively. Only 2 residues these regions, additionally allowed and generously allowed regions of the residues are not in correct and usual geometry. In count to protein residues are in the most favored regions. This means that majority of the protein residues i.e. 527 that make 80.5% of the total residues with potential to be mutated is listed in Table 5. The wild and mutated multi-epitope vaccine construct are provided in Fig. 3.

4.3. Disulphide engineering

Disulphide engineering is a key approach in biotechnology to increase thermal stability of protein under study and assist in understanding its dynamics [77]. The introduction of disulphide bonds in the protein is believed to lower the conformational entropy and at the same time increases the free energy of the denatured state [78]. Thus both these actions lead to overall stability of protein fold conformation. Total of 70 pair of residues were found that could be manipulated for the purpose of increasing protein stability. However, only pair of residues with combined energy of >2 kcal/mol were opted. The complete set of residues with potential to be mutated is listed in Table 5. The wild and mutated multi-epitope vaccine construct are provided in Fig. 3.

4.10. Adaptation of codon usage to Escherichia coli K 12 strain

The codon adaptation tool is a simple way to adjust the codon usage of the input sequence to the selected prokaryotic and eukaryotic system [79,80]. The difference in codon usage of the host from that organism from which the desired sequence stems from plays a major role in minor expression of that sequence. Codon adaptation of the vaccine construct primary sequence to E. coli K12 expression system is vital in ensuring its maximum expression. The codon adaptive index and GC content of the improve sequence are 0.86 and 57.75, respectively. The adaptive score of the vaccine and it’s in silico restriction cloning are provided in S-Fig. 3 and Fig. 4.

4.11. Molecular docking with TLR3

Molecular docking of the designed vaccine construct with TLR3 as a receptor was accomplished using PatchDock, which generated more than 20 complexes of TLR3 and vaccine construct conformations. The top 10 best predicted solutions were subjected to the fast interaction refinement in molecular docking (firedock) to refine the predicted solutions based the global energy. Lower the global energy of a solution implies the stable formation of complex and vice versa. As such solution 3 with global energy of −24.02 kcal/mol, attractive van der Waals energy of −38.57 kcal/mol, repulsive van der Waals energy of 28.88 kcal/mol, atomic contact energy of 11.49 kcal/mol and hydrogen bonding energy of −1268.5 kcal/mol. The docked conformation of the TLR3 and construct with respect to each is shown in Fig. 5. The top complex conformation of the TLR3-MEVC was also understand via ClusPro that complement good binding of the molecules with lowest energy of −1268.5 kcal/mol.

Table 6

| Rank | Solution number | Global energy (kcal/mol) | Attractive Van der Waals | Repulsive Van der Waals | Atomic contact energy (kcal/mol) | Hydrogen bonding | Rank | Solution number | Global energy (kcal/mol) | Attractive Van der Waals | Repulsive Van der Waals | Atomic contact energy (kcal/mol) | Hydrogen bonding |
|------|----------------|--------------------------|-------------------------|------------------------|--------------------------------|----------------|------|----------------|--------------------------|-------------------------|------------------------|--------------------------------|----------------|
| 1    | 3              | −24.02                   | −38.57                  | 28.88                  | 11.49                          | −4.00          | 2    | 7              | −18.39                   | −25.30                  | 19.27                  | −2.78                          | −1.87          |
| 3    | 10             | −8.94                    | −21.98                  | 11.85                  | 4.64                           | −1.22          | 4    | 4              | −8.93                    | −15.55                  | 6.26                   | 1.15                           | −1.13          |
| 5    | 9              | −6.05                    | −6.53                   | 1.51                   | −2.06                          | 0.00           | 6    | 6              | −3.57                    | −17.34                  | 10.10                  | 0.14                           | −0.98          |
| 7    | 5              | 5.95                     | −11.60                  | 14.02                  | 3.33                           | 0.00           | 8    | 8              | 13.59                    | −4.79                   | 1.11                   | 1.08                           | −1.43          |
| 9    | 1              | 17.90                    | −8.84                   | 28.80                  | 1.84                           | −1.58          | 10   | 2              | 2825.46                  | −44.83                  | 3622.84               | −0.61                          | −5.55          |

Fig. 5. Docked complex of the designed multi-epitope peptide vaccine construct (shown in red) and chain A of TLR3 (shown in yellow).
3.12. Immune simulation

The vaccine construct appears to produce robust IgM antibody response even in the absence of recurrent antigen exposure. The IgM antibody titer seen reach close to 700,000 scales (Fig. 6). Among the cytokines and interleukins, high level interferon gamma was reported to the antigen compared to others (Fig. 6). Additionally, appropriate and decent cellular activity with their respective memory from the host immune system were revealed along with proliferating dendritic cells and enhanced activity from macrophages (S-Fig. 4). Different humoral immunity isotypes can be easily noticed, an indicator of isotype switching and memory cell formation (S-Fig. 5). This immune profile leads to a conclusion of a good agreement on natural development of immunity against the virus and protection against the virus.

3.13. Molecular dynamics simulations

The multi-epitope peptide vaccine construct complex was seen stably docked with the TLR3 receptor and the epitopes are exposed throughout the simulation time for efficient and easy processing by the host immune system. The dynamics of the complex was evaluated through two structural parameters. First, root mean square deviation (rmsd) was investigated that measures the distance between the carbon alpha atoms of superimposed proteins [81]. The average rmsd estimated for the system is 5.2 Å. The system is reported with some major rmsd peaks that after visual inspection found because of the moving loops of the system. This may be an approach towards complex stability as both the interacting molecules are stably docked with each other’s. Secondly, we evaluated root mean square fluctuations (rmsf) for the system. RMSF sheds light on the fluctuating residues from there mean position [82]. The average rmsf for the system is 4.6 Å. As mentioned earlier, these variation are because of fluctuating loops of the system which is not altering intermolecular conformations and does not affect overall stability. The rmsd and rmsf plots of the system can be seen in Fig. 7. Additionally, we performed hydrogen bond analysis for the complex to decipher the strength of intermolecular interactions during simulation time. As can be seen in Fig. 8, the molecules are involved in good number of hydrogen bonds throughout the simulation time, enhancing stability of the complex.

3.14. MM-PBSA binding energies calculation

The docking prediction for the vaccine construct conformation with respect to the TLR3 receptor was validated by performing computationally cost effective and sufficiently reliable MMGBSA and MMPBSA approaches. Both these approaches were run over trajectories of the MD simulations. It is evident from the Table 7 that robust energies are noticed for the system and the values are quite low indicating a good agreement on the docking findings. The intermolecular interactions are dominated by both van der Waals and electrostatic energies with energy value of $-188.44$ kcal/mol and $-473.80$ kcal/mol, respectively. Summing, both these mentioned values are contributing factor to the
significantly low gas phase energy of $-626.25$ kcal/mol. Opposed to this, non-favorable energy of solvation was noticed for the system $(564.3822$ kcal/mol in MM-GBSA and $527.0717$ kcal/mol in MM-PBSA). The net binding energy of the system in MM-GBSA is $-61.8682$ kcal/mol whereas in MM-PBSA this energy is $-99.1787$ kcal/mol. Both these net values support highly stable system and excellent intermolecular binding affinity between the vaccine construct and TLR3 receptor. The net entropy concluded for the system is $57.76$ kcal/mol which again suggesting stable nature of the vaccine at the docked site. Additionally, to evaluate intermolecular affinity towards end of the simulation time, MM/GBSA of last 20-ns was performed (S-Table 1). The net MMGBSA and MMPBSA in this time period are $-77.4$ kcal/mol and $-107.4$ kcal/mol, respectively.

4. Conclusions

This study shows that membrane protein, envelope small membrane protein, non-structural protein ORF3, non-structural protein ORF5 and spike glycoprotein could act as potential subunit vaccine candidates for designing a subunit-based vaccine against MERS-CoV. Similarly, these proteins were mapped for novel set of antigenic B and T cell epitopes that could be used in the design of a multi-epitope peptide vaccine construct to circumvent the limitations of subunit vaccines. Further, we have shown that the designed multi-epitope vaccine construct docking affinity for TLR3 immune receptor, thus maximizing its presentation to the human immune system and ensuring the evoking potential of both humoral and cellular immune responses. The prediction of docking was validated by molecular dynamics simulation and binding free energies that validated complex high stability and intermolecular affinity. Lastly, the use of the subunit vaccine candidates and multi-epitope vaccine construct in experimental in vivo model is highly recommended to underpin their immune protective potential against MERS-CoV and the current pandemic COVID-19.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.molliq.2020.114706.

Table 7

| Energy component | Average   | Std. dev. | Std. err. of mean |
|------------------|-----------|-----------|------------------|
| MM-GBSA          | $-188.4473$ | $7.4753$  | $0.7475$          |
| VDWAALS          | $-437.8031$ | $15.8762$ | $1.5876$          |
| EEL              | $587.7871$  | $15.3673$ | $1.5367$          |
| ESURF            | $-23.4049$  | $0.9328$  | $0.0933$          |
| DELTA G gas      | $-626.2504$ | $16.5671$ | $1.6567$          |
| DELTA G solv     | $564.3822$  | $15.1083$ | $1.5108$          |
| DELTA TOTAL      | $-61.8682$  | $4.7152$  | $0.4715$          |
| MM-PBSA          | $-188.4473$ | $7.4753$  | $0.7475$          |
| VDWAALS          | $-437.8031$ | $15.8762$ | $1.5876$          |
| EEL              | $547.1448$  | $15.0112$ | $1.5011$          |
| ENPOLAR          | $-20.0731$  | $0.4982$  | $0.0498$          |
| EDISPER          | $0.0000$    | $0.0000$  | $0.0000$          |
| DELTA G gas      | $-626.2504$ | $16.5671$ | $1.6567$          |
| DELTA G solv     | $527.0717$  | $14.8625$ | $1.4862$          |
| DELTA TOTAL      | $-99.1787$  | $7.2585$  | $0.7258$          |

Fig. 8. Hydrogen bonds analysis for TLR3-MEPVC complex.

References

[1] R.S. Sikkema, E. Farag, M. Islam, M. Atta, C. Reusken, M.M. Al-Hajri, M.P.G. Koopmans, Global status of Middle East respiratory syndrome coronavirus in dromedary camels: a systematic review, Epidemiol. Infect. 147 (2019).
[2] A.R. Fehr, R. Channappanavar, S. Perlman, Middle East respiratory syndrome: emergence of a pathogenic human coronavirus, Annu. Rev. Med. 68 (2017) 387–399.
[3] B.L. Haagmans, S.H.S. Al Dhahiry, C.B.M. Reusken, V.S. Raj, M. Galiano, R. Myers, G.-J. Godeke, M. Jorges, E. Farag, A. Diab, et al., Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation, Lancet Infect. Dis. 14 (2014) 140–145.
[4] A.J. Hall, J.J. Tokars, S.A. Badreddine, Z. Bin Saad, E. Furukawa, M. Al Masri, L.M. Haynes, S.I. Gerber, D.T. Kuhar, C. Miao, et al., Health care worker contact with MERS patient, Saudi Arabia, Emerg. Infect. Dis. 20 (2014) 2148.
[5] A.N. Alshukairi, J. Zheng, J. Zhao, A. Nehdi, S.A. Baharoon, L. Layqah, A. Bokhari, S.M. Al Johani, N. Sannan, M. Boudjelal, et al., High prevalence of MERS-CoV infection in camel workers in Saudi Arabia, MBio. 9 (2018) e01985–e02018.
[6] J.F.W. Chan, S.K.P. Lau, K.K.W. To, V.C.C. Cheng, P.C.Y. Woo, K.-Y. Yuen, Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease, Clin. Microbiol. Rev. 28 (2015) 465–522.
[7] D. Gyits, L.M. Carvalho, A. Rambaut, T. Bedford, MERS-CoV spillover at the camel-human interface, Elife 7 (2018).
[8] K. Chang, M. Ki, E.G. Lee, S.Y. Lee, B. Yoo, J.H. Choi, MERS epidemiological investigation to detect potential mode of transmission in the 178th MERS confirmed case in Pyeongtaek, Korea, Epidemiol. Health 37 (2015).
[9] C. Gossner, N. Danielson, A. Gervelmeyer, F. Berthe, B. Faye, K. Kaasik Aaslav, C. Adlhoch, H. Zeller, P. Penttinien, D. Coulombier, Human–dromedary camel interactions and the risk of acquiring zoonotic Middle East respiratory syndrome coronavirus infection, Zoonoses Public Health 63 (2016) 1–9.

Declaration of Competing Interest

The authors declare no conflict of interest.
K. Guruprasad, B.V.B. Reddy, M.W. Pandit, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Protein Eng. Des. Sel. 4 (1990) 155–161.

A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (1980) 1895–1898.

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

L. Heo, H. Park, C. Seok, GalaxyRefine: protein structure refinement driven by side-chain repacking, Nucleic Acids Res. 41 (2013) W384–W388.

A.A. Dombkowski, K.Z. Sultana, D.B. Craig, Protein disulfide engineering, FEBS Lett. 588 (2014) 206–212.

D.B. Craig, A.A. Dombkowski, Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins, BMC Bioinformatics 14 (2013) 346.

E. Angov, Codon usage: nature’s roadmap to expression and folding of proteins, Biotechnol. J. 6 (2011) 650–659.

A. Grote, K. Hiller, M. Scheer, R. Münch, B. Nörtemann, D.C. Hempel, D. Jahn, JCat: a novel tool to adapt codon usage of a target gene to its potential expression host, Nucleic Acids Res. 33 (2005) W526–W533.

V.N. Maiorov, G.M. Crippen, Significance of Root-mean-Square Deviation in Comparing Three-dimensional Structures of Globular Proteins, 1994.

A. Kuzmanic, B. Zagrovic, Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors, Biophys. J. 98 (2010) 861–871.