Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach

Shafi Mahmud1,8, Md. Oliullah Rafi2,8, Gobindo Kumar Paul1, Maria Meha Promi3, Mst. Sharmin Sultana Shimu3, Suvro Biswas3, Talha Bin Emran4, Kuldeep Dhama5, Salem A. Alyami7, Mohammad Ali Moni6 and Md. Abu Saleh1

Currently, no approved vaccine is available against the Middle East respiratory syndrome coronavirus (MERS-CoV), which causes severe respiratory disease. The spike glycoprotein is typically considered a suitable target for MERS-CoV vaccine candidates. A computational strategy can be used to design an antigenic vaccine against a pathogen. Therefore, we used immunoinformatics and computational approaches to design a multi-epitope vaccine that targets the spike glycoprotein of MERS-CoV. After using numerous immunoinformatics tools and applying several immune filters, a poly-epitope vaccine was constructed comprising cytotoxic T-cell lymphocyte (CTL)-, helper T-cell lymphocyte (HTL)-, and interferon-gamma (IFN-γ)-inducing epitopes. In addition, various physicochemical, allergenic, and antigenic profiles were evaluated to confirm the immunogenicity and safety of the vaccine. Molecular interactions, binding affinities, and the thermodynamic stability of the vaccine were examined through molecular docking and dynamic simulation approaches, during which we identified a stable and strong interaction with Toll-like receptors (TLRs). In silico immune simulations were performed to assess the immune-response triggering capabilities of the vaccine. This computational analysis suggested that the proposed vaccine candidate would be structurally stable and capable of generating an effective immune response to combat viral infections; however, experimental evaluations remain necessary to verify the exact safety and immunogenicity profile of this vaccine.

Middle East respiratory syndrome coronavirus (MERS-CoV) originated in bats and causes an acute, infectious, viral respiratory disease characterized by symptoms including cough, fever, diarrhea, and occasional multi-organ failure1,2. MERS-CoV is covered by a single-stranded RNA and belongs to the Beta-coronavirus (β-CoV) genus in the family of Coronaviridae, which is distinct from the coronaviruses that cause the common cold coronavirus and severe acute respiratory syndrome (SARS)3–5. MERS-CoV was first identified in Saudi Arabia in 2012 and later spread to European countries. MERS-CoV is associated with an unusually high mortality rate of approximately 35%6–8. According to the WHO report released in May 2018, a total of 2,220 people contracted MERS-CoV, including 790 deaths, resulting in a mortality rate of 35.6%. In Saudi Arabia alone, 1,844 cases were identified, resulting in 716 deaths. In 2015, the South Korean population also became infected with this acute virus, resulting in 186 cases and 36 deaths9,9.

The genome for MERS-CoV encodes envelope (E), membrane (M), nucleocapsid (N), and spike (S) structural proteins, which are necessary to complete the structure of this viral particle10,11. The E protein is typically

1Microbiology Laboratory, Department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi 6505, Bangladesh. 2Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Jashore 7408, Bangladesh. 3Department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi 6505, Bangladesh. 4Department of Pharmacy, BGC Trust University Bangladesh, Chittagong 4381, Bangladesh. 5Division of Pathology, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly 243122, Uttar Pradesh, India. 6Faculty of Medicine, WHO Collaborating Centre on eHealth, UNSW Digital Health, School of Public Health and Community Medicine, UNSW Sydney, Sydney, NSW 2052, Australia. 7Department of Mathematics and Statistics, Imam Mohammad Ibn Saud Islamic University, Riyadh 11432, Saudi Arabia. 8These authors contributed equally: Shafi Mahmud and Md. Oliullah Rafi. 9email: m.monir@unsw.edu.au; saleh@ru.ac.bd
expressed when cells become infected and is distributed within the intracellular membranes of the endoplasmic reticulum (ER) and Golgi compartments\textsuperscript{12–14}. The N protein binds to the genome of the CoV-RNA, forming the nucleocapsid\textsuperscript{15}, whereas the M protein defines the shape of the viral envelope\textsuperscript{16}. The spike (S) glycoprotein is another important protein because it is involved in host recognition and facilitates host cell entry\textsuperscript{17,18}. The spike protein consists of S1 and S2 subunits\textsuperscript{19,20}. To initiate infection, the S1 subunit binds to the dipeptidyl peptidase 4 (DPP4) receptor on the host cell surface\textsuperscript{19,20,21}, and the S2 subunit mediates the fusion of the viral and host cell membranes\textsuperscript{20}. As of 2020, no specific vaccines against MERS-CoV are available. The S glycoprotein has been considered to be a potential vaccine candidate because neutralizing antibodies against this protein would block viral entry and prevent viral infection\textsuperscript{23–25}.

Using conventional methods, vaccines are designed using large proteins, and the use of inappropriate antigens can increase the potential for allergic reactions; however, a peptide-based multi-epitope vaccine that contains short antigenic peptide fragments, referred to as epitopes, might be able to overcome these limitations\textsuperscript{26}. Epitopes, the antigenic portion of the pathogen that is recognized by the host immune system, and innate immunity are elicited against it\textsuperscript{27}. The cell-mediated immune response is mainly dependent on the pattern recognition receptors recognizing the pathogen-associated molecular patterns of the pathogen\textsuperscript{28}. Toll-like receptors are considered pathogen recognition receptors (PRRs), TLRs family consists of eleven proteins, and each of them uniquely interacts with diverse PAMPs, they are expressed on the surface of the cells\textsuperscript{29}. Toll-like receptor-2 can recognize viral structural glycoproteins. The TLR4 has an essential role during host pathogenesis that can trigger the anti-viral host defense mechanisms against coronavirus\textsuperscript{30}. After infection, cytotoxic T-cell lymphocytes (CTLs) become activated and kill infected cells\textsuperscript{31}. Antigens bound to the major histocompatibility complex (MHC) are presented on the infected cell surface, allowing them to be recognized by CTLs\textsuperscript{32}. MHC class I molecules exhibit cytosolic peptides antigens of the infected cells and phagocytosed antigens are presented on MHC class II molecules\textsuperscript{33–36}. CD4 + T cells recognize the antigenic peptides that are displayed by class II MHC molecules whereas cytotoxic T-cell lymphocyte (CTL)-interact with class I MHC-peptide complexes\textsuperscript{37}. Activated CD4 + T cells secrete cytokines and are responsible for the further activation of B cells required for producing proper antibodies\textsuperscript{38}.

In this study, an immunoinformatics strategy was applied in a sequential manner to design a multi-epitope subunit vaccine to target the MERS-CoV S glycoprotein. This epitope-based vaccine prediction was designed to be antigenic and non-allergenic in nature. The final vaccine contained cytotoxic T-cell lymphocyte (CTL)-, helper T-cell lymphocyte (HTL)-, and interferon-γ (IFN-γ)-inducing and B-cell epitope sequences. Additional physicochemical, molecular docking and thermodynamic stability profiling were assessed to evaluate the safety and efficiency of the vaccine (Fig. 1).
Results

Sequences retrieval and analysis of phylogenetic tree. A phylogenetic tree was built using all of the spike glycoprotein sequences obtained from various MERS-CoV isolates identified in different countries (Fig. 2). The results showed that all of the S proteins assembled together into a single clade and were closely related to one another. Therefore, a MERS-CoV vaccine that was designed against one single strain could potentially be effective against all of the strains.

Antigenic and physicochemical evaluations of the target protein. The VaxiJen v2.0 predicted score was 0.4919, which signified that the sequence was potently antigenic. The ExPASy tool in ProtParam indicated that the virulent target protein being used in this study has an instability index score of 36.53, an aliphatic index score of 83.00, and a negative grand average of hydropathicity (GRAVY) value of −0.064 (Supplementary Table 1). The targeted spike glycoprotein contains 1353 amino acids with the molecular weight of 149,505.26 kDa which specifies the sequences to be antigenic. The estimated half-life was computed to be 30 h for mammalian-reticulocytes, > 20 h for yeast, > 10 h for *Escherichia coli*.

The prediction of T-cell epitopes and assessment. The CTL epitopes can be used to eliminate virus-infected cells, generate cellular immunity, and reduce the level of circulating virus. In contrast, HTL epitopes can develop both cellular and humoral immune responses and have the ability to activate B-cells to produce
antibodies. Therefore, an effective vaccine should contain receptor-specific cytotoxic T-cell lymphocyte, helper T-cell lymphocyte epitopes. In this study, the NetCTL-1.2 server and the Infectious Disease Epidemiology Bureau (IDEB) stabilized matrix method (SMM) were employed for the prediction of the CTL epitope, whereas the HTL epitope was predicted using the Net MHCII pan 3.2 server (Supplementary Tables 2 and 3). In addition, the predicted sequences were subjected to numerous immune filters, including antigenicity, high-binding affinity, non-allergenicity, and promiscuously towards MHC (MHC-I and MHC-II) alleles. Finally, the best candidate was selected, as shown in Tables 1, 2, and Supplementary Table 4.

Table 1. Selected promiscuous cytotoxic T-lymphocyte epitope. The epitopes listed in the table are non-allergic, non-toxic, and displayed 100% conservancy among target protein sequence (as predicted using the IDEB conservancy tool). Epitopes with half-maximal inhibitory concentration (IC50) values below 250 were considered to bind suitably with respective human leukocyte antigen (HLA) alleles. The VaxiJen v2.0 server was utilized at 0.4 thresholds to predict the antigenicity of the epitopes.
Vaccine construction, structure modeling, refinement, and validation. The criteria used for epitope selection during the construction of the linear vaccine included: (a) antigenic and non-allergic; (b) high-binding affinity with MHC alleles; (c) promiscuous; (d) containing overlapping CTL and HTL epitopes; (e) 100% conserved across S proteins; and (f) no overlapping capacity with any component of the human proteome. Epitopes that featured these properties were used to construct a linear vaccine sequence, comprising 7 CTL, 9 HTL, and 6 IFN-γ epitopes (Tables 1, 2, and Supplementary Tables 4–7). In addition, AYY and GPGPG linkers were added to the construction to prevent junctional epitope formation, facilitate the immune processing of the antigens and allow effective separation of each epitope within the human body40. To improve immunogenicity, cholera toxin subunit B (CTB) was added as an adjuvant attached to the N-terminal end (Fig. 3). The finalized vaccine sequence featured a molecular weight of 51.8 kDa and consisted of 489 amino acids (Supplementary Material SM1). The selected epitopes included in the vaccine were visualized on the 3D model of the S glycoprotein, which displayed the epitopes retaining their original positions (Fig. 4). The secondary structure prediction for the vaccine suggested that it would contain 24.34% alpha-helical regions, 27.20% extended strands, 4.50% beta turns, and 43.97% random coils (Fig. 5A). The 3D model of the finalized vaccine was prepared by the trRosetta web server which displayed five different models as per their Z-score ranging from 1.74 to 3.03 where the highest TM-score containing model (0.61 ± 0.05) was subjected to refinement. TM-score measures the similarity between two structures. A TM-score is higher than 0.5 is implying that a model of correct topology and less than 0.17 indicates random similarity41. Among all of the refined 3D structures, model 1 was selected which was predicted to be the best based on different parameters including MolProbity (2.72), GDT-HA (0.9059), and RMSD (0.519) (Fig. 5B and Supplementary Table 8). The poor rotamers and clash score of the finalized model were predicted to be 2.1 and 34.1, respectively. The global distance test-high accuracy (GDT-HA) is a global measure agreements between the experimental model and a predicted protein structure42. The Z-score of the finalized vaccine structure was −6.85, as predicted by Pro-SA web (Fig. 5C), and the score was within the range for comparably-sized proteins, which indicated that the predicted model was reliable43. The results of the Ramachandran plot analysis showed that 87%, 12%, 0.8%, and 0.3% residues were in favoured, additional, allowed, and disallowed regions, respectively (Fig. 5D and Supplementary Materials SM2). The ERRAT value of the structure was predicted to be 56.017 (Supplementary Fig. S1). The ERRAT score higher than 50 indicating the good quality of a protein model44.

Assessment of the antigenic, physicochemical, and allergenic profiles of the vaccine construct. The VaxiJen v2.0 server predicted a score of 0.5134, which confirms that the candidate vaccine is antigenic. Allergenicity was analyzed by AllerTOP v.2.0 and the AllergenFP v.1.0 server and was determined to

| Epitopes       | Peptide position | HLA-DRB1* alleles | Antigenicity |
|----------------|------------------|-------------------|--------------|
| TIKYYSIPHSIRSI | 289              | 01:01 03:01 04:01 | 0.556        |
| LNSTYFKLSIPTNFS | 773              | 01:01 03:01 04:01 | 0.4286       |
| KYYSIIPHSIRSIQS | 291              | 03:01 04:01 08:01 | 0.4516       |
| IKYYSIIPHSIRSIQ | 290              | 07:01 08:01 08:03 | 0.61         |
| NSTYFKLSIPTNFSF | 774              | 01:01 03:01 04:01 | 0.68         |
| QSYFRLNGVGITQQ  | 974              | 10:01 11:01 12:01 | 0.7573       |
| TYFKLSIPTNFSFGV | 776              | 13:01 14:01 15:01 | 0.933        |
| VSFVNPNGLYFMH   | 1124             |                   | 0.4899       |
| STYFKLSIPTNFSFG | 775              |                   | 0.8983       |

Table 2. Selected helper T-lymphocyte epitopes, as predicted by the NetMHC II pan 3.2 server. The strong, intermediate, and non-binding affinities are represented by red, green, and black boxes, respectively. The epitope antigenicity was predicted by the VaxiJen v2.0 server at the 0.4-threshold level. The epitopes are promiscuous, non-allergic, non-toxic, and displayed 100% conservancy with the target sequence.
be non-allergenic (Supplementary Fig. S2). The physicochemical properties of a vaccine must be determined to assess efficiency and safety\(^4\); therefore, several physical and chemical parameters of the vaccine construct were evaluated by ExPASy (Supplementary Materials SM3). The aliphatic index and theoretical pI values were found to be 71.10 and 9.74, respectively, which suggested that the candidate is thermostable. The estimated half-life was evaluated to be 30 h in mammalian reticulocytes, > 20 h in yeast, > 10 h in *E. coli*, The Grand average hydro-
Phathicity (GRAVY) score is -0.130 which specifies the candidate is hydrophilic in nature and has the ability to interact with the aqueous environment. The instability index was calculated as 26.58, which classified the protein as being stable. In addition, our designed vaccine does not contain any signal peptides that would either specify or inhibit protein localization (Supplementary Fig. S3). The TMHMM server predicted that no production difficulties would be associated with expression (Supplementary Fig. S4).

Post-translational modification analysis. A wide range of post-translational modifications was predicted in both targeted spike glycoprotein and the finalized vaccine construct. Analysis of glycosylation site within the target protein indicated that there are two N-glycosylation sites were existing (position 6 and 244) and one O-GlcNAc site was found in 878 positions (Supplementary Figs. S5 and S6) which specifies the epitopes which are used in vaccine designing are not overlapping in those regions (PTM sites). Additionally, there was no lipid PTMs as N-terminal glycines myristoyl and GPI-modification was predicted in the finalized vaccine construct. Prediction of phosphorylation modification displayed 45 sites (Ser: 24, Thr: 16, Tyr: 5) in the construct. The YinOYang server predicted one O_GlcNAc attachment site to exist in the vaccine construct and there is no N-glycosylation and N-acetylation was found by NetNGlyc 1.0 server and NetAct 1.0 server, respectively (Supplementary Table 9).

Prediction of the B-cell epitope. B-cells have significant roles in humoral immunity through the secretion of antibodies. A B-cell epitope can be recognized by the B-cell receptor, providing long-term immunity. The Ellipro server result included 10 linear and four conformational B-cell epitopes (Tables 3 and 4). PyMOL was used to visualize the mapped epitopes in the vaccine construct (Supplementary Figs. S7 and S8).

Population coverage. Successful HLA allele calculations are necessary to develop a vaccine that will be effective for the entire global population. The IDEB Population Coverage tool indicated that the selected epitope
shown a strong binding affinity. The nine HTL epitopes were docked individually with HLA-DRB1*01:01 and compared to the CTL positive (lowest energy -562.6) and negative (lowest energy 756.5) control (Supplementary Fig. 8) and also compared the binding affinity with positive and negative control. We found that 7 CTL epitopes monly occurring HLA alleles in the human population HLA-A*02:01, HLA-DRB1*01:01 and HLA-DRB1*15:01 selected CTL and HTL epitopes included in the finalized vaccine construct docked individually with three com-plexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
plexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
plexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
plexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
plexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
plexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
mplexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
mplexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
mplexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
mplexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
mplexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
patchDock server generates the interaction model of a receptor-ligand complex. The refinement of the best com-
mplexes was performed using the FireDock tool and found ten complexes for each TLR2-vaccine and TLR4-vac-
cine. Among the top-ten complexes generated when examining the interaction between the vaccine and TLR-4,
HLA-DRB1*15:0, the best model was NSTYFKLSIPTNFSF, which shows the lowest energy (-987.0 with HLA-DRB1*01:01 and -1125.6 with HLA-DRB1*15:0) compared to the positive and negative control (Supplementary Table 14). Thus, these T-cell epitopes ensured their suitable binding affinity to be used in multi-epitope vaccine design against MERS-CoV.

Analysis of molecular dynamics and simulation. The root means square deviation (RMSD) of the c-alpha atoms of the vaccine complex was calculated for two vaccine complexes. The average RMSD values for the vaccine and TLR-2 complex and for the vaccine and TLR-4 complex were 2.043 Å and 2.357 Å, respectively, which demonstrates the stable nature of both complexes. As shown in Fig. 9A, both the vaccine and TLR-2 and vaccine and TLR-4 complexes experience an initial increase in RMSD descriptors until the 15 ns time point, after which the upward trend ceases. A low degree of fluctuation was observed for both two complexes, which may be responsible for structural integrity. Therefore, the solvent-accessible surface area (SASA) values for both complexes were analyzed to understand changes in protein volume, based on the simulation trajectories. The average SASA value of the vaccine and TLR-2 complex was found to be 74,896.758Å², which is similar to the entire SASA profile (Fig. 9D) because no deviations were found. In contrast, the vaccine and TLR-4 complex had

Table 5. Population coverage of the selected epitope included in the vaccine construct. aCoverage of population on projected. bPopulation recognized by HLA combinations/epitope hits on the average number. c90% of the population recognized by HLA combinations/epitope hits on the minimum number.

| Population/area | Coveragea | Average_hitb | pc90c |
|-----------------|-----------|--------------|-------|
| Central Africa  | 99.34%    | 26.31        | 14.8  |
| East Africa     | 99.25%    | 25.78        | 14.64 |
| East Asia       | 100.0%    | 39.72        | 28.42 |
| Europe          | 100.0%    | 47.54        | 35.91 |
| North Africa    | 99.8%     | 30.56        | 17.21 |
| North America   | 100.0%    | 46.02        | 31.76 |
| Northeast Asia  | 99.1%     | 24.76        | 14.4  |
| Oceania         | 99.35%    | 24.71        | 14.5  |
| South Africa    | 99.82%    | 28.55        | 15.78 |
| South America   | 100.0%    | 31.57        | 21.27 |
| South Asia      | 100.0%    | 40.36        | 28.65 |
| Southeast Asia  | 100.0%    | 35.37        | 22.34 |
| Southwest Asia  | 98.64%    | 23.09        | 14.02 |
| West Africa     | 99.55%    | 27.53        | 15.21 |
| West Indies     | 99.46%    | 24.44        | 14.66 |
| World           | 100.0%    | 44.2         | 30.49 |
| Average         | 95.83     | 30.82        | 19.71 |
| Standard deviation | 15.27  | 10.49        | 8.55  |

Figure 6. The docked complex of the designed multi-epitope vaccine protein and TLR-4 immune receptor. The vaccine construct has been shown in CPK molecular model (pink) in the figure and the global energy value of the finalized docked complex was −21.16 kcal/mol.
Figure 7. The docked complex of the designed multi-epitope vaccine protein and TLR-2 immune receptor. The vaccine construct has been shown in CPK molecular model (pink) in the figure. The global energy of the finalized docked complex was −22.13 kcal/mol as predicted by the PatchDock server.

Figure 8. The interaction patterns of selected epitopes included in the vaccine constructs with HLA alleles. A1–A7 are the binding pattern of 7 CTL epitopes (red) with HLA-A*02:01 allele (magentas). The interaction pattern of 9 HTL epitopes (red) with the HLA-DRB1*01:01 (blue) allele has been shown in figure B1–B9. Figures C1–C9 represents the binding pattern of 9 HTL epitopes with HLA-DRB1*15:01 allele (green).
an average SASA value of 87,958.090 Å². Both simulated complexes displayed an initial rise in the SASA values, which indicated the expansion of the protein volume during the initial phase. The radius of gyration (Rg) for the simulation trajectory provides information regarding the compressed nature of the protein, in which a higher Rg profile denotes less rigidity in the biological system. As shown in Fig. 9B, the Rg profile of the two vaccine complexes could be demonstrated, and an initial increase in the Rg profile for the vaccine and TLR-2 complex was observed. Therefore, the Rg descriptor trend for the vaccine and TLR-2 complex was similar from 0 to 50 ns, although a few fluctuations were observed. In contrast, the Rg value of the vaccine and TLR-4 complex was similar until 35 ns, followed by a larger deviation observed from 35–38 ns, which might be responsible for the loose packaging of the system. The vaccine molecules and their degrees of stability were further assessed by examining the total hydrogen bond number during the entire simulation period. The formation of hydrogen bonds and the number of substantial changes in the simulated complexes can be used to define the rigidity of the complexes. The vaccine and TLR-2 and the vaccine and TLR-4 complexes both featured stable numbers of hydrogen bonds, as shown in Fig. 9C, indicating relative complex stability. Moreover, the protein flexibility across the amino acid residues was evaluated by assessing the root mean square fluctuation (RMSF) profile. The RMSF profiles of the TLR-2 and TLR-4 complexes indicated that most of the amino acid residues from both complexes had RMSF profiles below 2.5 Å, and larger changes were only observed for a few residues. This result, as shown in Fig. 10, defines the vaccine complex stability and stiffness. Furthermore, secondary structure content from TLR-4-vaccine, TLR-2, and vaccine complexes was analyzed to understand their structural integrity (Supplementary Fig. S9). The secondary structure content; alpha helix, beta-sheet, coil and turn the region from both biological complexes were steady state in whole molecular dynamics simulation.

**In silico immune simulation.** The immunogenic profile of the chimeric peptide vaccine is shown in Fig. 11 and Supplementary Fig. S10. The immune simulation results showed that the secondary and tertiary responses were considerably more immunogenic than the primary response. The antigenic concentration decreased, and the immunoglobulin activity became significantly increased during the secondary and tertiary responses (Fig. 11A). In addition, the formation of multiple B-cell isotypes was predicted (Fig. 11B). A similarly increased response was indicated by HTL and CTL populations during vaccination (Fig. 11C and D). Increased macrophage activity was detected in the form of natural killer (NK) and dendritic cells (Supplementary Fig. S10 and Fig. 11E). High levels of IFN-γ supported the activation of an adequate immune response, and IL-2 secretion
was also elicited during the simulation (Supplementary Fig. S10E). This profile demonstrates the development of immune memory subsequently increased.

**Discussion**

Although MERS is a highly infectious disease with a high mortality rate, no approved vaccines and or treatment drugs are currently available to prevent MERS-CoV infection. Vaccines are urgently necessary for the prevention and treatment of this disease. The application of an immunoinformatics strategy and the associated tools represents a rationally effective approach to the development of a peptide-based vaccine. This approach is time-consuming, cost-effective, and can allow researchers to predict antigenic epitopes that may be potential candidates for use in an effective multi-epitope vaccine. In this investigation, our aim was to design a peptide-based poly-epitope vaccine against MERS-CoV. By applying a cutting-edge immunoinformatics strategy against the MERS-CoV proteome, the sequence of the most recognizable virulent factor, the S glycoprotein, was targeted. The antibody-mediated neutralization of the viral S protein is a primary goal in vaccine development because the spike glycoprotein is associated with viral attachment and entry to host cells. Here, we constructed a multi-epitope vaccine by using several computational tools. The vaccine was designed to provide immunity based on the use of several small antigenic peptide fragments; in contrast to vaccines that utilize the whole genome or large proteins, this approach does not produce any allergic reactions in the host. In addition, our designed
vaccine has several advantages compared with conventional and single-epitope vaccines owing to the following distinctive characteristics: (a) the presence of multiple MHC epitopes, allowing the vaccine to be recognized by several T-cell receptors; (b) the use of overlapping CD4+ and CD8+ T-cell epitopes; (c) the presence of multiple epitopes from the targeted virulent protein; and (d) the inclusion of an immunostimulator (adjuvant) for producing long-term immunity. The design of a vaccine using a similar method has been demonstrated to result in the development of protective efficiency in vivo, and some of these vaccines have entered the clinical trial phase. The selection of antigenic epitopes is crucial for vaccine development. The cytotoxic T-cell lymphocyte, helper T-cell lymphocyte, and interferon-gamma inducing epitopes of the S protein were identified based on the use of several immune filters. Epitopes were screened to identify antigenic but non-allergenic sequences that were capable of interacting with multiple HLA alleles were conserved 100% among the various S protein sequences and did not overlap with any components of the human proteome to reduce the possibility of autoimmunity. Through the application of various in silico analysis techniques, the designed poly-epitope (CTL, HTL, IFN-γ inducing)-containing vaccine was constructed with the addition of an N-terminal-linked Cholera toxin B (CTB) adjuvant. Adjuvant and epitopes (CTL-HTL-IFN-γ) are ordered in such way have been shown to antigen-specific immune response stimulation reported in many studies, the adjuvant is attached in the N-terminal of the construct because binding of it with TLR displays enhanced results in producing immune-response whereas TLRs associated to the activation of the humoral and cellular immunity. Cholera toxin B is potentially used as a viral adjuvant. The use of specialized sequences including linkers has the ability to improve vaccine construct. Previously many studies demonstrated that GPGPG and AAY linkers were added between predicted HTL and CTL epitopes sequences respectively and produced junctional immunogenicity, consequently allowing the rational design construction of a potent poly-epitope vaccine. Arai et al. reported that the EAAAK linker was incorporated between epitopes and adjuvant for improving bioactivity of fused protein and reaching a high level of expression and increasing the stability of the vaccine construct. Similarly, Bazan and colleagues designed a T-cell-based multi-epitope subunit vaccine against the Ebola virus. They used the Immune Epitope Database (IDEB) to predict antigenic epitopes that were used to construct a vaccine candidate that was found to be immunogenic when expressed in mice.

Figure 11. The in silico immune simulation profile of vaccine which injected 12 doses over a period of 12 months. (A) shows both antigen concentration and relative antibody responses. The efficacy of the vaccination showed by the presence of protective IgGs. (B) shows the corresponding count of antibody generating plasma cells while (C–E) show the activity of macrophages cytotoxic T, and helper T cells.
During vaccine development, allergenicity can be a major problem. In our final vaccine, allergenicity was not detected. Various physicochemical properties were determined using the ProtParam ExPASy tool, which indicated an instability index value of 26.58, which indicated that the vaccine would be stable. The theoretical PI value was calculated as 9.74. The aliphatic index of the vaccine was evaluated to be 71.10, which indicates that the protein would be thermostable. The GRAVY value was calculated to be 0.130, which suggested that the vaccine can interact with water, has a polar nature, and a high degree of solubility. Foroutan and other researchers designed a vaccine and then validated their candidate through experimental evaluation, reporting that their vaccine was able to induce strong cellular and humoral immune responses in mice. The aliphatic and instability index values for our designed vaccine were better than those reported for the vaccine candidate designed and tested by Foroutan et al. The Ramachandran plot analysis for the vaccine showed that 87%, 12.0%, 0.8%, and 0.3% residues were found in the favored, allowed, additionally allowed, and disallowed regions, respectively, with an ERRAT value and a Z-score of 56.017 and −6.85, respectively, which indicates that the protein falls in the plot of Ramachandran, which consists of the Z-scores of the already determined structures solved by NMR and X-ray crystallographic experiments. TLR-4 is expressed in monocytes, granulocytes, and immature dendritic and macrophage cells.

The direct interaction between CTB and TLR-4 facilitates the activation of TLR-4 by CTB. An enzyme-linked immunosorbent assay (ELISA) indicated that CTB is capable of inducing the activation of NF-κB in TLR-4 receptor cells through direct binding. The viral E glycoprotein can also be recognized by TLR-2. Molecular docking studies were used to analyze the molecular interactions and binding affinity patterns between the vaccine and both TLR-2 and TLR-4.

The global energies of the best-docked complexes between the vaccine and TLR-4 and TLR-2 were −21.16 kJ/mol and −22.13 kJ/mol, respectively, which denoted favorable binding affinities. The molecular dynamics simulation study of the vaccine candidate and TLR-4 and TLR-2 complexes were conducted to confirm their stable nature at atomic conditions. The simulation data by combining RMSD, RMSF, SASA, Prg descriptors from trajectories correlates with the structural rigidity of the vaccine complexes. The RMSD and RMSF profile of the vaccine candidates were below 2.5 Å for most of the simulation time. These results define the vaccine complexes’ integrity and less mobility at the simulation conditions. The immune simulation study indicated that our designed vaccine candidate could likely generate an appropriate immune response during secondary exposure after the final injection (Fig. 11).

Many researchers have recently applied similar immuninformatics approaches to the design of multi-epitopic vaccine candidates against Kaposi\'s sarcoma, Klebsiella pneumoniae, Pseudomonas aeruginosa, dengue, Nipah virus, SARS-CoV-2, Hendra virus, and malaria. A similar strategy has also been used to develop a vaccine against cancerous antigens. Therefore, the construction of a vaccine using epitopes appears to be capable of inducing the activation of immune cells in the host, which may further trigger other immune cells via a complex signaling pathway.

Materials and methods

Retrieval of protein sequence and the analysis of the phylogenetic tree. The NCBI database was utilized to retrieve the S glycoprotein amino acid sequences for MERS-CoV (Accession no: ANI69889.1), which were stored in a FASTA format. The sequences were aligned using the MUSCLE tool, and a phylogenetic tree was constructed by using MEGA-X. The Jones-Taylor-Thornton (JTT) model was used to an estimated algorithm for a matrix of pairwise distances to select superior log-likelihood values.

Antigenic and physicochemical evaluation of the target protein. The prediction of antigenicity is an important step when attempting to isolate the most antigenic protein sequences. The stored protein sequence was submitted to the VaxJen v2.0 Server in plain sequence format to determine antigenicity, using default parameters. The physicochemical properties of the protein were analyzed by the ExPaSy tool.

T-cell epitope prediction and assessment. CTL epitope prediction. Cytotoxic (CD8\(^+\)) T-cell epitopes were predicted by submitting FASTA sequences of the target protein to the NetCTL-1.2 server where thresholds were set for epitope identification, TAP transport efficiency, and proteasomal C-terminal cleavage were 0.75, 0.05, and 0.15, respectively. NetCTL 1.2 can predict epitope from the query sequence on the basis of the structural rigidity of the vaccine complexes. The Rmsd and Rmsf profile of the vaccine candidates were below 2.5 Å for most of the simulation time. These results define the vaccine complexes’ integrity and less mobility at the simulation conditions.

HTL epitope prediction. The Net MHCI pan 3.2 server was utilized to predict 15-mer-long epitopes capable of recognizing human leukocyte antigen class II DRB1 alleles:01:01, 03:01, 04:01, 07:01, 08:01, 08:03, 10:01, 11:02, 12:01, 13:02, 14:01, and 15:01. The server uses artificial neuron networks to predict the peptide that binding to HLA-DQ, HLA-DR, and HLA-DP alleles, and the server to predict the epitope was performed based on receptor affinity, which is assigned a percentile rank for each predicted output. The predicted epitopes were divided into strong, weak, and non-binder, based on percentile scores of less than 2%, 2%–10%, and greater than 10%, respectively. The VaxJen v2.0 server was used for assessing the antigenicity of each epitope at 0.4 thresholds. AllerTOP v.2.0, AllergenFP v.1.0, AllerCatPro v.1.7, and ToxinPred server were applied to screen out epitopes based on non-allergenicity and non-toxicity. The visualization of the epitope in the S glycoprotein was performed using PyMOL.
Prediction of B-cell epitopes. Linear and conformational B-cell epitopes were predicted by the ElliPro server90 of IDEB. This method is driven by three algorithms that perform the calculation of the protein shape as an ellipsoid, the residues protrusion index (PI) calculation, and neighboring residues clustering on the basis of PI values. For the output of each B-cell epitope, ElliPro gives a score described as PI values for each residues. The residue Protrusion Index (PI) was also calculated. ElliPro indicated that 90% of protein residues were associated with PI values of 0.9 and 10% of residues contain without Ellipsoid. Ellipsoid and ElliPro are considered the most significant prediction tools for all proteins90.

Prediction of interferon-gamma inducing epitopes. In both cell-mediated and adaptive immune systems, interferon-gamma (IFN-γ) cytokine plays a significant role to stimulate natural killer cells and macrophages for exerting immunity against viral and bacterial infections. For the prediction of IFN-γ in the target protein, we used the IFN epitope server (http://crdd.osdd.net/raghava/ifnepitope/)94, which is based on an IFN-γ dataset. The main aim of this server is to design and predict IFN-γ inducing containing peptide sequences from the query protein and it works based on a training dataset of 10,433 experimentally validated helper T-cell epitopes from the IDEB database91. From the input protein, the server generates overlapping IFN-γ inducing epitopes as displayed by the numerical score. The support vector machine (SVM) and a motif hybrid method were used to perform the prediction.

Population coverage and epitope conservancy calculation. The expression and distribution of HLA alleles could vary throughout the world according to the difference in regions. Thus, it is necessary to assess the HLA allele distribution around the world population. Population coverage of the selected epitope was analyzed by the IEDB population coverage tool92 and the conservancy of each epitope was evaluated by the IEDB analysis resource.

Molecular interaction pattern analysis of selected epitopes with HLA alleles. The three-dimensional structure of selected CTL and HTL epitopes was generated by an online server named PEPFOLD 3.593 and the 3D crystal structure of the three commonly occurring HLA alleles in the human population HLA-A*02:01, HLA-DRB1*01:01, and HLA-DRB1*15:01 was downloaded from the RCSB PDB database ID- 1QEW, 2GH9, and 1BX2, respectively. The HLA class I and class II alleles used in the docking study are expected to cover more than 95% of the worldwide population94. The PDB structure was prepared before running P-P docking by removing water and ligand and energy minimization of the structure was carried out. In order to analyze the interaction pattern of screened out best epitopes with HLA alleles, ClusPro 2.095 P-P docking server was employed. This web-based docking tool accomplishes rigid docking by sampling billions of conformations, energy minimization, and pairwise RMSD of the complexes and estimate binding energy score of the Protein–Protein docked complex on the basis of shaped complementarity. Deocys as references states, and desolvation employed. This web-based docking tool accomplishes rigid docking by sampling billions of conformations, energy minimization, and pairwise RMSD of the complexes and estimate binding energy score of the Protein–Protein docked complex on the basis of shaped complementarity. This web-based docking tool accomplishes rigid docking by sampling billions of conformations, energy minimization, and pairwise RMSD of the complexes and estimate binding energy score of the Protein–Protein docked complex on the basis of shaped complementarity. Deocys as references states, and desolvation employed. Molecular docking was carried out under the hydrophobic environment and the best cluster (complex) of epitopes and alleles was selected based on the lowest docking energy score. Finally, PyMol was used to visualize the interaction between epitopes and alleles.

Vaccine construction, structure modeling, and validation. The selected CTL, HTL, and IFN-γ epitopes were linked by using AYY and GPGPG linkers, and an adjuvant CTB was added to the N-terminal of the sequences via an EAAAK linker. Secondary structural features of the vaccine construct including extended strands, alpha-helical regions, random coils, and beta turns were predicted by the SOPMA server. The trRosetta (transform-restrained Resta) online tool96 (https://yanglab.nankai.edu.cn/trRosetta/) was applied to generate the three-dimensional model of the linear vaccine constructs, and the GalaxyRefine web server97 was used to refine the model. The trRoseta server has been considered to be the accurate and fast algorithms for de novo protein structure prediction and the principle of this algorithm is to attempt to mimic the interplay of global and local interaction in defining protein model. On the basis of direct energy minimizations with a retrained Rosetta this computational tool generates three dimensional structure from the input protein sequence. The restrained comprises orientation distributions and inter-residues distance98. Galaxy Refine web server primarily reconstructs side chains and accomplishes repacking of the side chain and then utilizes molecular dynamic simulation to attain overall structure relaxation. This algorithm improves the quality of local structure according to CASP10 calculation techniques99. Validation of the vaccine structure was performed based on the ERRAT and Z-score41. ProSA-web determines and assigns Z-score for input protein structure which is shown on the plot in the context of all known protein models whereas structures have been evaluated by X-ray crystallography and NMR, this tool also displayed any problematic part of the structure as highlighted in three-dimensional molecule viewer. When the score was outside the range for comparably-sized proteins it specifies structure error41. The ERRAT web server evaluates non-bonded atom–atom interaction compared to reliable high-resolution crystallography structures. Finally, the overall structural quality was validated by a Ramachandran plot analysis, followed by the PROCHECK server100.

Assessment of the physicochemical and antigenic properties of the vaccine. Antigenicity was measured by the Vaxijen v2.0 web tool84, where allergenicity was checked by the AllerTOP v.2.096 and Aller-genFP v.1.096 tools. Vaxijen v2.0 is a freely available server that can predict the antigenicity of a sequence on the basis of auto- and cross-covariance (ACC) transformation of protein sequences into uniform vectors of principal amino acid properties. AllerTOP v2.0 classifies the allergens of a protein using amino acid E-descriptors, k nearest neighbor machine learning methods, and the auto- and cross-covariance (ACC) transformation.
Alternatively, AllergenFP is a descriptor-based and alignment-free strategy to recognize allergens and non-allergens. It is essential to determine physical and chemical parameters associated with the vaccine construct. Therefore, various physicochemical properties of the vaccine like instability index, aliphatic index, molecular weight, GRAVY values, isoelectric point, and half-life were evaluated by using the ExPASy ProtParam tool. SignalP 4.1 and TMHMM v2.0 servers were used to check any transmembrane helices in the vaccine and the existence of any signal peptides.

**Analysis of post-translational modification.** In order to analyze post-translational modification in both the target spike glycoprotein and the designed vaccine construct including glycosylation, phosphorylation, and acetylation, we have used YinOYang 1.2, NetNGlyc 1.0, NetPhos 3.0, and NetAct 1.0 servers which are available at http://www.cbs.dtu.dk/services. The NetPhos 3.2 web tool can predict phosphorylation sites at threonine, tyrosine, and serine for amino acid sequences. The NetNGlyc and YinOYang server use an artificial neural network to predict N-terminal glycosylation and O-GlcNAc attachment sites in mammalian protein.

**Molecular docking and refinement.** To calculate the binding affinity and interaction patterns between the designed multi-epitope vaccine and Toll-like receptor 2 (PDB: 2Z7X) and Toll-like receptor 4 (PDB: 3FXI), the structures were retrieved from the RCSB PDB database in pdb format. The PatchDock server was used to perform molecular docking, and the refinement of the best complex was performed by the FireDock server. PatchDock server calculates surface fix coordinating scores, separating scores, and portrayal of atomic shape for protein–protein molecular docking. This algorithm splits the TLRs receptor and the vaccine molecules into small patches in agreement with the surface. These small patches resemble distinctive shapes, which can separate puzzle pieces visually. The lowest docking energy score was used to select the top-ranked vaccine-TLRs complex structure. In addition, molecular interactions were visualized through PyMol and Discovery Studio 2020.

**Molecular dynamic simulation.** The YASARA software package was implemented to perform molecular dynamic simulations of the vaccine/TLR complexes. The AMBER14 force field was used for the system, and the vaccine complex was initially cleaned and optimized for hydrogen bond formation. The simulation system was established with the aid of a cubic simulation cell. For the initial energy minimization process, a simulated annealing method was applied with steep gradient approaches. The system was neutralized by the addition of water molecules and 0.9% NaCl salt. The physiological conditions of the system were set to pH 7.4 and a temperature of 310 K. The system temperature was maintained with a Berendsen thermostat. The Particle Mesh Ewald (PME) method was applied for the calculation of long-range electrostatic interactions and short-range Columb and vdW interactions. A cut-off radius was established at 8 Å. The simulation time step was set to 1.25 fs, and simulation trajectories were saved after every 100-ps interval. Finally, the simulation was performed for 50 ns, and the RMSE, RMSD, Rg, and SASA values and the numbers of hydrogen bonds formed during the trajectories were analyzed to reveal the stabilities of the vaccine complexes.

**In silico immune simulation.** The C-IMMSIM webservice was utilized to perform immune stimulations, assess immunogenicity, and determine the immune response profile for our vaccine. This server is an agent-based model that uses position-specific scoring matrices to predict peptides derived from machine learning techniques for predicting immune interactions. It simulates three compartments which are three anatomical regions in mammals: (1) the bone marrow (the simulation region of hematopoietic stem cell for producing new myeloid and lymphoid cells); (2) the thymus (where native T-cells are selected to avoid autoimmunity); and (3) lymph node. The minimum recommended time between dose 1 and dose 2 for most of the vaccines currently in use, is 4 weeks. The entire simulation ran for 1,400-time steps which are about 15 months (a time step is about 8 h). Two peptide injections were given four weeks apart at time step 10, 94, 178, 262, 346, 430, 514, 598, 682, 766, 850, 934.

**Conclusions**
Multi-epitope vaccines have already gained importance and demonstrated protective efficiency, capable of generating immunity in vivo, with some entering clinical trials. The present study was based on an immunoinformatics-driven method that was used to identify potential antigenic epitopes for use in a vaccine candidate against MERS-CoV. Three antigen categories, including CTL, HTL, and IFN-γ epitopes, of the S glycoprotein, were used to construct a multi-epitope vaccine. The physicochemical and antigenic profiles of the vaccine were studied computationally. The stability profile and molecular interactions between the designed vaccine and immune receptors were assessed through molecular dynamics simulations and molecular docking studies. In silico immune simulation indicated the vaccine’s ability to trigger an immune response. A series of immunoinformatics strategies were applied sequentially to design and evaluate a vaccine that may produce a protective immunity against viral infection, however, the experimental evaluation is required to assess the exact efficiency. The experimental assay may include the synthesis of the vaccine followed by the in vitro and in vivo assay. Additionally, we propose more studies that include synthesis and biological activities of the designed multi-epitope vaccine.

**Data availability**
All data generated or analysed during this study are included in this published article (and its Supplementary Information files).
40. Saadi, M., Karkkänen, A. & Nouri, H. R. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. Infect. Genet. Evol. 51, 227–234 (2017).

41. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. Proteins 57, 702–710. https://doi.org/10.1002/prot.20264 (2004).

42. Huang, Y. J., Mao, B., Arabini, J. M. & Montelione, G. T. Assessment of template-based protein structure predictions in CASP10. Proteins 82(Suppl 2), 43–56. https://doi.org/10.1002/prot.24488 (2014).

43. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407–W410. https://doi.org/10.1093/nar/gkm290 (2007).

44. Messoudi, A., Belguith, H. & Ben Hamida, I. Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase. Theor. Biol. Med. Model. 10, 22. https://doi.org/10.1186/1742-4682-10-22 (2013).

45. Dey, A. K., Malyala, P. & Singh, M. Physicochemical and functional characterization of vaccine antigens and adjuvants. Expert Rev. Vaccines 13, 671–685. https://doi.org/10.1586/14766508.13.3.671 (2014).

46. Walker, J. M. The proteinscopes handbook (Springer, 2005).

47. Carty, M. & Bowie, A. G. Recent insights into the role of Toll-like receptors in viral infection.

48. Lester, S. N. & Li, K. Toll-like receptors in antiviral innate immunity.

49. Memish, Z. A. Physicochemical and functional characterization of vaccine antigens and adjuvants.

50. Valenzuela, M. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

51. Jabbar, B. et al. Physicochemical and functional characterization of vaccine antigens and adjuvants.

52. Lu, C. et al. Evaluation of tandem Chlamydia trachomatis MOMP multi-epitope vaccines in BALB/c mouse model.

53. Jiang, P. et al. Evaluation of tandem Chlamydia trachomatis MOMP multi-epitope vaccines in BALB/c mouse model.

54. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality.

55. He, R. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

56. Cao, Y. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

57. Cao, Y. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

58. Hou, J. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

59. Lennerz, V. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

60. Slingluff, C. L. J. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

61. Toledo, H. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

62. Vaure, C. & Liu, Y. A comparative review of toll-like receptor 4 expression and functionality in different animal species.

63. Bazhan, S. I. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

64. Boehme, K. W. & Compton, T. Innate sensing of viruses by toll-like receptors.

65. Mizel, S. B. & Bates, J. T. Flagellin as an adjuvant: cellular mechanisms and potential.

66. Khatoon, N. et al. Exploring Leishmania secretory proteins to design B and T cell multi-epitope vaccines.

67. Hou, J. et al. Identification of a CD4 T-cell epitope in the hemagglutinin stalk domain of influenza A virus.

68. Kim, H. J., Kim, J.-K., Seo, S. B., Lee, H. J. & Kim, H.-J. Intranasal vaccination with peptides and cholera toxin subunit B as a potent systemic adjuvant for HIV-1 DNA vaccination intramuscularly in mice.

69. Ali, M. et al. Evaluation of tandem Chlamydia trachomatis MOMP multi-epitope vaccines in BALB/c mouse model.

70. Ali, M. et al. Evaluation of tandem Chlamydia trachomatis MOMP multi-epitope vaccines in BALB/c mouse model.

71. Ali, M. et al. Evaluation of tandem Chlamydia trachomatis MOMP multi-epitope vaccines in BALB/c mouse model.

72. Vaure, C. & Liu, Y. A comparative review of toll-like receptor 4 expression and functionality in different animal species.

73. Phongsiray, V., Iizasa, E., Hara, H. & Yoshida, H. Evidence for TLR4 and FcRy-CARD9 activation by cholera toxin B subunit and its direct bindings to TREM2 and LIR1 receptors. Mol. Immunol. 66, 463–471. https://doi.org/10.1016/j.molimm.2015.05.005 (2015).

74. Boehme, K. W. & Compton, T. Innate sensing of viruses by toll-like receptors. J. Virol. 78, 7867–7873. https://doi.org/10.1128/JVI.78.15.7867-7873.2004 (2004).

75. Dar, H. A. et al. Immunoinformatics-aided design and evaluation of a potential multi-epitope vaccine against klebsiella pneumoniae. Vaccines 7, 10. https://doi.org/10.3390/vaccines7030088 (2019).

76. Solanki, V., Tiwari, M. & Tiwari, V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. Sci. Rep. 9, 1–19 (2019).
116. 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* 5, 1–14. https://doi.org/10.1371/journal.pone.0009862 (2010).

117. Castiglione, F., Mantile, F., De Berardinis, P. & Prisco, A. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Comput. Math. Methods Med.* 2012, 842329. https://doi.org/10.1155/2012/842329 (2012).

**Author contributions**

S.M., M.O.R., G.K.P. and M.A.S. together planned and designed the research. M.A.S. and M.A.M. arranged the whole facilities for the research and supervised the whole research. S.M. and M.O.R. imparted in study design and interpreted the results putting efforts on statistical analysis with M.M.P., M.S.S.S., S.B. and T.B.E. S.M., M.O.R., and G.K.P. participated in the manuscript draft and has thoroughly checked. T.B.E., K.D.S.A.A., M.A.M., and M.A.S. revised the manuscript for necessary changes in format, grammar and English standard. All authors read and agreed on the final version of the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-92176-1.

**Correspondence** and requests for materials should be addressed to M.A.M. or M.A.S.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021