Engineering Novel Epitope-based Subunit Vaccine against SARS-CoV-2 by Exploring the Immunoinformatics Approach

Bishajit Sarkar^{1,2}, Md. Asad Ullah^{1,2}, Yusha Araf^{1,3}, Mohammad Shahedur Rahman^{1,2*}

¹COVID Research Cell (CRC), Wazed Miah Science Research Center (WMSRC), Jahangirnagar University, Savar, Dhaka, Bangladesh

²Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh

³Department of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

*For correspondence:

Professor Dr. Mohammad Shahedur Rahman,

Deputy Director

Wazed Mia Science Research Center, Jahangirnagar University,

Savar, Dhaka-1342, Bangladesh.

Email: rahmanms@juniv.edu

Abstract

As the number of infections and deaths caused by the recent COVID-19 pandemic is increasing dramatically day-by-day, scientists are rushing towards developing possible counter-measures to fight the deadly virus, SARS-CoV-2. Although many efforts have already been put forward for designing and developing potential vaccines, however, most of them are proved to possess negative consequences. Therefore, in this study, the methods of immunoinformatics were exploited to design novel epitope-based subunit vaccine against the SARS-CoV-2, targeting four essential proteins of the virus i.e., spike glycoprotein, nucleocapsid phosphoprotein, membrane glycoprotein, and envelope protein. The highly antigenic, non-allergenic, non-toxic, non-human homolog and 100% conserved (across other isolates from different regions of the world) epitopes were used for constructing the vaccine. In total, fourteen CTL epitopes and eighteen HTL epitopes were used to construct the vaccine. Thereafter, several *in silico* validations i.e., the molecular docking, molecular dynamics simulation (including the RMSF and RMSD studies), and immune simulation studies were also performed which predicted that the designed vaccine should be quite safe, effective, and stable within the biological environment. Finally, *in silico* cloning and codon adaptation studies were also conducted to design an effective mass production strategy of the vaccine. However, more *in vivo* and *in vitro* studies are required on the predicted vaccine to finally validate its safety and efficacy.

Keywords: COVID-19; SARS-CoV-2; *In silico*; Immunoinformatics; Vaccine designing

1. Introduction

The Coronavirus Disease-2019 or COVID-19 is caused by a virus known as the Severe Acute Respiratory Syndrome Coronavirus- 2 (SARS-CoV-2). This disease was originated in Wuhan City of Hubei Province of China in December, 2019. But now, its spread is abruptly accelerated on a worldwide scale from its origin (Ullah et al., 2020a, Islam et al., 2020). Coronaviruses (CoVs) are a group of immensely diversified RNA viruses with some general features i.e. these are enveloped and positive-sense viruses, containing ssRNA (single-stranded RNA). They are responsible for several diseases of respiratory, hepatic, neurological as well as enteric systems. Actually, the severity of their infection varies from animal to human (Zumla et al., 2016; Chan et al., 2013). As reported by serology and genome phylogeny, CoVs can be categorized into four genera named as Alpha, Beta, Gamma, and Delta CoVs and to date, seven strains of CoVs are discovered that are known to infect human. This group of Human Coronaviruses (HCoVs) includes two alpha CoVs (HCoV-229E and HCoV-NL63) and five beta CoVs (HCoV-OC43, HCoV-HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2 (Zhu et al., 2018). But most importantly, the SARS-CoV, MERS-CoV, and SARS-CoV-2, these three HCoVs have appeared during the last two decades (Cheng et al., 2007; Chan et al., 2015). Although the SARS-CoV and MERS-CoV epidemics didn't have much impact worldwide, however, the recent COVID-19 pandemic causing SARS-CoV-2 has already caused much sufferings and claimed the lives of hundreds of thousands of people, let alone infecting millions of people around the world. As of August 14, 2020, the COVID-19 has disseminated in around 213 countries and territories all over the world. Not only that, it has put an end to nearly 770,000 lives with almost 21 million infected cases. (https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1? Accessed on: 14 August 2020).

Perhaps the development and evaluation of specified drugs for the treatment of COVID-19 disease are going to take couple of years. Regardless of this, broad range of existing host-directed therapies are in course of investigation at present (Amanat and Krammer, 2020). Apace with drug, at least 40 amongst the prestigious pharmaceutical and research institutions from numerous countries have demonstrated their engagement in developing COVID-19 vaccines passionately after the sequenced genome of SARS-CoV-2 has been unraveled and released officially on 11 January 2020. Moreover, some institutions have already started efficacy evaluation of their potential candidates in animals as well as clinical trials (Zhang et al., 2020, Moin et al., 2020). For example, safety clinical trials have been conducted on two vaccines i.e. Moderna's candidate vaccine mRNA-1273 and DNA vaccine called INO-4800. But both these vaccines are still not evaluated for efficacy thoroughly (Inovio Pharmaceuticals, 2020; Kaiser Permanente Washington Health Research Institute, 2020). Moreover, the Phase I clinical trial of the most current vector-based vaccine, LV-SMENP DC engaging 100 patients was performed on March 24, 2020. But the completion date of the study is estimated to be December 31, 2024 (NCT04276896) (<http://clinicaltrials.gov/>. Accessed on: 16 August 2020). Furthermore, a conventional vaccination strategy has also been displayed using whole inactivated or live-attenuated virus vaccines. Researchers from University of Hong Kong have come up with a live influenza vaccine which expresses proteins of SARS-CoV-2 (Cheung, 2020). Besides these, "Codon de-optimization" technology has been developed by Codagenix. Basically, this technology attenuate viruses and the company is now probing COVID-19 vaccination strategies (Shieber J, 2020). Nonetheless, the main drawback of attenuated vaccines is that secondary mutations can promote virulence during reversion and can give rise to more serious conditions. In contrast with such vaccines, the prime advantage of subunit vaccines is they are safe because the constituents of such vaccines are only

the recombinant proteins or synthetic peptides from any infectious virus or pathogen. These vaccines do not include the whole infectious agent. For this reason, when it comes to subunit vaccines, the probability of inducing any adverse effect after administration is nearly zero (Sarkar et al., 2019).

In this study, an effective subunit vaccine was focused to develop against various isolates of SARS-CoV-2 from different countries across the world, utilizing the immunoinformatics approach. In immunoinformatics, the novel antigens of a pathogen or virus are identified by dissecting its genomic data and then different tools of *in silico* biology and bioinformatics are used for vaccine designing and development by analyzing its genome (Chong & Khan, 2019; Rappuoli, 2000; Sarkar et al., 2020a). In our study, a blue print of epitope-based vaccine was designed which might produce substantial immune response towards SARS-CoV-2, isolated from different countries around the world, targeting the spike glycoprotein, nucleocapsid phosphoprotein, membrane glycoprotein, and envelope protein of the virus. The protein sequences of SARS-CoV-2, isolated from Bangladesh, was used as model to construct the vaccine. Only those epitopes which were found to be 100% conserved in some other selected countries were used for vaccine construction. As a result, the designed vaccine is expected to be effective in different regions around the world.

CoV spike glycoproteins promote and facilitate the entry of the virus into the host cells and these proteins are the prime target of antibodies. The nucleocapsid phosphoprotein is vital for packaging the viral genome into a helical ribonucleocapsid (RNP) and it plays an elementary role during viral self-assembly. Also, the membrane and envelope proteins are important for viral entry, replication, budding, and particle assembly within the host cells (Schoeman and Fielding, 2019; Alsaadi and Jones, 2019). Therefore, these four proteins were used as potential targets in this study to design

the vaccine with the purpose of interfering the viral life cycle. **Figure 01** represents the step-by-step procedure used in this study to design the vaccine.

Figure 01. The step-by-step procedure adapted in the vaccine designing study.

2. Materials and Methods

2.1. Strain identification and Retrieval of the Protein Sequences

The SARS-CoV-2 virus isolated from Bangladesh was identified and four target proteins of the virus i.e. spike glycoprotein, nucleocapsid phosphoprotein, membrane glycoprotein, and envelope protein were retrieved from the National Center for Biotechnology Information or NCBI (<https://www.ncbi.nlm.nih.gov/>) database.

2.2. Antigenicity Prediction and Physicochemical Property Analysis of the Proteins

The antigenicity of the retrieved protein sequences of SARS-CoV-2 was predicted by the online tool, VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), using the prediction accuracy threshold of 0.4 because the 0.4 threshold has been proved to increase the prediction accuracy. The server uses different statistical methods to provide reliable results in terms of antigenicity (Doytchinova and Flower, 2007, Ullah et al., 2020b). Thereafter, various physicochemical properties of the selected antigenic protein sequences were predicted by the ExPASy's online tool ProtParam (<https://web.expasy.org/protparam/>), where all the parameters were also kept at their default values (Gasteiger et al., 2005).

2.3. Prediction of T-cell Epitopes

An effective multi-epitope subunit vaccine must comprise of cytotoxic T-cell and helper T-cell epitopes so that after the administration, the vaccine would be able to stimulate the immune cells and generate substantial immune responses (Chaudhri et al. 2009; Zhu and Paul, 2008). The MHC class-I epitopes of the selected protein sequences were predicted using the online epitope prediction server, NetCTL 1.2 (<http://www.cbs.dtu.dk/services/NetCTL/>) (Larsen et al., 2007). This server uses the NetCTL 1.2 prediction method for predicting the possible epitopes from a given query protein sequence. And the MHC class-II epitopes were predicted using another online epitope prediction server, Immune Epitope Database or IEDB (<https://www.iedb.org/>). The IEDB server houses a huge amount of data on antibody and T cell epitopes, experimented in humans,

non-human primates, and other animal species in terms of infectious disease, allergy, auto-immunity, and transplantation (Vita et al., 2018, Sarkar et al., 2020b). The MHC class-II restricted CD4⁺ helper T-lymphocyte (HTL) epitopes were obtained for the full HLA set, using the IEDB recommended 2.22 prediction method. The top MHC class-I and MHC class-II epitopes which were found to be common for all the selected reference HLA alleles were taken for further analyses.

2.4. Antigenicity, Allergenicity, Toxicity, and Transmembrane Topology Determination

In the study, a few criteria were set to select the most promising epitopes from all the epitopes predicted by the NetCTL 1.2 and IEDB servers. Only those epitopes that were found to be highly antigenic, non-allergenic, non-toxic, 100% conserved (among the selected sequences from different countries around the world), and non-homolog to the human proteome, were considered as the most promising or best selected epitopes and only these epitopes were used in the vaccine construction process. The antigenicity of the selected epitopes was determined using the VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) server again, keeping the prediction accuracy parameter threshold 0.4 to get better predictions. Thereafter, the allergenicity of the selected epitopes was predicted using two different online tools, AllerTOP v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) and AllergenFP v1.0 (<http://ddg-pharmfac.net/AllergenFP/>) to get better prediction accuracy. However, the results generated by AllerTOP v2.0 were given priority since the server has better prediction accuracy of 88.7% than AllergenFP server (87.9%) (Dimitrov et al., 2014; Dimitrov et al., 2013). After that, the toxicity prediction of the epitopes was conducted using the ToxinPred server (<http://crdd.osdd.net/raghava/toxinpred/>) and finally, the transmembrane topology of the selected

epitopes was determined using the transmembrane topology of protein helices determinant, TMHMM v2.0 server (<http://www.cbs.dtu.dk/services/TMHMM/>) (Gupta et al., 2013).

2.5. Conservancy and Human Homology Prediction of the Epitopes

The conservancy analysis of the selected epitopes was performed using the ‘epitope conservancy analysis’ module of the IEDB server (<https://www.iedb.org/conservancy/>) (Bui et al., 2007). The epitopes, that were found to be 100% conserved among the selected isolates of SARS-CoV-2 from different countries across the world, were considered for vaccine construction (along with some mentioned criteria) because this will ensure the broad-spectrum activity of the designed vaccine over the selected isolates and so the vaccine would also be effective at different regions and countries of the world. Thereafter, the homology of the epitopes to the human proteome was determined to find out any epitope that had homology with the human proteome. Only the non-homolog epitopes were taken because if any epitope is a homolog to the human proteome, then that epitope might not have the capability to induce a potent immune response as a foreign antigenic particle. The protein BLAST module (blastP) of BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) tool was used in the human homology determination, where *Homo sapiens* (taxid: 9606) was used as the comparing organism, keeping all other parameter default. An e-value cut-off of 0.05 was set in the experiment and the epitopes that had no hits below the e-value inclusion threshold were selected as non-homologous peptides (Mehla and Ramana, 2006).

2.6. Cytokine Inducing Ability Prediction of the Epitopes

The helper T-cells generate several types of cytokines including IFN-gamma, IL-4 (interleukin-4), and IL-10 (interleukin-10) to activate and stimulate different types of immune cells i.e. cytotoxic T-cells, macrophages, B-cells etc. (Luckheeram et al. 2012). So, prediction of the cytokine

inducing capability of the HTL or MHC class-II epitopes is an important criteria for vaccine construction. The IFN-gamma inducing capability of the predicted HTL epitopes was determined using IFNepitope (<http://crdd.osdd.net/raghava/ifnepitope/>) server. The Hybrid (Motif + SVM) prediction approach was used for determining the IFN-gamma inducing ability which is one of the highly accurate approaches for the prediction of IFN-gamma inducing capability of the epitopes (Dhanda et al., 2013a). Furthermore, IL-4 and IL-10 inducing properties of the HTL epitopes were determined using IL4pred (<https://webs.iitd.edu.in/raghava/il4pred/index.php>) and IL10pred (<http://crdd.osdd.net/raghava/IL-10pred/>) servers, respectively (Dhanda et al. 2013b; Nagpal et al., 2017). SVM method was used in both the IL4pred and IL10pred predictions keeping the threshold at default values.

2.7. Population Coverage Analysis

To design a multi-epitope vaccine, it is a crucial prerequisite to consider the distribution of specific HLA alleles among different population and ethnicities around the world because the expression of different HLA alleles may vary from one population to another. The IEDB population coverage tool (<http://tools.iedb.org/population/>) was used to determine the population coverage of the most promising epitopes across different HLA alleles in different regions around the world (Bui et al., 2007; Adhikari & Rahman, 2017).

2.8. Cluster Analysis of the MHC Alleles

Cluster analysis of the MHC alleles was carried out in order to identify of the relationship of the MHC class-I and class-II alleles used in the study. The cluster analysis of the MHC alleles was conducted using the online tool MHCcluster 2.0 (<http://www.cbs.dtu.dk/services/MHCcluster/>) (Thomsen et al., 2013). During the analysis, all the parameters of the server were kept default and

all the used HLA super-type representatives (MHC class-I), and HLA-DR representatives (MHC class-II) were selected.

2.9. Vaccine Construction

In this step, the most promising epitopes from 2.4 subsection were conjugated with one another for constructing a possible vaccine. Human beta-defensin-3, as an adjuvant sequence, was connected to the epitopes by EAAAK linkers. Adjuvants are important to enhance the antigenicity, immunogenicity, stability, and longevity of the constructed vaccines (Meza et al., 2017; Lee and Nguyen, 2015). The pan HLA-DR epitope (PADRE) sequence was also associated with the adjuvant and epitopes. The PADRE sequence provokes the immune response by improving the capacity of the CTL epitopes of the vaccines (Pandey et al., 2016, Wu et al., 2010). Thereafter, the AAY and GPGPS linkers were used to conjugate the CTL and HTL epitopes. The EAAAK linkers are proved to provide a partition of the domains of a bifunctional fusion protein (Arai et al., 2001), whereas the GPGPS linkers are used to generate the junctional epitopes and also to enhance the immune processing and presentation (Saadi et al., 2017). And the AAY linker is also used widely in the *in silico* vaccine designing experiments because this linker conjugates the epitopes effectively and efficiently (Yang et al., 2015). **Figure 02** represents a graphical illustration of how the vaccine was designed in this study.

Figure 02. Schematic representation of the possible vaccine construct with linkers (EAAAK, AAY, and GPGPG), PADRE sequence, adjuvant (human beta-defensin-3) and epitopes (CTL, HTL) in sequential and appropriate manner.

2.10. Antigenicity, Allergenicity, and Physicochemical Property Analysis

The antigenicity of the constructed vaccine was predicted by the online server VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), keeping the prediction threshold set at 0.4 (Doytchinova and Flower, 2007). Thereafter, both AlgPred (<http://crdd.osdd.net/raghava/algpred/>) and AllerTop v2.0 (<https://www.ddg-pharmfac.net/AllerTOP/>) servers were used to determine the allergenicity of the vaccine construct (Saha and Raghava, 2006). MEME/MAST motif prediction approach was used in the allergenicity prediction of the vaccine by this server. After that, various physicochemical properties of the vaccine were analyzed by the online server, ProtParam (<https://web.expasy.org/protparam/>) (Gasteiger et al., 2005). Finally, the solubility of the vaccine construct was determined by the SOLpro module of the SCRATCH protein predictor (<http://scratch.proteomics.ics.uci.edu/>) and further clarified by the Protein-sol server (<https://protein-sol.manchester.ac.uk/>). Both these servers use machine learning techniques to provide much reliable predictions. During conducting the predictions, all the parameters were kept at their default values (Magnan et al., 2009; Hebditch et al., 2017).

2.11. Secondary and Tertiary Structure Prediction of the Vaccine Constructs

After the physiochemical analysis of the vaccine, it was subjected to the secondary structure prediction. Several online tools i.e., PRISPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) (using PRISPRED 4.0 prediction method), GOR IV (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) and SIMPA96 (https://npsaprabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_simpa96.html) were used to run the prediction keeping all the parameters default (Buchan and Jones, 2019; Jones, 1999; Garnier et al., 1996; Geourjon and Deleage, 1995; Montgomerie et al., 2006; Levin et al., 1986). After that, the tertiary or 3D structure of the vaccine was predicted using RaptorX (<http://raptorx.uchicago.edu/>) online server. The server predicts the tertiary or 3D structure of a protein using an efficient template-based method (Källberg et al., 2012).

2.12. 3D Structure Refinement and Validation

The 3D structures generated by computational methods may not provide the true or native structures of the proteins. Therefore, the 3D structure refinement is performed to enhance the resolution of the computationally predicted models so that they can closely resemble the native protein structures. The generated 3D structure of the designed vaccine was refined by the GalaxyRefine module of the GalaxyWEB server (<http://galaxy.seoklab.org/>), which uses CASP10 tested refinement method for protein structure refinement (Nugent et al., 2014; Ko et al., 2012). After refining the structure, validation process of the vaccine construct was carried out by analyzing the Ramachandran plot, generated by PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>) server (Laskowski et al., 2006; Morris et al., 1992) and z-score provided by another online tool, ProSA-

web (<https://prosa.services.came.sbg.ac.at/prosa.php>). A z-score within the range of the z-scores of all the experimentally determined protein chains in the PDB database corresponds to a higher quality of a query protein (Wiederstein and Sippl, 2007).

2.13. Vaccine Protein Disulfide Engineering

The vaccine protein disulfide engineering was conducted by the online tool, Disulfide by Design -2 v12.2 (<http://cptweb.cpt.wayne.edu/DbD2/>) (Craig and Dombkowski, 2013). The server determines the possible sites within a protein structure which have the greater possibility to form the disulfide bonds. During the prediction, the intra-chain, inter-chain, and C β for glycine residue were selected. The χ_3 and C α -C β -S γ angles were kept at -87° or $+97^\circ \pm 5$ and $114.6^\circ \pm 10$, respectively and the amino acid pairs with less than 2.2 kcal/mol bond energy were selected for mutation into cysteine residues to form the disulfide bonds among themselves. The 2.2 kcal/mol was used as a threshold in this study because almost 90% of the naturally formed disulfide bonds have energy value of less than 2.2 kcal/mol (Petersen et al., 1999; Craig and Dombkowski, 2013).

2.14. Protein-Protein Docking

In protein-protein docking, the vaccine construct was docked against multiple TLRs. It's highly important for a vaccine to have a good binding affinity with the TLRs because, TLR proteins generate potential immune responses after recognizing the vaccines which mimic the original viral infections. Thus, they facilitate to produce a strong immunity towards a particular virus or pathogen (Stern LJ and Calvo-Calle, 2009). In this study, the vaccine construct was docked with TLRs i.e. TLR-1 (PDB ID: 6NIH), TLR-2 (PDB ID: 3A7C), TLR-3 (PDB ID: 2A0Z), TLR-4 (PDB ID: 4G8A), TLR-8 (PDB ID: 3W3M), and TLR9 (PDB ID: 4QDH). The protein-protein docking was conducted using three different online tools to improve the prediction accuracy. Initially, the

ClusPro v2.0 (<https://cluspro.bu.edu/login.php>) was used for docking where the lower energy score corresponds to the better binding affinity (Vajda et al., 2017; Kozakov et al., 2017; Kozakov et al., 2013). The ClusPro server calculates the energy score based on the following equation:

$$E = 0.40E_{rep} + (-0.40E_{att}) + 600E_{elec} + 1.00E_{DARS} \quad (\text{Kozakov et al., 2017; Kozakov et al., 2013})$$

Then the docking was again performed using PatchDock (<https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php>) server and the results were refined by FireDock server (<http://bioinfo3d.cs.tau.ac.il/FireDock/php.php>). Both these servers are easy and user-friendly servers that provide much reliable results of molecular docking (Duhovny et al., 2002; Atapour et al., 2019; Schneidman-Duhovny et al., 2005). Thereafter, the final round of docking was performed by the HawkDock server (<http://cadd.zju.edu.cn/hawkdock/>) along with the Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) study (Weng et al., 2019). Then the vaccine docked with the TLR (with the highest binding affinity) was visualized using Discovery Studio Visualizer (Biovia et al., 2016).

2.15. Conformational B-cell Epitope Prediction

The humoral immunity, along with the cell mediated immunity, is very important to fight the pathogens inside the body. The humoral immunity of the body depends on the B-cells that produce antibodies when they encounter an antigen. Therefore, the constructed vaccine should have effective conformational B-cell epitopes so that it can provide more potent immunity. The conformational B-cell epitopes of the constructed vaccine protein were predicted by IEDB ElliPro tool (<http://tools.iedb.org/ellipro/>), keeping all the parameters default (Ponomarenko et al., 2008).

2.16. Molecular Dynamics (MD) Simulation

For observing the state changes and effects of the environment on the crystalized structure of the vaccine protein as well as protein-TLR complex structure, MD simulation was carried out. The

GROMACS (GRONingen MACHine for Chemical Simulations) a Linux based command-line program had been used for this purpose (Abraham et al., 2015). Two separate simulations were run for crystalized and complex protein structures. For conducting the two simulations, at first the crystalized and complex protein 'pdb' files were cleaned to remove any environmental substrates generated by the server. After that the pdb2gmx was run using OPLS-AA (Optimized Potential for Liquid Simulation-All Atom) force field to generate the topology for both the structures (Kaminski et al., 2001). The structures were positioned in the center of a 2nm sized cube, being 1nm from each edge during their simulations. Solvation was performed by filling the box with water molecules spatially placed with a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Then the energy minimization had been conducted to stabilize the structures. From **Supplementary Figure S1**, we can see that the potential energy of the structure had quickly reduced below the order of 10^6 for the protein complex. From this, we inferred that the complex system was stable enough to conduct further simulations. NVT (Number Volume Temperature) equilibration had been performed for 100 ps to stabilize the temperature. Afterwards, the NPT (Number Pressure Temperature) equilibration was done for 100 ps and pressure as well as density were calculated. The crystalized protein structure was also prepared in the same way. Then the resulting stabilized structures had then been subjected to MD simulation of 20 ns. The RMSD, RMSF of backbone of the energy minimized structures was predicted and radius of gyration was also calculated for the two structures. All plots and simulation graphs had been analyzed using the Xmgrace and QtGrace tool (Turner, 2005). Finally, the two simulations were compared with each other.

2.17. Immune Simulation

The immune simulation study of the constructed vaccine was carried out to predict its immunogenicity and immune response profile using the C-ImmSim server

(<http://150.146.2.1/CIMMSIM/index.php>). The server predicts the real-life-like immune interactions by means of machine learning techniques and position-specific scoring matrix (PSSM) (Rapin et al., 2010). All the parameters except the time steps were kept at their default values during the experiment. However, the time steps were kept at 1, 84, and 170, and the number of simulation steps was set at 1050. So, three injections ought to require four weeks apart because the recommended interval between two doses of most of the commercial vaccines is proved to be four weeks (Castiglione et al., 2012). The Simpson's Diversity index, D was calculated from the figures.

2.18. Codon Adaptation and *In Silico* Cloning

Codon adaptation and *in silico* cloning are two essential steps in vaccine designing by immunoinformatics. An amino acid can be encoded by more than one codon in different organisms because the cellular machinery of an organism can be completely different from another organism, this phenomenon is known as codon bias. Therefore, codon adaptation is performed to predict the suitable codon that efficiently encodes a specific amino acid in a particular organism. In the codon adaptation study, the vaccine protein was reverse translated to the possible DNA sequence which was expected to encode the amino acids of the designed vaccine protein (Khatoon et al., 2017; Angov, 2011). The Java Codon Adaptation Tool or JCat server (<http://www.jcat.de/>) was used for the codon adaptation study of the constructed vaccine (Grote et al., 2005). In the server, the prokaryotic *E. coli* strain K12 was selected as the target organism and rho-independent transcription terminators, prokaryotic ribosome binding sites, and EaeI and StyI cleavage sites of restriction enzymes were avoided at the server. Finally, the SnapGene restriction cloning software was used for inserting the newly adapted DNA sequence between the EaeI and StyI restriction sites of the pETite vector (Solanki and Tiwari, 2018). The pETite vector plasmid contains

ubiquitin-like modifier or SUMOtag and 6X-His tag which facilitates the solubilization and viable affinity purification of the recombinant protein (Biswal et al., 2015).

2.19. Analysis of the Vaccine mRNA

The mRNA secondary structure prediction was carried out using two different online tools i.e., Mfold (<http://unafold.rna.albany.edu/?q=mfold>) and RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RFold.cgi>). Both servers predict the mRNA secondary structures thermodynamically and generate minimal free energies for the structures (Zuker et al., 2003; Mathews et al., 1999; Mathews et al., 2007; Gruber et al., 2008). To analyze the mRNA folding and secondary structure of the vaccine, the optimized DNA sequence from the JCat server was converted to the possible RNA sequence by the DNA<->RNA->Protein tool (<http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm>). Then the RNA sequence was pasted in the Mfold and RNAfold servers for lower minimum free energy prediction using the default parameters.

3. Results

3.1. Identification, Selection and Retrieval of Viral Protein Sequences

The four proteins i.e. spike glycoprotein, nucleocapsid phosphoprotein, membrane glycoprotein, and envelope Protein of the SARS-CoV-2 of Bangladeshi isolate were selected and then the protein sequences were retrieved from the NCBI database. **Table 01** lists the GenBank accession IDs of the selected protein sequences.

Table 01. List of the proteins used in this study with their GenBank accession numbers.

No	Virus Strain	Name of the protein	GenBank accession number
01	SARS-CoV-2	Spike glycoprotein	QLF97699.1
02		Nucleocapsid phosphoprotein	QLF97707.1
03		Membrane Glycoprotein	QLF97702.1
04		Envelope Protein	QLF97701.1

3.2. Antigenicity Prediction and Physicochemical Property Analysis of the Protein Sequences

All the query proteins were found to be antigenic in the physicochemical property analysis. Again, all of them had a similar half-life of 30 h in the mammalian cell culture system but only the nucleocapsid phosphoprotein was predicted to be unstable, having instability index over 40. All the proteins were also found to have quite high aliphatic indexes. Furthermore, the spike glycoprotein had the lowest GRAVY value of -0.077. The results of the physicochemical property analysis are listed in **Supplementary Table S1**.

3.3. Epitope Prediction and Filtering the Most Promising Epitopes

The MHC class-I and class-II epitopes were predicted from the target protein sequences for constructing the vaccine. Since the servers generated a lot of possible epitopes, so some criteria i.e., high antigenicity, non-allergenicity, non-toxicity, non-human homology, 100% conservancy

(across other selected isolates of SARS-CoV-2), were used as filters. The epitopes that were found to follow all these criteria, were finally selected for vaccine construction considering as the most promising epitopes (**Table 02**). The 100% conservancy of the epitopes among the selected isolates from different regions of the world ensured that this possible vaccine should be effective in controlling the SARS-CoV-2 infections in various regions around the world along with Bangladesh. **Supplementary Table S2** lists the potential epitopes of spike glycoprotein, **Supplementary Table S3** lists the potential epitopes of nucleocapsid phosphoprotein, and **Supplementary Table S4** and **Supplementary Table S5** list the potential T-cell epitopes of membrane glycoprotein and envelope protein, respectively. **Supplementary Table S6** lists the proteins from different countries (with their GenBank IDs) which were used in the conservancy analysis.

Table 02. List of the epitopes finally selected for vaccine construction (selection criteria: antigenicity, non-allergenicity, non-toxicity, 100% conservancy and non-homolog to the human proteome).

Protein name	MHC class-I epitopes	MHC class-II epitopes
Spike Glycoprotein	VLPFNDGVY	QSLIVNNATNVVIK
	WTAGAAAYY	VLSFELLHAPATVCG
	GAAAYYVGY	VVLSFELLHAPATVC
	QLTPTWRVY	-
	STECNLLL	-
	VLKGVKLHY	
Nucleocapsid phosphoprotein	SSPDDQIGY	IAQFAPSASAFFGMS
	SPDDQIGYY	GTRNPANNAIIVLQL
	DLSPRWYFY	-
	LSPRWYFYY	-
Membrane Glycoprotein	LVGLMWLSY	IKLIFLWLLWPVTLA
	ATSRTLSTYY	VGLMWLSYFIASFRL
	AGDSGFAAY	KLIFLWLLWPVTLAC
Envelope Protein	VSLVKPSFY	LLFLAFVVFLVTLA
	-	VLLFLAFVVFLVTL
	-	LFLAFVVFLVTLAI
	-	LLFLAFVVFLVTLA
	-	LFLAFVVFLVTLAI
	-	NSVLLFLAFVVFLV
	-	SVLLFLAFVVFLVT
	-	VLLFLAFVVFLVTL
	-	FLAFVVFLVTLAIL
	-	LAFVVFLVTLAILT
	-	

3.4. Cytokine Inducing Ability Prediction of the Epitopes

IFN-gamma, IL-4, and IL-10 inducing capacity prediction of the HTL epitopes showed that many of the selected HTL epitopes had the capability of inducing at least one of these cytokine. Moreover, all the most promising epitopes were also found to have at least one cytokine production capability (**Supplementary Table S2, S3, S4, and S5**).

3.5. Population Coverage Analysis

The population coverage analysis showed that the MHC class-I and class-II epitopes covered 90.42% and 93.23%, respectively and the combined MHC class-I and class-II covered 84.56% of the world population. India was found to possess the highest percentage of population coverage of both the MHC class-I epitopes (93.49%) and MHC class-II epitopes (91.20%). And the highest percentage of population coverage of the MHC class-I and class-II epitopes in combination was obtained by China (85.67%) (**Supplementary Figure S2**).

3.6. Cluster Analysis of the MHC Alleles

The cluster analysis of the possible MHC class-I and MHC class-II alleles, that may interact with the predicted epitopes of the selected proteins, was carried out using the online tool MHCcluster 2.0. The tool demonstrates the relationship of the clusters of the alleles in a phylogenetic manner. **Supplementary Figure S3** illustrates the result of the experiment where the red zone indicates a strong interaction and the yellow zone represents a weaker interaction.

3.7. Vaccine Construction

The vaccine has been constructed which could be used to fight against the selected viral strains effectively. Beta-defensin-3 was used as adjuvant and the PADRE sequence was also used for the vaccine construction. EAAAK, AAY, and GPGPG linkers were used at their appropriate positions

during conjugating the epitopes. The newly constructed vaccine were designated as: CV (Table 03)

Table 03. Vaccine constructed against the SARS-CoV-2. The bolded letters represent the linker sequences.

Name of the vaccine	Vaccine Construct
“CV” vaccine	EAAAK GIINTLQKYYCRVRGGRC AVL SCLPK EEQIGK CSTRGRKCCRRKKE EAAAK AKFVAAWTLKAAAAAYVLPFNDGVYAAAYWTAGAAAYYAAYGAAAYYVG YAAAYQLTPTWRVYAAAYSTEC SNLLLAAYVLKGVKLHYAAAYSSPDDQIGYAAAYSPDDQIGYYAAAYDLSPRWYFYAAAYLSPRWYFYAAAYLVGLMWLSYAAAYATSRTL SYAAAYAGDSGF AAYYAAAYVSLVKPSFY GPGPG QSLIVNNATNVVIK GPGPG VLSFELLHAPATVCG GPGPG VVLSFELLHAPATVCG GPGPG IAQFAPSASAFFGMS GPGPG GTRNPANNA AIVLQ LGP GPG IKLIFLWLLWPVTLA GPGPG VGLMWLSYFIASFRL GPGPG KLIFLWLLWPVTLAC GPGPG LLFLAFVVFLLVTLA GPGPG VLLFLAFVVFLLVTL GPGPG LFLAFVVFLLVTLA I GPGPG LLFLAFVVFLLVTLA GPGPG LFLAFVVFLLVTLA I GPGPG NSVLLFLAFVVFLLV GPGPG SVLLFLAFVVFLLV TGPGPG VLLFLAFVVFLLVTL GPGPG FLAFVVFLLVTLA I L VT L A I L GPGPG LAFVVFLLVTLA I L VT L A I L TGPGPG

3.8. Antigenicity, Allergenicity, and Physicochemical Property Analysis of the Vaccine Constructs

The CV vaccine was predicted to be a potent antigen as well as a non-allergen. In the physicochemical property analysis of the vaccine, a high (basic) theoretical pI, half-life of 1 h in the mammalian cells, and more than 10 h in the *E. coli* cell culture system were predicted. Moreover, the vaccine construct was also found to be quite stable as well as soluble upon overexpression in *E. coli* cell culture system. Furthermore, it had an aliphatic index of 114.48 and a quite low GRAVY value of -0.884 (**Table 04**).

Table 04. Antigenicity, allergenicity and physicochemical property analysis of the vaccine construct. AN; antigenicity, AG; allergenicity, II; instability index, AI; aliphatic index, GRAVY; grand average of hydropathicity.

Name of the protein	AN	AG	pI	Total number of negatively charged residues	Total number of positively charged residues	Ext. coefficient	Half-life	II	AI	GRAVY	Solubility
CV	Antigenic	Non-allergenic	9.22	14	29	117745	1 hours (mammalian reticulocytes, 20 min (yeast, >10 hours (Escherichia coli	Stable (31.49)	114.48	- 0.884	Soluble (SolPro: 0.760, Protein-Sol: 0.668)

495

3.9. Secondary and Tertiary Structure Prediction of the Vaccine Constructs

In the secondary structure analysis of the CV vaccine, the amino acid percentage of α -helix, β -strand, and coil structure of the vaccine protein was predicted using four different online servers. The vaccine had the highest percentage of the amino acids in the coil structure, considering the results from all the servers. The amino acid percentages of the vaccine are given with a tabular representation (**Supplementary Table S7 and Supplementary Figure S4**).

The 3D structure of the SV vaccine construct was predicted by the online server RaptorX. The vaccine had 5 domains with quite low p-value of 3.56e-08, which declared that the quality of the

prediction 3D structure was quite good. The homology modeling of the vaccine construct was performed using 1KJ6A as template from the Protein Data Bank. The 3D structure of SV is illustrated in **Figure 03**.

Figure 03. The tertiary structure of the CV vaccine.

3.10. Protein 3D Structure Refinement and Validation of the Vaccine

The protein structure generated by the RaptorX server was subjected to the refinement process using the Galaxy-web server, which was then analyzed by Ramachandran plot generated by the PROCHECK server and the z-score predicted by the ProSA-web server. The Ramachandran plot analysis depicted that CV vaccine had 86.2% of the amino acids in the most favored region, 11.7% of the amino acids in the additional allowed regions, 0.8% of the amino acids in the generously

allowed regions, and 1.3% of the amino acids in the disallowed regions. Moreover, CV had the z-score of -5.75 which lies within the range of experimentally proven X-ray crystal structures of proteins from the Protein Data Bank (**Supplementary Figure S5**).

3.11. Vaccine Protein Disulfide Engineering

In protein disulfide engineering, the DbD2 server is used which identifies the pairs of amino acids that have the capability to form disulfide bonds. In this experiment, only those amino acid pairs were selected which had a bond energy value of less than 2.2 kcal/mol. The CV generated four pairs of amino acids with bond energy less than 2.2 kcal/mol i.e. 93 Ala and 431 Phe, 369 Tyr and 396 Cys, 435 Thr and 528 Ala, 459 Gly and 468 Phe. The selected amino acid pairs formed the mutant version (with disulfide bonds) of the original vaccine in the DbD2 server (**Supplementary Figure S6**). Since CV was predicted to have four possible pairs of amino acid residues with the capability to form potential disulfide bonds, therefore, it can be considered as a quite stable vaccine construct.

3.12. Protein-Protein Docking Study

The protein-protein docking study was performed to analyze the ability of the designed vaccine to interact with different TLRs which would occur during an actual immune response. CV vaccine showed a very high binding affinity when docked by ClusPro 2.0. For better prediction, it was further docked using two other servers i.e. PatchDock and HawkDock servers, respectively. ClusPro 2.0 server showed the highest binding affinity with TLR-8 (-1261.2 kcal/mol). Again, when analyzed by PatchDock and FireDock server, CV vaccine again showed the best global energy score with TLR-8 (-32.29 kcal/mol). Furthermore, the docking and MM-GBSA analyses by HawkDock server also predicted the highest binding affinity of the vaccine protein with TLR-8 (**Table 05**). Since all of the servers predicted very good binding affinities of the CV vaccine with

the TLR-8, therefore, the visualization of the interaction (**Figure 04**) and molecular dynamics (MD) simulation was performed only for the CV-TLR-8 docked complex.

Table 05. Results of molecular docking study of the CV vaccine interacted with the TLR-8.

Target TLRs (with PDB IDs)	ClusPro energy score	Global energy (PatchDock server)	HawkDock score (the lowest score)	MM-GBSA (binding free energy, in kcal mol ⁻¹)
TLR-1 (6NIH)	-1131.7	-8.30	-5701.54	-43.28
TLR-2 (3A7C)	-1021.3	-24.33	-4102.30	-66.69
TLR-3 (2A0Z)	-923.5	-11.73	-3708.21	-49.22
TLR-4 (4G8A)	-955.1	-2.92	-5568.56	-65.08
TLR-8 (3W3M)	-1261.2	-32.29	-5819.85	-112.41
TLR-9 (4QDH)	-1139.0	-6.02	-3988.37	-109.86

Figure 04. The interaction between TLR-8 (receptor protein on left) and the CV vaccine (ligand protein on right in yellow color). Here the interacting amino acid pairs are: Leu 342 (receptor)-Glu 340 (ligand), Ala 99 (receptor)- Glu 555 (ligand), Ala 97 (receptor)- Ile 496 (ligand), Phe 486 (receptor)- Leu 372 (ligand), Leu 342(receptor)- Val 488 (ligand), Leu 535 (receptor)-Phe 526

(ligand), Tyr 563 (receptor)-Asp 543 (ligand), Lys 333 (receptor)-Asp 543 (ligand), Ile 409 (receptor)- Leu 514 (ligand), Lys 412 (receptor)- Leu 515 (ligand), Leu 433 (receptor)- Ile 496 (ligand).

3.13. Conformational B-cell Epitope Prediction

The conformational B-cell epitope prediction of the CV vaccine determined four potential regions of the vaccine with scores ranging from 0.537 to 0.774 and covering total 329 amino acids (**Supplementary Table S8**).

3.14. Molecular Dynamics (MD) Simulation

The CV-TLR-8 protein complex was generated through docking. One chain had a mass of 85724.286 amu, charge -10.00 e, and 751 residues. The other chain had a mass of 62965.010 amu, charge 15.00 e, and 601 residues. A total of 92459 water molecules was added to the system after solvation from which 5 were replaced by CL ions during ionization to neutralize the system's charge. Energy minimization had been completed in 2383 steps when the steepest descent had converged and the force had reached <1000 KJ/mol. The average potential energy was calculated to be -5.2177795e+06 KJ/mol with a maximum force of 9.6646100e+02 on atom 8176. From the temperature equilibration plot of **Supplementary Figure S7 (a)**, it is clear that the target minimization value of 300K remained stable over the remainder of the equilibration and fluctuated by only ± 1.37 K. The pressure value was also predicted which showed fluctuations around 0 bar with a range of ± 100 bar and the average pressure had been found to be - 1.64581 bar. Similarly, density had also been calculated over 100 ps and the average density had been found as 1067.77 kgm³ (**Supplementary Figure S7**). The density values are mostly stable over time, indicating that the system had been well equilibrated.

Trajectory analysis was conducted and thereafter the RMSD, RMSF, and the radius of gyration were also calculated after completion of the 20 ns simulation (**Figure 05**). A plot of the RMSD backbone has been depicted in **Figure 05(a)** which revealed that RMSD levels had gone up to ~0.6 nm. The black line of the graph refers to the RMSD relative to the structure present in the minimized, equilibrated system and the red line is the RMSD relative to the crystal structure. Since both these plots are almost highly correlated, so it can be declared that the structure remained quite stable during the experiment. The RMSF and radius of gyration graphs also point to the fact that the CV-TLR-8 complex was quite stable during the experiment (**Figure 05b & 05c**). Like the complex structure, the single vaccine crystal structure also generated very good results in the MD simulation. In the RMSD graph of **Figure 05d**, it can be noticed that the RMSD value of the crystal structure had gone up to ~0.8 nm, which is quite normal for such single crystalized structure. Again, from the RMSF and radius of gyration graphs of the structure also pointed towards the fact that the crystalized structure don't have any so significant region with the possibility to undergo deformation, so it can be deduced that the crystalized vaccine CV might be quite stable in the biological environment (**Figure 05e & 05f**). So, overall, both the structures (complex and crystalized vaccine structure) showed quite acceptable and sound performance in the MD simulation study.

630

Figure 05. The results of the MD simulation of the CV-TLR-8 complex and also the only crystal structure of the vaccine. Here, (a) RMSD plot of backbone of the complex showing the complex structure had maintained a stable structure with minimum fluctuations. (b) RMS Fluctuations of all the atoms of the complex at their average positions. The peaks and dips in the graph denote the flexibility of the corresponding region in the molecular structure. (c) Radius of gyration of the complex. (e) RMSD plot of backbone of the crystal structure showing the structure had maintained a stable structure with minimum fluctuations. (b) RMS Fluctuations of all the atoms of the crystal structure at their average positions. (c) Radius of gyration of the crystalized vaccine structure.

3.15. Immune Simulation

The immune simulation of the CV vaccine was analyzed by the C-ImmSimm server which determines the stimulation of adaptive immunity as well as the immune interactions of the epitopes with their specific targets (Luckheeram et al., 2012). The immune simulation study expressed that, after the administration of the vaccine, the primary immune response against the vaccine was found to be stimulated significantly as indicated by the gradual increase in the levels of different immunoglobulins (**Figure 06a**). Again, the successive augmentation in the concentrations of active B-cell (**Figure 06b and Figure 06c**), plasma B-cell (**Figure 06d**), helper T-cell (**Figure 06e and Figure 06f**), regulatory T-cell, and cytotoxic T-cell (**Figure 06g, Figure 06h and Figure 06i**) were also predicted, which represents the capability of the vaccine to create a very potent secondary immune response. Furthermore, the increase in the concentration of macrophages and dendritic cells indicated a very potent presentation of antigen by these antigen-presenting cells (APCs) (**Figure 06j and Figure 06k**). The CV vaccine was also predicted to produce different

types of cytokines (**Figure 06I**). Henceforth, the overall immune simulation study revealed that, the CV vaccine might be able to provoke strong immunogenic responses after its administration.

3.16. Codon Adaptation, *In Silico* Cloning, and Analysis of the mRNA Structure

For *in silico* cloning and plasmid construction, at first the protein sequence of the CV vaccine was adapted by the JCat server. The codon adaptation index (CAI) value of CV was found to be 0.903 which indicates that the DNA sequence contained a higher proportion of the codons that are most likely to be used in the target organism *E. coli* strain K12 (codon bias) for efficient production the CV vaccine (**Supplementary Figure S8**) (Zuker, 2003; Mathews et al., 1999). A good GC content of 56.79% was also recorded for the adapted sequence. After codon adaptation, the predicted DNA sequence of CV vaccine was inserted into the pETite vector plasmid. The plasmid contains sequences of the SUMO tag and 6X His tag, for this reason, these sequences are expected to facilitate the purification of the CV vaccine during the downstream processing (Carbone et al., 2003) (**Figure 07**). Thereafter, the secondary structure of the CV mRNA was predicted by Mfold and RNAfold servers. The Mfold server generated a minimum free energy score of -517.50 kcal/mol, which was in agreement with the prediction of the RNAfold server (-531.10 kcal/mol). **Supplementary Figure S9** depicts the vaccine mRNA structure predicted by the RNAfold server.

695

698

699

700 **Figure 06.** C-IMMSIMM representation of the immune simulation of the best predicted vaccine,
701 CV. (a) The immunoglobulin and immunocomplex response to the CV vaccine inoculations (lines
702 colored in black) and specific subclasses are indicted by colored lines, (b) Rise in the B-cell
703 population over the course of the three injections, (c) Increment in the B-cell population per state
704 over the course of vaccination, (d) Increase in the plasma B-cell population over the course of
705 the injections, (e) Enhancement of the helper T-cell population over the course of the three
706 injections, (f) Increment in the helper T-cell population per state over the course of the vaccination,
707 (g) Elevation in the regulatory T lymphocyte over the course of the three injections, (h) Increment
708 in the cytotoxic T lymphocyte population over the course of the injections, (i) Increase in the
709 active cytotoxic T lymphocyte population per state over the course of the three injections, (j)
710 Rise in the active dendritic cell population per state over the course of the three injections,
711 (k) Increment in the macrophage population per state over the course of the injections, (l)
712 Augmentation in the concentrations of different types of cytokines over the course of the three
713 injections.

714

715

716

Figure 07. Constructed pETite vector plasmid with the CV vaccine insert (marked in red color).

4. Discussions

Vaccines are widely produced pharmaceutical products that are currently being administered worldwide to combat various pathogenic infections. Developing vaccines by conventional means is a time consuming process which sometimes takes several years to achieve desired results (María et al., 2017). However, the finesse of modern technology and the accessibility of the genomic data of about all the pathogens have saved the time of vaccine development and made it possible to develop the novel peptide-based "subunit vaccines", which are comprised of only the antigenic protein segments of the target pathogen and thus the toxic or allergenic parts of an antigen can be eliminated during vaccine designing and development (Sarkar et al., 2020c). Again, vaccine designing through these computation-based methods also saves time and cost of designing and developing processes to a greater extent (Rappuoli, 2000; Rappuoli et al., 2016). In this study, the methods of immunoinformatics were exploited to design epitope-based polyvalent vaccines against the SARS-CoV-2 virus, isolated from Bangladesh, targeting the spike glycoprotein, nucleocapsid phosphoprotein, membrane glycoprotein, and envelope protein of the virus.

After identifying the target proteins, their antigenicity and physicochemical properties were analyzed. In the antigenicity test, all the selected proteins were found to be antigenic which should aid in the antigenic response of the vaccine. The theoretical pI refers to the pH at which a protein contains no integrated charge. All the proteins except the spike glycoprotein were found to possess basic theoretical pI (pH more than 7.0), which should be quite achievable. The aliphatic index of a protein determines the relative volume of the amino acids in its side chains occupied by the aliphatic amino acids i.e. valine, alanine, etc. and higher aliphatic index of a protein represents its more thermostable state (Ikai, 1980). All the proteins had quite high aliphatic indexes, so they were considered to be quite thermostable. The GRAVY value determines the hydrophilic and

761 hydrophobic nature of a compound. The negative GRAVY value represents the hydrophilic
762 characteristics, whereas the positive value indicates the hydrophobic characteristics of a compound
763 (Kyte & Doolittle, 1982; Chang & Yang, 2013). With the negative GRAVY value, all the proteins
764 were found to be hydrophilic in nature so they should be easily soluble in water. All the proteins
765 generated sound and satisfactory results in the physicochemical property analysis.

766 After analyzing the physicochemical properties, the possible T-cell epitopes were determined by
767 the NetCTL 1.2 and IEDB servers to construct the vaccine. The cytotoxic T-cells recognize the
768 antigens, whereas the helper T-cells are involved in activating the B-cell, macrophages, and even
769 the cytotoxic T-cells (Chaudhri et al., 2009; Zhu & Paul, 2008). Furthermore, the cell mediated
770 immune response also provides a life-long immunity by secreting antiviral cytokines and by
771 recognizing and destroying the infected cells (Cano & Lopera, 2013; Sarkar et al., 2020d). A set of
772 filters were set to select the most promising T-cell epitopes from all the epitopes predicted by the servers.

773 Only the highly antigenic (so that the epitopes would generate sufficient immunogenic response), non-
774 allergenic (so that the epitopes won't be able to cause any unwanted allergenic reaction), non-toxic, 100%
775 conserved (across the selected isolates from different countries around the world), and non-homolog (to the
776 human proteome, the homolog epitopes might not be able to induce a potent immune response)
777 epitopes were considered for constructing the vaccine. Again, cytokines i.e. the IFN-gamma, IL-10, and
778 IL-4 are very important to establish a network among the cells of the immune system during
779 immunogenic responses as they activate and mediate proper functioning of the immune cells
780 (Luckheeram et al., 2012). Therefore, the prediction of the cytokine producing ability of the HTL
781 epitopes is important before the vaccine construction. Almost all the HTL epitopes as well as the
782 all the most promising epitopes were predicted to be the inducer of at least one cytokine (among
783 IFN-gamma, IL-10, and IL-4). This capability would highly contribute in the immunogenic

activities of the constructed vaccine. Again, the population coverage analysis demonstrated that a very good portion of the world population has occupied the selected MHC alleles so the designed vaccine should be effective in various regions of the world.

The most promising epitopes obtained by filtering from the previous step, were conjugated together using different linkers i.e. EAAAK, AAY, and GP GPG. During the vaccine construction, human beta-defensin-3 was used as an adjuvant. The PADRE sequence was also used to provoke the immunogenic response of the vaccines (Lee & Nguyen, 2015; Wu et al., 2010). The final vaccine was designated as “CV”. In the antigenicity and allergenicity analyses, the CV vaccine was found to be antigenic as well as non-allergenic. Hence, the vaccine might generate robust immune response as well as cause no allergic reaction within the body. Moreover, in the physicochemical property analysis of the vaccine, it was found to be basic with a high theoretical pI, which should be achievable. The aliphatic index refers to the protein’s thermal stability and thus, the higher the aliphatic index of a protein is, the more thermostable state it acquires (Panda & Chandra et al., 2012). Therefore, the CV vaccine could be considered as a thermostable vaccine protein due to having a high aliphatic index (114.48). Again, the vaccine was found to have negative GRAVY value which revealed that it might be hydrophilic. And along with the hydrophilic nature of the vaccine, it was also found to have a half-life of more than 10 hours in *E. coli*, for this reason, the mass production and purification of the vaccine in the *E. coli* cell culture system should be much easier. Furthermore, the vaccine protein was found to be soluble upon over-expression in *E. coli* by both of the servers (SolPro and Protein-Sol). Solubility is an important factor to be considered for the post-production studies of vaccines because higher solubility of a protein also aids in efficient purification during the downstream processing. And the instability index (31.49) of the vaccine protein dictated that it might be quite stable in the

biological environment because a compound with instability index less than 40 is considered to be stable (Guruprasad et al., 1990). Considering all these aspects of the physicochemical property analysis, it can be considered that the predicted vaccine might be quite effective and responsive as a potential vaccine candidate.

After the physicochemical property analysis, the secondary and tertiary structures of the CV vaccine were determined and thereafter, the protein refinement and validation steps were also performed. The refined vaccine protein was predicted to have very good amount of the amino acids in the favored regions of the Ramachandran plot analysis and the protein also had a remarkable z-score which pointed towards a good quality structure of the protein. Thereafter, the disulfide engineering of the vaccine was performed where four amino acid pairs with less than 2.2 kcal/mol bond energy were selected for mutation into cysteine residues to form the disulfide bonds among themselves. The threshold of the bond energy was kept at 2.2 kcal/mol in this study because 90% of native disulfide bonds are commonly found to have an energy value of less than 2.2 Kcal/mol (Craig & Dombkowski, 2013). Having four possible pairs of amino acid residues i.e. 93 Ala and 431 Phe, 369 Tyr and 396 Cys, 435 Thr and 528 Ala, 459 Gly and 468 Phe, CV can be considered as a stable vaccine construct thereby.

The protein-protein docking of the vaccine construct with different TLRs was performed using several online tools. The docking step is one of the necessary steps in the vaccine designing experiment because it determines the possible interaction of the constructed CV vaccine with different TLRs, which might occur during an original immune response. The docking experiment, performed by all the servers showed that the CV vaccine had good capability to interact with all the selected TLRs, with the highest score generated by TLR-8. Therefore, after the docking study, the MD simulation study of the CV-TLR-8 was carried out because CV vaccine generated the best

docking results with the TLR-8. MD simulation is performed to simulate a biological environment for the vaccine and analyze the physical movements and interactions of the atoms of a protein complex with environment molecules for a fixed length of time, which provides a view of the dynamic evolution of the system. The results of an MD simulation show how stable the vaccine complex (in this case, CV-TLR-8) is in terms of changing pressure, temperature, and motion. In our experiment, a low average potential energy of -5.2177795×10^6 KJ/mol, an average temperature fluctuation of only ± 1.3 K, and also quite desired RMSD, RMSF as well as radius of gyration were obtained which pointed towards the fact that the energy minimized CV-TLR-8 structure might be quite stable in the biological environment. The MD simulation had also predicted that the crystalized vaccine CV structure should also be stable within the biological environment with quite acceptable values in the experiment.

Thereafter, the immune simulation study of the CV vaccine was performed which predicted that the vaccine might generate an immune response which is consistent with the typical and natural immune system. After the vaccine was administrated, the primary immune response was found to be stimulated, which might provoke the secondary immune response in later stages. The CV vaccine was found to stimulate both humoral and cell-mediated mediated immune responses as indicated by the increase in the levels of the cytotoxic T-cells, helper T-cells, memory B-cells, plasma B-cells, and different antibodies (Almofti et al., 2018; Carvalho et al., 2002). Again, the rise in the concentration of APCs like the macrophages and dendritic cells also indicated a very good antigen presentation. Moreover, the augmentation in the cytokine profile was also reported which represents the possibility of the vaccine to produce the broad-spectrum immunity against the viral invasions (Shey et al., 2019; Hoque et al., 2019; Kambayashi & Laufer, 2014). The negligible Simpson index (D) predicted a diverse immune response of the vaccine CV (Rapin et

al., 2010). Overall, the CV vaccine was found to be effective and satisfactory in the results of the immune simulation study.

Finally, the codon adaptation and *in silico* cloning were performed to design a recombinant plasmid which might be used for the mass production of the CV vaccine in the *E. coli* strain K12. *E. coli* cell culture system is considered as a recommended system for the production of recombinant proteins. In the codon adaptation study of the vaccine, the obtained results were found to be satisfactory with the CAI value of 0.903 and the GC content of 56.79%, because any CAI value over 0.80 and the GC content within 30% to 70% are considered to be good scores (Shey et al., 2019; Morla et al., 2016; Khatoon et al., 2017). Following that, the optimized CV DNA sequence was inserted between the *EaeI* and *StyI* restriction sites of the pETite plasmid. The newly constructed recombinant plasmid contains the sequences for SUMO and 6X His tags, so the vaccine protein was expected to be expressed in fusion with these tags, which might facilitate its purification process in the later stages. Moreover, during the stability prediction of the vaccine mRNA secondary structure, both Mfold and RNAfold servers generated negative and much lower minimal free energies of -517.50 kcal/mol and -531.10 kcal/mol, respectively. Therefore, with these lower minimal free energies, it can be declared that the predicted CV vaccine might be quite stable upon transcription *in vivo* (Hamasaki-Katagiri et al., 2017).

Overall, this study suggests the designed CV vaccine as a safe and effective epitope based subunit vaccine to fight against the SARS-CoV-2 virus, based on the strategies employed in the study. However, more *in vivo* and *in vitro* studies are required to finally confirm the outcomes of this study.

5. Conclusion

The recent COVID-19 pandemic has caused the sufferings of millions of people worldwide, while claiming the lives of hundreds of thousands of people. Although researchers are rushing to develop potential counter-measure to fight the virus, however, till now, all the efforts have showed potential drawbacks. Different types of antiviral therapies are also being tried all over the world, but their outcomes are also raising serious questions about their usage. Moreover, the vaccines that are now under development or various stages of trials, are inactivated or live-attenuated vaccines, which can have serious consequences if any unwanted reversal back to the virulence form occurs. Therefore, in this study, a blue-print of potential vaccine against the SARS-CoV-2 has been designed targeting four distinctive proteins i.e., spike glycoprotein, nucleocapsid phosphoprotein, membrane glycoprotein, and envelope protein. During the vaccine construction, only those epitopes which were found to be highly antigenic (so that the vaccine would be able to generate robust immune response), non-allergenic (so that the vaccine won't cause any harmful reaction within the body), non-toxic, non-human homolog, and 100% conserved epitopes (so that vaccine would be effective against different isolates around the world) were used. Numerous *in silico* validations, conducted in the study, also indicated that the designed vaccine might be quite safe and effective as well as well responsive to use. If satisfied results are achieved in the wet-lab based studies, then this epitope-based vaccine candidate might provide a relatively cheap and effective option to reach the entire world to combat the COVID-19 pandemic.

6. Conflict of Interest

The authors declare that they have no conflict of interest.

898

899 **7. Acknowledgements**

900 The authors are thankful to the COVID Research Cell (CRC), Wazed Mia Science Research
901 Center, Jahangirnagar University, for providing the support to conduct this research.

902

903 **8. References**

904

905 Adhikari, U.K. and Rahman, M.M., 2017. Overlapping CD8+ and CD4+ T-cell epitopes
906 identification for the progression of epitope-based peptide vaccine from nucleocapsid and
907 glycoprotein of emerging Rift Valley fever virus using immunoinformatics approach. *Infection,
908 Genetics and Evolution*, 56, pp.75-91. DOI: <https://doi.org/10.1016/j.meegid.2017.10.022>

909

910 Almofti, Y.A., Abd-elrahman, K.A., Gassmallah, S.A.E. and Salih, M.A., 2018. Multi Epitopes
911 Vaccine Prediction against Severe Acute Respiratory Syndrome (SARS) Coronavirus Using
912 Immunoinformatics Approaches. *American Journal of Microbiological Research*, 6(3), pp.94-114.
913 DOI: <https://doi.org/10.12691/ajmr-6-3-5>

914

915 Alsaadi, E. A., & Jones, I. M. Membrane binding proteins of coronaviruses. *Futur Med* [Internet].
916 2019 Apr 1 [cited 2020 Jun 10]; 14 (4): 275–86.

917

918 Amanat, F., & Krammer, F. (2020). SARS-CoV-2 vaccines: status report. *Immunity*.

919

920 Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular
921 docking. *Proteins: Structure, Function, and Bioinformatics*. 2007 Oct;69(1):139-59. DOI:
922 10.1002/prot.21495

923
924 Angov, E., 2011. Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol.*
925 *J.* 6, 650-659. DOI: <https://doi.org/10.1002/biot.201000332>

926 Arai, R., Ueda, H., Kitayama, A., Kamiya, N. and Nagamune, T., 2001. Design of the linkers
927 which effectively separate domains of a bifunctional fusion protein. *Protein engineering*, 14(8),
928 pp.529-532. DOI: <https://doi.org/10.1093/protein/14.8.529>

929
930 Atapour, A., Mokarram, P., MostafaviPour, Z., Hosseini, S.Y., Ghasemi, Y., Mohammadi, S.,
931 Nezafat, N., 2019. Designing a fusion protein vaccine against HCV: an in silico approach. *Int. J.*
932 *Pept. Res. Thera.* 25, 861-872. DOI: <https://doi.org/10.1007/s10989-018-9735-4>

933
934 Bacchetta, R., Gregori, S. and Roncarolo, M.G., 2005. CD4+ regulatory T cells: mechanisms of
935 induction and effector function. *Autoimmunity reviews*, 4(8), pp.491-496. DOI:
936 <https://doi.org/10.1016/j.autrev.2005.04.005>

937
938 Biovia, D.S., Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N. and Richmond,
939 T.J., 2000. Dassault Systèmes BIOVIA, Discovery Studio Visualizer, v. 17.2, San Diego: Dassault
940 Systèmes, 2016. *J Chem Phys.* 10, 0021-9991

941

- 942 Biswal, J.K., Bisht, P., Mohapatra, J.K., Ranjan, R., Sanyal, A. and Pattnaik, B., 2015. Application
943 of a recombinant capsid polyprotein (P1) expressed in a prokaryotic system to detect antibodies
944 against foot-and-mouth disease virus serotype O. *Journal of virological methods*, 215, pp.45-51.
945 DOI: <https://doi.org/10.1016/j.jviromet.2015.02.008>
946
- 947 Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic acids*
948 *research*. 2019 Jul 2;47(W1):W402-7. DOI: <https://doi.org/10.1093/nar/gkz297>
949
- 950 Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis
951 tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC bioinformatics*. 2007
952 Dec 1;8(1):361. DOI: <https://doi.org/10.1186/1471-2105-8-361nehka>
953
- 954 Cano, R.L.E. and Lopera, H.D.E., 2013. Introduction to T and B lymphocytes. In *Autoimmunity:*
955 *From Bench to Bedside [Internet]*. El Rosario University Press.
956
- 957 Carbone A, Zinovyev A, Képes F. Codon adaptation index as a measure of dominating codon bias.
958 *Bioinformatics* 2003;19:2005-15. DOI: doi.org/10.1093/bioinformatics/btg272
959
- 960 Carvalho, L.H., Sano, G.I., Hafalla, J.C., Morrot, A., De Lafaille, M.A.C. and Zavala, F., 2002.
961 IL-4-secreting CD4⁺ T cells are crucial to the development of CD8⁺ T-cell responses against
962 malaria liver stages. *Nature medicine*, 8(2), pp.166-170. DOI: [https://doi.org/10.1038/nm0202-](https://doi.org/10.1038/nm0202-166)
963 [166](https://doi.org/10.1038/nm0202-166)
964

Castiglione, F., Mantile, F., De Berardinis, P. and Prisco, A., 2012. How the interval between prime and boost injection affects the immune response in a computational model of the immune system. *Computational and mathematical methods in medicine*, 2012. DOI: <https://doi.org/10.1155/2012/842329>

Chan JF, Lau SK, To KK, Cheng VC, Woo PC, Yuen KY. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin Microbiol Rev.* 2015;28(2):465-522

Chan JF, Lau SK, Woo PC. The emerging novel Middle East respiratory syndrome coronavirus: the "knowns" and "unknowns". *J Formos Med Assoc.* 2013;112(7):372-381

Chang, K.Y. and Yang, J.R., 2013. Analysis and prediction of highly effective antiviral peptides based on random forests. *PloS one*, 8(8). DOI: <https://doi.org/10.1371/journal.pone.0070166>

Chaudhri, G., Quah, B.J., Wang, Y., Tan, A.H., Zhou, J., Karupiah, G. and Parish, C.R., 2009. T cell receptor sharing by cytotoxic T lymphocytes facilitates efficient virus control. *Proceedings of the National Academy of Sciences*, 106(35), pp.14984-14989. DOI: <https://doi.org/10.1073/pnas.0906554106>

Cheng VC, Lau SK, Woo PC, Yuen KY. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin Microbiol Rev.* 2007;20(4):660-694

Cheung E. China coronavirus: Hong Kong researchers have already developed vaccine but need time to test it, expert reveals. South China Morning Post. <https://www.scmp.com/news/hong-kong/health-environment/article/3047956/china-coronavirus-hongkong-researchers-have>.

Accessed 28 Feb 2020

Chong, L.C. and Khan, A.M., 2019. Vaccine Target Discovery. DOI: <https://doi.org/10.1016/b978-0-12-809633-8.20100-3>

Cooper, N.R. and Nemerow, G.R., 1984. The role of antibody and complement in the control of viral infections. *Journal of investigative dermatology*, 83(s 1), pp.121-127. DOI: <https://doi.org/10.1038/jid.1984.33>

Craig, D.B. and Dombkowski, A.A., 2013. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC bioinformatics*, 14(1), p.346. DOI: <https://doi.org/10.1186/1471-2105-14-346>

Dhanda SK, Gupta S, Vir P, Raghava GP. Prediction of IL4 inducing peptides. Clinical and Developmental Immunology. 2013 Oct;2013. DOI: <https://doi.org/10.1155/2013/263952>

Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. Biology direct. 2013 Dec;8(1):30. DOI: <https://doi.org/10.1186/1745-6150-8-30>

- 1010 Dimitrov, I., Flower, D.R. and Doytchinova, I., 2013, April. AllerTOP-a server for in silico
1011 prediction of allergens. In *BMC bioinformatics* (Vol. 14, No. 6, p. S4). BioMed Central. DOI:
1012 <https://doi.org/10.1186/1471-2105-14-s6-s4>
1013
- 1014 Dimitrov, I., Naneva, L., Doytchinova, I. and Bangov, I., 2014. AllergenFP: allergenicity
1015 prediction by descriptor fingerprints. *Bioinformatics*, 30(6), pp.846-851. DOI:
1016 <https://doi.org/10.1093/bioinformatics/btt619>
1017
- 1018 Doytchinova, I.A., Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens,
1019 tumour antigens and subunit vaccines. *BMC Bioinform.* 8, 4. DOI: [https://doi.org/10.1186/1471-](https://doi.org/10.1186/1471-2105-8-4)
1020 [2105-8-4](https://doi.org/10.1186/1471-2105-8-4)
1021
- 1022 Duhovny, D., Nussinov, R., Wolfson, H.J., 2002. Efficient unbound docking of rigid molecules.
1023 In *International workshop on algorithms in bioinformatics* (185-200). Springer, Berlin,
1024 Heidelberg. DOI: https://doi.org/10.1007/3-540-45784-4_14
1025
- 1026 Garcia, K.C., Teyton, L. and Wilson, I.A., 1999. Structural basis of T cell recognition. *Annual*
1027 *review of immunology*, 17(1), pp.369-397. DOI:
1028 <https://doi.org/10.1146/annurev.immunol.17.1.369>
1029
- 1030 Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., 2005.
1031 Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols*
1032 *handbook* (pp. 571-607). Humana press. DOI: <https://doi.org/10.1385/1-59259-890-0:571>

1033

1034 Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. JCat: a novel tool to
1035 adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res*
1036 2005;33:W526-31. DOI: doi.org/10.1093/nar/gki376

1037

1038 Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R. and Hofacker, I.L., 2008. The vienna RNA
1039 websuite. *Nucleic acids research*, 36(suppl_2), pp.W70-W74. DOI:
1040 <https://doi.org/10.1093/nar/gkn188>

1041

1042 Gu, Y., Sun, X., Li, B., Huang, J., Zhan, B. and Zhu, X., 2017. Vaccination with a paramyosin-
1043 based multi-epitope vaccine elicits significant protective immunity against *Trichinella spiralis*
1044 infection in mice. *Frontiers in microbiology*, 8, p.1475. DOI:
1045 <https://doi.org/10.3389/fmicb.2017.01475>

1046

1047 Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A. and Kumar, R., 2013. Consortium, OSDD;
1048 Raghava, GPS In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS One*, 8,
1049 p.e73957

1050

1051 Guruprasad, K., Reddy, B.B. and Pandit, M.W., 1990. Correlation between stability of a protein
1052 and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from
1053 its primary sequence. *Protein Engineering, Design and Selection*, 4(2), pp.155-161. DOI:
1054 10.1093/protein/4.2.155

1055

1056 Hamasaki-Katagiri, N., Lin, B.C., Simon, J., Hunt, R.C., Schiller, T., Russek-Cohen, E., Komar,
 1057 A.A., Bar, H. and Kimchi-Sarfaty, C., 2017. The importance of mRNA structure in determining
 1058 the pathogenicity of synonymous and non-synonymous mutations in haemophilia. *Haemophilia*,
 1059 23(1), pp.e8-e17. DOI: <https://doi.org/10.1111/hae.13107>
 1060
 1061 Hebditch, M., Carballo-Amador, M.A., Charonis, S., Curtis, R. and Warwicker, J., 2017. Protein–
 1062 Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*, 33(19), pp.3098-
 1063 3100. DOI: <https://doi.org/10.1093/bioinformatics/btx345>
 1064
 1065 Hoque, M.N., Istiaq, A., Clement, R.A., Sultana, M., Crandall, K.A., Siddiki, A.Z. and Hossain,
 1066 M.A., 2019. Metagenomic deep sequencing reveals association of microbiome signature with
 1067 functional biases in bovine mastitis. *Scientific reports*, 9(1), pp.1-14. DOI:
 1068 <https://doi.org/10.1038/s41598-019-49468-4>
 1069
 1070 <http://clinicaltrials.gov/>. Accessed on: 16 August 2020.
 1071
 1072 https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1? Accessed on: 16
 1073 August 2020.
 1074
 1075 Ikai, A., 1980. Thermostability and aliphatic index of globular proteins. *The Journal of*
 1076 *Biochemistry*, 88(6), pp.1895-1898. DOI: <https://doi.org/10.1093/oxfordjournals.jbchem.a133168>
 1077

1078 Inovio Pharmaceuticals . 2020. Inovio collaborating with Beijing advaccine to advance INO-4800
 1079 vaccine against new coronavirus in China.[http://ir.inovio.com/news-and-media/news/press-](http://ir.inovio.com/news-and-media/news/press-release-details/2020/IVI-INOVIO-and-KNIH-to-Partner-with-CEPI-in-Phase-12-Clinical-Trial-of-INOVIOS-COVID-19-DNA-Vaccine-in-South-Korea/default.aspx)
 1080 [release-details/2020/IVI-INOVIO-and-KNIH-to-Partner-with-CEPI-in-Phase-12-Clinical-Trial-](http://ir.inovio.com/news-and-media/news/press-release-details/2020/IVI-INOVIO-and-KNIH-to-Partner-with-CEPI-in-Phase-12-Clinical-Trial-of-INOVIOS-COVID-19-DNA-Vaccine-in-South-Korea/default.aspx)
 1081 [of-INOVIOS-COVID-19-DNA-Vaccine-in-South-Korea/default.aspx](http://ir.inovio.com/news-and-media/news/press-release-details/2020/IVI-INOVIO-and-KNIH-to-Partner-with-CEPI-in-Phase-12-Clinical-Trial-of-INOVIOS-COVID-19-DNA-Vaccine-in-South-Korea/default.aspx)
 1082 Islam H, Rahman A, Masud J, Shweta DS, Araf Y, Ullah MA, Sium SM, Sarkar B. A Generalized
 1083 Overview of SARS-CoV-2: Where Does the Current Knowledge Stand?. Electron J Gen Med.
 1084 2020; 17 (6): em251. DOI: 10.29333/ejgm/8258
 1085
 1086 Johnston C, Gottlieb SL, Wald A. Status of vaccine research and development of vaccines for
 1087 herpes simplex virus. Vaccine. 2016 Jun 3;34(26):2948-52. DOI: 10.1016/j.vaccine.2015.12.076
 1088 Jones DT. Protein secondary structure prediction based on position-specific scoring matrices.
 1089 Journal of molecular biology. 1999 Sep 17;292(2):195-202. DOI:
 1090 <https://doi.org/10.1006/jmbi.1999.3091>
 1091
 1092 Kaiser Permanente Washington Health Research Institute . Kaiser Permanente, Washington Health
 1093 Research Institute; Seattle: 16 March 2020. Kaiser Permanente launches first coronavirus vaccine
 1094 trial. [Internet] Retrieved: 23 March 2020.
 1095
 1096 Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., Xu, J., 2012. Template-based protein
 1097 structure modeling using the RaptorX web server. Nat. protoc. 7, 1511. DOI:
 1098 <https://doi.org/10.1038/nprot.2012.085>
 1099

- 1100 Kambayashi, T. and Laufer, T.M., 2014. Atypical MHC class II-expressing antigen-presenting
 1101 cells: can anything replace a dendritic cell?. *Nature Reviews Immunology*, 14(11), pp.719-730.
 1102 DOI: <https://doi.org/10.1038/nri3754>
 1103
- 1104 Khatoon, N., Pandey, R.K. and Prajapati, V.K., 2017. Exploring Leishmania secretory proteins to
 1105 design B and T cell multi-epitope subunit vaccine using immunoinformatics approach. *Scientific*
 1106 *reports*, 7(1), pp.1-12. DOI: <https://doi.org/10.1038/s41598-017-08842-w>
 1107
- 1108 Ko, J., Park, H., Heo, L. and Seok, C., 2012. GalaxyWEB server for protein structure prediction
 1109 and refinement. *Nucleic acids research*, 40(W1), pp.W294-W297. DOI:
 1110 <https://doi.org/10.1093/nar/gks493>
 1111
- 1112 Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Vajda S. How good is
 1113 automated protein docking?. *Proteins: Struct, Funct, Bioinf* 2013;81:2159-66. DOI:
 1114 <https://doi.org/10.1002/prot.24403>
 1115
- 1116 Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro
 1117 web server for protein–protein docking. *Nat Protoc* 2017;12:255. DOI:
 1118 <https://doi.org/10.1038/nprot.2016.169>
 1119
- 1120 Kumar H, Frischknecht F, Mair GR, Gomes J. In silico identification of genetically attenuated
 1121 vaccine candidate genes for Plasmodium liver stage. *Infection, Genetics and Evolution*. 2015 Dec
 1122 1;36:72-81. PMID: **26348884**, DOI: 10.1016/j.meegid.2015.09.002

- 1123
- 1124 Kyte, J., Doolittle. RF 1982. *Mol. Biol*, 157, pp.105-132. DOI: <https://doi.org/10.1016/0022->
- 1125 2836(82)90515-0
- 1126 Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. Larsen MV,
- 1127 Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M. BMC Bioinformatics. Oct 31;8:424.
- 1128 2007
- 1129
- 1130 Laskowski, R.A., MacArthur, M.W. and Thornton, J.M., 2006. PROCHECK: validation of
- 1131 protein-structure coordinates. DOI: <https://doi.org/10.1107/97809553602060000882>
- 1132
- 1133 Lee, S. and Nguyen, M.T., 2015. Recent advances of vaccine adjuvants for infectious diseases.
- 1134 *Immune network*, 15(2), pp.51-57. DOI: <https://doi.org/10.4110/in.2015.15.2.51>
- 1135
- 1136 Lee, S., & Nguyen, M. T. (2015). Recent advances of vaccine adjuvants for infectious diseases.
- 1137 *Immune network*, 15(2), 51-57.
- 1138
- 1139 Luckheeram, R.V., Zhou, R., Verma, A.D. and Xia, B., 2012. CD4+ T cells: differentiation and
- 1140 functions. *Clinical and developmental immunology*, 2012. DOI:
- 1141 <https://doi.org/10.1155/2012/925135>
- 1142
- 1143 Magnan, C.N., Randall, A. and Baldi, P., 2009. SOLpro: accurate sequence-based prediction of
- 1144 protein solubility. *Bioinformatics*, 25(17), pp.2200-2207. DOI:
- 1145 <https://doi.org/10.1093/bioinformatics/btp386>

- 1146
- 1147 María, R.R., Arturo, C.J., Alicia, J.A., Paulina, M.G., Gerardo, A.O., 2017. The impact of
- 1148 bioinformatics on vaccine design and development. InTech, Rijeka, Croatia.
- 1149 Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H., 1999. Expanded sequence dependence of
- 1150 thermodynamic parameters improves prediction of RNA secondary structure. *Journal of*
- 1151 *molecular biology*, 288(5), pp.911-940. DOI: <https://doi.org/10.1006/jmbi.1999.2700>
- 1152
- 1153 Mathews, D.H., Turner, D.H. and Zuker, M., 2007. RNA secondary structure prediction. *Current*
- 1154 *protocols in nucleic acid chemistry*, 28(1), pp.11-2. DOI:
- 1155 <https://doi.org/10.1002/0471142700.nc1102s28>
- 1156
- 1157 Mehla, K. and Ramana, J., 2016. Identification of epitope-based peptide vaccine candidates against
- 1158 enterotoxigenic Escherichia coli: a comparative genomics and immunoinformatics approach.
- 1159 *Molecular BioSystems*, 12(3), pp.890-901. DOI: <https://doi.org/10.1039/c5mb00745c>
- 1160
- 1161 Meza, B., Ascencio, F., Sierra-Beltrán, A.P., Torres, J. and Angulo, C., 2017. A novel design of a
- 1162 multi-antigenic, multistage and multi-epitope vaccine against Helicobacter pylori: an in silico
- 1163 approach. *Infection, Genetics and Evolution*, 49, pp.309-317. DOI:
- 1164 <https://doi.org/10.1016/j.meegid.2017.02.007>
- 1165
- 1166 Moin AT, Sakib MN, Araf Y, Sarkar B, Ullah MA. Combating COVID-19 Pandemic in
- 1167 Bangladesh: A Memorandum from Developing Country. Preprints. 2020 May 27. DOI:
- 1168 10.20944/preprints202005.0435.v1

- 1169
- 1170 Morla, S., Makhija, A. and Kumar, S., 2016. Synonymous codon usage pattern in glycoprotein
1171 gene of rabies virus. *Gene*, 584(1), pp.1-6. DOI: <https://doi.org/10.1016/j.gene.2016.02.047>
- 1172
- 1173 Morris, A.L., MacArthur, M.W., Hutchinson, E.G. and Thornton, J.M., 1992. Stereochemical
1174 quality of protein structure coordinates. *Proteins: Structure, Function, and Bioinformatics*, 12(4),
1175 pp.345-364. DOI: <https://doi.org/10.1002/prot.340120407>
- 1176
- 1177 Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, Raghava GP. Computer-aided
1178 designing of immunosuppressive peptides based on IL-10 inducing potential. *Scientific reports*.
1179 2017 Feb 17;7:42851. DOI: <https://doi.org/10.1038/srep42851>
- 1180
- 1181 Nugent, T., Cozzetto, D. and Jones, D.T., 2014. Evaluation of predictions in the CASP10 model
1182 refinement category. *Proteins: Structure, Function, and Bioinformatics*, 82, pp.98-111. DOI:
1183 <https://doi.org/10.1002/prot.24377>
- 1184
- 1185 Panda, S. and Chandra, G., 2012. Physicochemical characterization and functional analysis of
1186 some snake venom toxin proteins and related non-toxin proteins of other chordates.
1187 *Bioinformation*, 8(18), p.891. DOI: <https://doi.org/10.6026/97320630008891>
- 1188
- 1189 Pandey, R.K., Sundar, S., Prajapati, V.K., 2016. Differential expression of miRNA regulates T cell
1190 differentiation and plasticity during visceral leishmaniasis infection. *Front. Microbial*. 7, 206. DOI:
1191 <https://doi.org/10.3389/fmicb.2016.00206>

1192

1193 Perlman S. Another Decade, Another Coronavirus. *N Engl J Med*. 2020 Feb 20; 382(8):760-762

1194 Petersen, M.T.N., Jonson, P.H., Petersen, S.B., 1999. Amino acid neighbours and detailed
1195 conformational analysis of cysteines in proteins. *Protein Eng.* 12, 535-548. DOI:
1196 <https://doi.org/10.1093/protein/12.7.535>

1197

1198 Ponomarenko, J., Bui, H.H., Li, W., Fusseder, N., Bourne, P.E., Sette, A. and Peters, B., 2008.
1199 ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics*,
1200 9(1), p.514.

1201

1202 Purcell, A.W., McCluskey, J., Rossjohn, J., 2007. More than one reason to rethink the use of
1203 peptides in vaccine design. *Nat. Rev. Drug Discov.* 6, 404-414. DOI:
1204 <https://doi.org/10.1038/nrd2224>

1205

1206 Rapin, N., Lund, O., Bernaschi, M. and Castiglione, F., 2010. Computational immunology meets
1207 bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune
1208 system. *PLoS One*, 5(4). DOI: <https://doi.org/10.1371/journal.pone.0009862>

1209

1210 Rappuoli R. Reverse vaccinology. *Current opinion in microbiology*. 2000 Oct 1;3(5):445-50. DOI:
1211 [https://doi.org/10.1016/s1369-5274\(00\)00119-3](https://doi.org/10.1016/s1369-5274(00)00119-3)

1212

- 1213 Rappuoli, R., Bottomley, M.J., D'Oro, U., Finco, O., De Gregorio, E., 2016. Reverse vaccinology
 1214 2.0: Human immunology instructs vaccine antigen design. J. Exp. Med. 213, 469-481. DOI:
 1215 <https://doi.org/10.1084/jem.20151960>
 1216
- 1217 Saadi, M., Karkhah, A. and Nouri, H.R., 2017. Development of a multi-epitope peptide vaccine
 1218 inducing robust T cell responses against brucellosis using immunoinformatics based approaches.
 1219 *Infection, Genetics and Evolution*, 51, pp.227-234. DOI:
 1220 <https://doi.org/10.1016/j.meegid.2017.04.009>
 1221
- 1222 Saha S, Raghava GP. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes.
 1223 Nucleic acids research. 2006 Jul 1;34(suppl_2):W202-9. DOI: <https://doi.org/10.1093/nar/gkl343>
 1224
- 1225 Sarkar B, Islam SS, Zohora US, Ullah MA. Virus like particles-A recent advancement in vaccine
 1226 development. *미생물학회지*. 2019 Dec;55(4):327-43. DOI: 10.7845/kjm.2019.9089
- 1227 Sarkar B, Ullah MA, Johora FT, Taniya MA, Araf Y. Immunoinformatics-guided designing of
 1228 epitope-based subunit vaccine against the SARS Coronavirus-2 (SARS-CoV-2). Immunobiology.
 1229 2020a. DOI: 10.1016/j.imbio.2020.151955
 1230
- 1231 Sarkar B, Ullah MA, Araf Y. A systematic and reverse vaccinology approach to design novel
 1232 subunit vaccines against dengue virus type-1 and human Papillomavirus-16. Informatics in
 1233 Medicine Unlocked. 2020b May 16:100343. DOI: 10.1016/j.imu.2020.100343
 1234

1235 Sarkar B, Ullah MA, Araf Y, Das S, Hosen MJ. Blueprint of epitope-based multivalent and
 1236 multipathogenic vaccines: targeted against the dengue and zika viruses. *Journal of Biomolecular*
 1237 *Structure and Dynamics*. 2020c Aug 7:1-21. DOI: 10.1080/07391102.2020.1804456

1238
 1239 Sarkar B, Ullah MA, Araf Y, Das S, Rahman MH, Moin AT. Designing novel epitope-based
 1240 polyvalent vaccines against herpes simplex virus-1 and 2 exploiting the immunoinformatics
 1241 approach. *Journal of Biomolecular Structure and Dynamics*. 2020d Aug 6:1-21. DOI:
 1242 10.1080/07391102.2020.1803969

1243
 1244 Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers
 1245 for rigid and symmetric docking. *Nucleic acids research*. 2005 Jul 1;33(suppl_2):W363-7. DOI:
 1246 10.1093/nar/gki481

1247
 1248 Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: current knowledge.
 1249 *Virology journal*, 16(1), 1-22.

1250
 1251 Shey, R.A., Ghogomu, S.M., Esoh, K.K., Nebangwa, N.D., Shintouo, C.M., Nongley, N.F., Asa,
 1252 B.F., Ngale, F.N., Vanhamme, L. and Souopgui, J., 2019. In-silico design of a multi-epitope
 1253 vaccine candidate against onchocerciasis and related filarial diseases. *Scientific reports*, 9(1), pp.1-
 1254 18. DOI: <https://doi.org/10.1038/s41598-019-40833-x>

1255

1256 Shieber J. Codagenix raises \$20 million for a new flu vaccine and other therapies. Tech Crunch.
1257 <https://techcrunch.com/2020/01/13/codagenix-raises-20-million-for-a-new-flu-vaccine-and->
1258 [othertherapies/](https://techcrunch.com/2020/01/13/codagenix-raises-20-million-for-a-new-flu-vaccine-and-othertherapies/). Accessed on: 16 August 2020.

1259

1260 Solanki, V. and Tiwari, V., 2018. Subtractive proteomics to identify novel drug targets and reverse
1261 vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. *Scientific*
1262 *reports*, 8(1), pp.1-19. DOI: <https://doi.org/10.1038/s41598-018-26689-7>

1263

1264 Stern, L.J., Calvo-Calle, J.M., 2009. HLA-DR: molecular insights and vaccine design. *Curr.*
1265 *Pharm. Des.* 15, 3249-3261.

1266

1267 Thomsen M, Lundegaard C, Buus S, Lund O, Nielsen M. MHCcluster, a method for functional
1268 clustering of MHC molecules. *Immunogenetics*. 2013 Sep 1;65(9):655-65. DOI:
1269 <https://doi.org/10.1007/s00251-013-0714-9>

1270

1271 Ullah MA, Araf Y, Sarkar B, Moin AT, Reshad RA, Hasanur MD. Pathogenesis, Diagnosis and
1272 Possible Therapeutic Options for COVID-19. 2020a. DOI: 10.20944/preprints202004.0372.v1

1273

1274 Ullah A, Sarkar B, Islam SS. Exploiting the reverse vaccinology approach to design novel subunit
1275 vaccine against ebola virus. *Immunobiology*. 2020b May 1:151949. DOI:
1276 10.1016/j.imbio.2020.151949

1277

- 1278 Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Kozakov D. New
 1279 additions to the C l u s P r o server motivated by CAPRI. *Proteins: Struct, Funct, Bioinf*
 1280 2017;85:435-44. DOI: <https://doi.org/10.1002/prot.25219>
 1281
- 1282 Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters
 1283 B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 2018 Oct 24. doi:
 1284 10.1093/nar/gky1006
 1285
- 1286 Wen, D., Foley, S.F., Hronowski, X.L., Gu, S. and Meier, W., 2013. Discovery and investigation
 1287 of O-xylosylation in engineered proteins containing a (GGGS) n linker. *Analytical chemistry*,
 1288 85(9), pp.4805-4812. DOI: <https://doi.org/10.1021/ac400596g>
 1289
- 1290 Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T. HawkDock: a web server to predict and
 1291 analyze the protein–protein complex based on computational docking and MM/GBSA. *Nucleic*
 1292 *Acids Res* 2019. DOI: <https://doi.org/10.1093/nar/gkz397>
 1293
- 1294 Wiederstein, M. and Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of
 1295 errors in three-dimensional structures of proteins. *Nucleic acids research*, 35(suppl_2), pp.W407-
 1296 W410. DOI: <https://doi.org/10.1093/nar/gkm290>
 1297
- 1298 Wu, C.Y., Monie, A., Pang, X., Hung, C.F., Wu, T.C., 2010. Improving therapeutic HPV
 1299 peptidebased vaccine potency by enhancing CD4+ T help and dendritic cell activation. *J. Biomed.*
 1300 *Sci.* 17, 88. DOI: <https://doi.org/10.1186/1423-0127-17-88>

- 1301
- 1302 Wu, C.Y., Monie, A., Pang, X., Hung, C.F., Wu, T.C., 2010. Improving therapeutic HPV
- 1303 peptidebased vaccine potency by enhancing CD4+ T help and dendritic cell activation. *J. Biomed.*
- 1304 *Sci.* 17, 88. DOI: <https://doi.org/10.1186/1423-0127-17-88>
- 1305
- 1306 Yadav, D. K., Yadav, N., & Khurana, S. M. P. (2014). Vaccines: present status and applications.
- 1307 In *Animal Biotechnology* (pp. 491-508). Academic Press.
- 1308
- 1309 Zhang, L., 2018. Multi-epitope vaccines: a promising strategy against tumors and viral infections.
- 1310 *Cellular & molecular immunology*, 15(2), pp.182-184. DOI: <https://doi.org/10.1038/cmi.2017.92>
- 1311
- 1312 Zhang N, Li C, Hu Y, Li K, Liang J, Wang L, Du L, Jiang S. Current development of COVID-19
- 1313 diagnostics, vaccines and therapeutics. *Microbes and Infection*. 2020 May 6. DOI:
- 1314 10.1016/j.micinf.2020.05.001
- 1315
- 1316 Zhu Y, Li C, Chen L, Xu B, Zhou Y, Cao L, Shang Y, Fu Z, Chen A, Deng L, Bao Y. A novel
- 1317 human coronavirus OC43 genotype detected in mainland China. *Emerging microbes & infections*.
- 1318 2018 Dec 1;7(1):1-4.
- 1319
- 1320 Zhu, J. and Paul, W.E., 2008. CD4 T cells: fates, functions, and faults. *Blood, The Journal of the*
- 1321 *American Society of Hematology*, 112(5), pp.1557-1569. DOI: [https://doi.org/10.1182/blood-](https://doi.org/10.1182/blood-2008-05-078154)
- 1322 2008-05-078154
- 1323

- 1324 Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic*
1325 *acids research*, 31(13), pp.3406-3415. DOI: <https://doi.org/10.1093/nar/gkg595>
1326
1327 Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. Coronaviruses -drug discovery and therapeutic
1328 options. *Nat Rev Drug Discov*. 2016;15(5):327-347
1329
1330
1331