Sains Malaysiana 51(9)(2022): 2985-2997
http://doi.org/10.17576/jsm-2022-5109-19

Construction of Multi-Epitopes Vaccine Candidate against SARS-CoV-2 D614G Variant

(Pembinaan Calon Vaksin Epitop Pelbagai terhadap Varian D614G SARS-CoV-2)

Siti Asmaa Mat jusoh[1ǂ], Danesh Thangeswaran[1ǂ], Muhammad Syahir Hakimi Mohd Hazli[1], Mohd Firdaus Raih[3], Nurulasma Abdullah[1] & Shaharum Shamsuddin[1,2,*]

[1]School of Health Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan Darul Naim, Malaysia
[2]Universiti Sains Malaysia (USM)-RIKEN Interdisciplinary Collaboration for Advanced Sciences (URICAS), 11700 Gelugor, Penang, Malaysia
[3]Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

Received: 28 November 2021/Accepted: 23 March 2022

ǂ *These authors contributed equally to this work*

ABSTRACT

COVID-19 caused by the SARS-CoV-2 virus has become a real threat due to the emergence of new variants which are more deadly with higher infectivity. Vaccine constructs that target specific SARS-CoV-2 variants are needed for stemming COVID-19 fatality. The spike (S) glycoprotein is the major antigenic component that triggers the host immune response. Reverse vaccinology strategy was applied to the S protein of COVID-19 variant D614G to identify highly ranked antigenic proteins. In this study, a multi-epitope synthetic gene was designed using computational strategies for the COVID-19 D614G variant. The SARS-CoV-2 D614G variant protein sequence was retrieved from the NCBI database. The prediction of linear B-cell epitopes was carried out using Artificial Neural Network (ANN)-based ABCpred and BepiPred 2.0 software. The top 15 highly antigenic epitopes sequences were then selected. Propred 1 and Propred servers were used to identify major histocompatibility complex (MHC) class I and class II binding epitopes within pre-determined B-cell epitopes to predict T-cell epitopes. The top 5 MHC class I and class II were selected. Further *in-silico* testing for its solubility, allergenicity, antigenicity, and other physiochemical properties was analyzed using Bpred. The constructed gene was subjected to assembly PCR and the gene product was confirmed by Sanger sequencing. The findings from this study suggested that a highly antigenic specific region of the SARS-CoV-2 D614G variant can be predicted *in-silico* and amplified using the assembly PCR method. The designed synthetic gene was shown to elicit specific humoral and cell-mediated immune responses towards the SARS-CoV-2 variants.

Keywords: Assembly PCR; D614G variant; gene construct; multi-epitopes; SARS-CoV-2 spike

ABSTRAK

COVID-19 yang disebabkan oleh virus SARS-CoV-2 telah menjadi ancaman sebenar kerana kemunculan varian baharu yang lebih mematikan dengan jangkitan yang lebih tinggi. Pembinaan vaksin yang mensasarkan varian SARS-CoV-2 tertentu diperlukan untuk membendung kematian disebabkan COVID-19. Glikoprotein lonjakan (S) adalah komponen antigen utama yang mencetuskan tindak balas imun perumah. Strategi vaksinologi terbalik diterapkan pada protein S daripada varian D614G COVID-19 untuk mengenal pasti protein antigen kelas tinggi. Dalam kajian ini, gen sintetik epitop pelbagai direka bentuk menggunakan strategi komputasi untuk varian D614G COVID-19. Urutan protein varian SARS-CoV-2 D614G diambil daripada pangkalan data NCBI. Ramalan epitop sel B linear dilakukan dengan menggunakan perisian ABCpred dan BepiPred 2.0 berasaskan rangkaian neural tiruan (ANN). Lima belas urutan epitop antigen teratas kemudian dipilih. Perisian Propred 1 dan Propred digunakan untuk mengenal pasti epitop pengikat kelas I dan kelas II kompleks kehistoserasian utama (MHC) dalam epitop sel B yang telah ditentukan untuk meramalkan

epitop sel T. Lima kelas I dan kelas II MHC teratas dipilih. Ujian *in-sillico* lebih lanjut untuk kelarutan, kealergenan, keantigenan dan sifat fisiokimia lain dianalisis menggunakan Bpred. Gen yang dibina dikenakan himpunan PCR dan produk gen tersebut disahkan oleh penjujukan Sanger. Hasil daripada kajian ini menunjukkan bahawa kawasan yang sangat khusus antigen varian SARS-CoV-2 D614G dapat diramalkan dalam silika dan dikembangkan menggunakan kaedah himpunan PCR. Gen sintetik yang dibangunkan menunjukkan penghasilan tindak balas imun khusus humoral dan sel yang dimediasi khusus terhadap varian SARS-CoV-2.

Kata kunci: Epitop pelbagai; himpunan PCR; pembinaan gen; peningkatan SAR-CoV-2; varian D614G

Introduction

Coronavirus Disease 2019 (COVID-19) was first reported in a 55-year old patient in Hubei province, China in December 2019 (Ma 2020). The infection has since escalated into a global pandemic with over 4 million reported deaths at the time of writing (as of 18 May 2021 with all the data extraction and analysis were conducted on May 2021 until September 2021). The disease is caused by the newly described severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from the Coronaviridae family that has been associated with acute respiratory distress syndromes, such as Middle East Respiratory Syndrome (MERS) and severe acute respiratory syndrome (SARS) (Xu et al. 2020). SARS-CoV-2 has been identified to have several structural proteins that play important roles during binding to host cells. The virus contains a positive-sense single-stranded RNA within an envelope composed of spike glycoprotein (S), membrane protein (M), nucleocapsid (N), and envelope protein (E) (Koyama et al. 2020). The full-length S protein and fragments of it such as the receptor-binding domain (RBD), S1 subunit, S2 subunit, membrane fusion peptide and N-terminal domain have been used as antigens for SARS-CoV-2 vaccines (Zhang et al. 2020). The S1 subunit and S2 subunit represents an important target for vaccine development because they are involved in binding fusion between the virus and the host cells (Dutta et al. 2008). These subunits contain the RBD that promotes viral entry through binding to the human angiotensin-converting enzyme 2 (hACE2) (Hoffmann et al. 2020). The S1 subunit engages with the hACE2 receptor during the viral entry for the replication process. Meanwhile, the S2 subunit assists the fusion binding between the S1 subunit and the hACE2 receptor by triggering several cascade reactions for membrane fusion (Lan et al. 2020). The full genetic sequence of S protein was available since January 10th, 2020 for the development of the vaccine (Walls et al. 2020). At the time of writing (as of 1 June 2021), there are at least 35 vaccine candidates utilizing the spike S protein that were undergoing clinical trials (World Health Organization 2021).

SARS-CoV-2 has been proven to undergo antigenic shift mutations that produce new variants with greater infection rates. The Centers for Disease Control and Prevention (CDC) has developed a variant classification scheme that groups variants using their PANGO lineages into the following three tiers: Variant of Interest (VOI) - B.1.525, B.1.526, B.1.526.1, B.1.617, B.1.617.1, B.1.617.2, B.1.617.3, and P.2; Variant of Concern (VOC) - B.1.1.7, B.1.351, B.1.427, P.1, B.1.617.1 and B.1.617.2; and Variant of High Consequence (VOHC) (CDC. 2021). Up to September 2021, there were yet to be any variants that were classified as VOHC. Variants of interest and variants of concern such as alpha (B.1.1.7), beta (B.1.351), gamma (P.1), kappa (B.1.617.1) and delta (B.1.617.2) all contained the key spike D614G that has been shown to increase the rate of binding fusion between the protein monomer with the hACE2 receptor during replication process (Galloway et al. 2021; Khateeb et al. 2021; Korber et al. 2020).

Lineages that contained the D614G mutation are found in 99% of the virus genomes sequenced, thus, displacing the variants that carried the original D614 found in the Wuhan reference sequence (Plante et al. 2021). The D614G replacement does not appear to alters the binding affinity towards the hACE2 receptor or neutralizing antibody, yet it increases the spike density to preserve spike integrity and to avoid S1 shedding (Zhang et al. 2020). Thus, resulting in increased infectivity and increased fitness (Hou et al. 2020). Hence, it is crucial that any future use of the spike protein as a vaccine candidate should contain this D614G. It is known that the components of adaptive immunity generated by B-cell and T-cell elicit a more specific and personalized immunity response. B-cells recognize solvent-exposed antigens through B-cell receptors (BCR), whereas T-cells recognize antigen by their specific surface T-cell receptor (TCR) on the antigen-presenting cell (APC) bound to major histocompatibility complex (MHC) class I and MHC class II. The B and T-cells epitopes are the antigen portion that binds to the immunoglobulin and MHC class I and II, respectively, and therefore targets for epitope-

based vaccines construction (Sanchez-Trincado et al. 2017).

The present study aims to identify highly antigenic regions from the spike glycoprotein of SARS-CoV-2 carrying the D614G mutation via *in silico* analysis, to construct a synthetic gene that encodes for fragments of the SARS-CoV-2 spike carrying the D614G mutation, and to assemble the gene using assembly polymerase chain reaction (PCR).

Materials and Methods

In silico preparation of oligonucleotides sequence: retrieving the protein sequences

The complete amino acid sequence of the spike protein of mutated SARS-CoV-2 D614G variant and non-mutated full spike of SARS-CoV-2 was downloaded from Protein Data Bank (PDB). The Protein Data Bank (PDB) entry with the database code 6XS6 (containing the D614G replacement), was used to extract the complete amino acid sequence without RBD of SARS-CoV-2. The amino acid sequences were then subjected to a Basic Local Alignment analysis (BLAST) search using the non-redundant option of the GenBank database was carried out to compare the open reading frame (ORF) for the spike glycoprotein sequence to confirm their similarity with other sequences of SARS-CoV-2 spikes and to identify the RBD position on the linear peptide sequence (Figure 1).

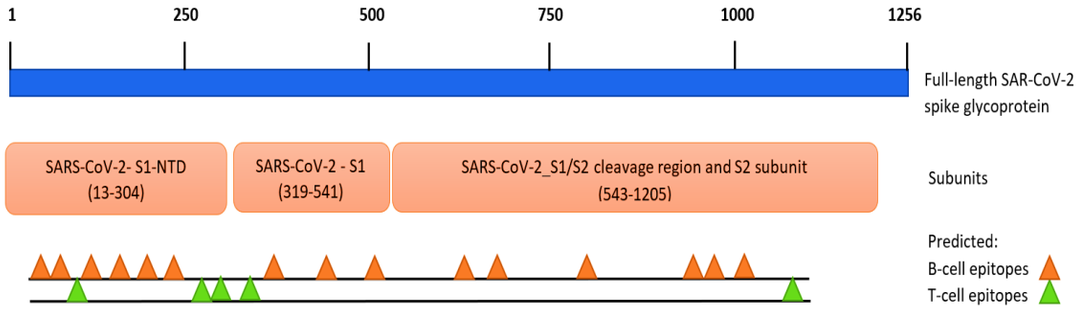

FIGURE 1. Location of the predicted B-cell and T-cell epitopes predicted by BepiPred 2.0

(NTD= N-terminal domain)

Prediction of linear B-cell epitopes using BepiPred 2.0 software

The prediction of linear B-cell epitopes was carried out using the Artificial Neural Network (ANN) based ABCpred and BepiPred 2.0 (http://crdd.osdd.net/raghava/bcepred/) (Saha & Raghava 2006). The complete amino acid sequence of the SARS-CoV-2 D614G variant and its wildtype were uploaded as FASTA format in the ABCpred server. The default threshold value of 0.51 and the window length of 20 was fixed to predict the B-cell epitopes from the complete amino acid sequences. The top 15 highly antigenic epitopes sequences were selected.

Prediction of T-cell epitopes using Propred 1 and Propred servers

Propred 1 and Propred servers were used to identify T-cell epitopes which have a binding affinity towards MHC class I and MHC class II alleles within the pre-determined B-cell epitopes. This strategy accelerates cytotoxic T-cell and helper T-cell mediated immune responses. ProPred 1 is a matrix-based approach that utilizes matrices obtained from the Bioinformatics and Molecular Analysis Section (BIMAS). Meanwhile, the quantitative matrices obtained from published literature are used by Propred (Singh & Raghava 2003).

Identification of MHC-class I epitopes

The MHC class I epitopes identification was assessed with available 47 different alleles in Propred 1 server. The pre-determined B-cell epitopes amino acid sequences were uploaded to Propred 1 server. All the available 47 alleles were selected for MHC class I identification. The server threshold was set at 4%. To increase the chances of finding correct epitopes, the option of immunoproteasome

and proteasome filters was chosen, and the threshold was fixed at 5% (Singh & Raghava 2003). Then, the common MHC class I epitopes were submitted to the VaxiJen v.2.0 server for antigenic propensity analysis with a default threshold value of 0.4 (Doytchinova & Flower 2007). The top 5 MHC class I epitopes was selected.

IDENTIFICATION OF MHC-CLASS II EPITOPES

The MHC class II epitopes identification was carried out using the pre-identified B-cell epitopes against 51 different MHC Class II alleles available in the Propred server. For MHC class II identification, the epitopes were chosen based on epitopes predicted by at least 10 different alleles from the server. The amino acid sequence of pre-determined B-cell epitopes was uploaded to the Propred server, and all 51 alleles from the server were selected. The threshold value was fixed at 0.4. The common MHC class II was submitted to the VaxiJen v.2.0 server for antigenic propensity analysis with a default threshold value of 0.4. The top 5 MHC class II was selected.

PREDICTION OF BINDING AFFINITY WITH MHCPRED

The prediction of the binding affinity of the chosen MHC class I and MHC class II epitopes were carried out using the MHCPred server. The predicted MHC class I and MHC class II epitopes with a VaxiJen score of >1.0 was selected from the predicted epitopes list. The chosen epitopes were then subjected to further assessment for binding affinity against HLA A*1101 and DRB1*0101, respectively, using MHCPred version 2.0 (Guan et al. 2003). Epitopes with a half maximum inhibitory concentration (IC50) value <100 nM were shortlisted as a strong candidate for high immunogenicity-associated multi-epitope gene construction. The top 5 from the list of MHC class I and MHC class II epitopes were selected to proceed with multi-epitope gene construction.

SYNTHETIC GENE CONSTRUCT DESIGN

The selected epitope candidates from MHC class I and MHC class II in amino acid form were connected using peptide linkers; AAY and GPGPG. The chosen top 5 epitopes from MHC class I was associated using the AAY peptide linker, followed by the top 5 epitopes from MHC class II using the GPGPG peptide linker. The whole connected multi-epitope sequences were coupled with restriction enzymes (RE) at the 5' and 3' end of the sequence. This step is important as the usage of suitable RE cleaves multiple-cloning sites in a vector and for proper binding during the further cloning process. The suitable RE was chosen by the RE analysis, carried out by the GenScript server. The designed multi-epitope sequence was uploaded to the GenScript server. The server analyzed the enzyme cutting sequences and gives a list of enzymes free from cutting the sequences. The RE; *Nde*I and *Bam*HI were selected. *Nde*I was added at 5' of the sequence and *Bam*HI at 3' of the sequence.

CODON OPTIMIZATION

Codon optimization was carried out using the ExpOptimizer tool from NovoProLabs for the maximum expression of any target proteins in any mainstream expression hosts. Dozens of primary factors of gene transcription and translation are considered by its extensive codon optimization algorithm. The raw amino acid sequence of the designed multi-epitope gene sequence was uploaded in the ExpOptimizer tool. The expression host *Escherichia coli* was selected, and the restriction enzyme *Nde*I and *Bam*HI were excluded from the list of enzymes listed in the server. The multi-epitopes gene sequence was optimized based on the codon adaptation index (CAI) value and GC content before the expression host, *E. coli*. The refined gene sequence was downloaded in a nucleotide sequence format.

EVALUATION OF ANTIGENICITY, ALLERGENICITY, SOLUBILITY, AND PHYSICOCHEMICAL PROPERTIES

The antigenicity of the gene sequence was evaluated using the VaxiJen v.2.0 server with the threshold was fixed at 0.4. Allergenicity screening of the final vaccine construct was verified using AllerTOP v. 2.0 and AllergenFP v.1.0 servers (Dimitrov et al. 2014). This is essential as the vaccine should not cause sensitization and allergic reaction within the body. Using Protein-Sol server, the solubility of vaccine build upon expression in *E. coli* was evaluated. Solubility testing is important for the production of recombinant protein in biotherapeutic (Hebditch et al. 2017). Physicochemical parameters of the construction were evaluated by Expasy ProtParam server (Wilkins et al. 1999) to assess the number of amino acids, molecular weight, theoretical isoelectric point, the total number of atoms, chemical formulae, estimated half-life, instability index, aliphatic index, and grand average of hydropathicity (GRAVY).

TERTIARY STRUCTURE PREDICTION, REFINEMENT, VALIDATION OF PROTEIN AND CONFORMATIONAL B-CELL EPITOPE PREDICTION

The multi-epitope-based gene's three-dimensional (3D) structure was created using the 3Dpro SCRATCH suite server (Cheng et al. 2005). Refinement of the model

vaccine structures was later carried out through the GalaxyRefine webserver (Heo et al. 2013). Using ProSA-web and PROCHECK server, this server initially reforms the side chains and performs sidechain repacking and eventually overall structural relaxation by molecular dynamics simulation. The refined model was finally validated to identify any possible errors. The DiscoTope 2.0 of the IEDB server tool was used to evaluate conformation B-cell epitopes. As an input, using the validated 3D structure of the multi-epitope-based gene, the server introduces a new description of the spatial neighborhood as a surface indicator of inclination scores and half-sphere exposure (Kringelum et al. 2012). The multi-epitope 3D gene design was uploaded as PDB file format in DiscoTope 2.0 of the IEDB server tool. The threshold was set to default at -3.70. The confirmative B-cell epitopes are analyzed.

OLIGONUCLEOTIDES AND PRIMERS SYNTHESIS AND PREPARATION

The *in silico* oligonucleotides and primers designing was carried out using FastPCR 6.7 server. The optimized nucleotide sequences were uploaded to the server with the minimal and maximal oligonucleotide lengths were fixed at 40 and 80, respectively. The minimal overlapping length was fixed at 20. The list of oligonucleotides and primers that was obtained with its respective calculated melting temperature were synthesized by Integrated DNA Technologies (IDT), Singapore, summarized in Table 1. The oligonucleotides and primers in the lyophilized powder were dissolved in TE buffer to a final concentration of 250 µM (as a stock solution). To prepare a working concentration, 30 µL of the stock solution (250 µM) was diluted 1:5 with 120 µL of TE buffer to make up the final concentration of 50 µM. The working solution and stock solution were stored at -20 °C for later use.

TABLE 1. List of oligonucleotides and primers for assembly PCR designed by FastPCR 6.7

Primer ID	Sequence (5'-3')	Length (nt)	Tm (°C)	Linguistic complexity (%)
1F1_1-50	CATATGGTTGTCGTTCTGTCCTTTGAACTGCTGGCCGCCTATAAAATCGC	50	73.8	86
1R2_31-80	CAATAAGCAGCCAGTTTGTAGTTGTAATCAGCGATTTTATAGGCGGCCAG	50	71.9	76
1F3_61-113	TACAAACTGGCTGCTTATTGTGTTATCGCAGGCACGATCACCGACGCAGCTTA	53	75.5	82
1R4_94-143	GCCAGGGATTGGGTCTTGCTGTCCAGGGTGTAAGCTGCGTCGGTGATCGT	50	79.2	78
1F5_124-173	AGCAAGACCCAATCCCTGGCGGCTTACCTGACGACCCTGGATTCTAAGAC	50	76.8	78
1R6_154-214	TCAGGGTCGGCTGTACACGGAAACCAGGGCCCGGACCCTGGGTCTTAGAATCCAGGGTCGT	61	81.9	77
1F7_195-258	CCGTGTACAGCCGACCCTGATCGTGGGTCCGGGTCCGGGTGTAGTCTTCCTGCACGTTACTTAC	64	81.1	74
1R8_239-300	ACGGTATGCCATCTGCATTGCGAAACCCGGGCCCGGACCAACGTAAGTAACGTGCAGGAAGA	62	80.4	79
1F9_281-339	CAATGCAGATGGCATACCGTTTCGGTCCGGGCCCGGGCCTGGTTAAAAACAAATGCGTG	59	80.1	79
1R10_320-393	GGATCCTACTTCGGTGCAGTTGACGCCGTTGAAACCCGGGCCCGGGCCGAAGTTCACGCATTTGTTTTTAACCA	74	81.8	80
1F_1-28	CATATGGTTGTCGTTCTGTCCTTTGAAC	28	60.6	87
1R_393-373	GGATCCTACTTCGGTGCAGTT	21	57.9	97

ASSEMBLY PCR

Assembly PCR was carried out in two steps. The first step combined all the oligonucleotides to create a larger DNA fragment. The DNA fragment was then amplified in the second PCR using a pair of specific primers to check if a correct fragment was amplified from the first PCR product. In the first step, a 25 µL reaction mixture containing 1.25 µL (2.5 µM) of oligonucleotides mixture, 1.0 µL (10mM) of the dNTPs mixture, 2.5 µL of 10× PCR buffer, 7.5 µL (1.5mM) of MgSO4, 0.5 µL (1 unit/µL) of *Taq* polymerase and 12.25 µL of sterilised deionised distilled water was prepared. The mixture was then subjected to first PCR (55 cycles of 95 °C for 60 s, 57 °C for 60 s and 72 °C for 60 s). In the second PCR, a reaction mixture containing 5 µL of the first PCR product, 2.5 µL (2.5 µM) of forward and reverse primers, 1.0 µL (10 mM) of the dNTPs mixture, 2.5 µL of 10× PCR buffer, 7.5 µL (1.5 mM) of MgSO4, 0.5 µL (1 unit/µL) of Taq polymerase and 6.0 µL of sterilised deionised distilled water was prepared to make up a final volume of 25 µL. The mixture was then subjected to second PCR (35 cycles of 95 °C for 60 s, 60 °C for 60 s and 72 °C for 60 s). The products from the first and second PCR were electrophoresed by 1.5% (w/v) agarose gel electrophoresis and the band was observed under fluorescence transilluminator.

SANGER SEQUENCING

The amplicons from the second PCR were purified using QIAquick® PCR Purification kit (Qiagen, Germany) and directly send for Sanger Sequencing (Apical Scientific, Selangor, Malaysia). The result was aligned and verified using Clustal Omega software. The similarities and gaps of the sequences were calculated using Water EMBOSS software.

RESULTS AND DISCUSSION

PREDICTION OF LINEAR B-CELL AND T-CELL EPITOPES

The complete 1257 amino acid sequence of S protein of SARS-CoV-2 D614G variant was retrieved from the PDB entry with the code 6XS6 as input for the prediction of linear B-cell and T-cell epitopes. A total of 102 linear B-cell epitopes of 20 amino acid window length is retrieved within the spike protein of the SARS-CoV-2 D614G variant. The chosen 15 linear B-cell epitopes with higher immunogenicity and antigenicity score are shown in Table 2. Immunogenicity is an essential criterion to identify epitopes that provoke an immune response and to determine the magnitude of an immune response. On the other hand, antigenicity was taken into consideration to identify epitopes that bind specifically with T cell receptors or other receptors that have adaptive immunity (Schindewolf & Menachery 2019). Meanwhile, a total of 50 and 25 B-cell derived MHC class I and class II epitopes, respectively, are predicted by ≥ 5 MHC I and II alleles using Propred 1. Out of the respective total epitopes, 5 epitopes were found with above threshold VaxiJen and binding affinity against HLA A*1101 and DRB1 *0101 predicted by MHC class I and class II, respectively, as shown in Table 3. The locations of the predicted B-cell and T-cell epitopes predicted by BepiPred 2.0 are illustrated in Figure 1. HLA A*1101 and DRB1*0101 were chosen because they are the common alleles for MHC molecules in the human population (Bhattacharya et al. 2020). The focus is set to identify potential B-cell derived T-cell epitopes to produce a potential synthetic gene that can be used as a vaccine construct. This strategy will help the human body to induce both cellular and humoral immune responses simultaneously. The B-cell derived epitopes will stimulate the immune system to elicit a humoral response, while T-cells epitopes derived from B-cells will stimulate a cell-mediated immune response. Thus, it is vital to critically choose the epitopes, which is must be accessible by both MHC class I and MHC class II molecules along with B-cells (Patra et al. 2020). The top 15 B-cell epitopes were chosen based on the immunogenicity and antigenicity score. The MHC class I and class II peptides were combined and re-constructed by adding AAY and GPGPG peptide linkers, respectively as shown below:

Newly constructed MHC class I peptide with linkers:
VVVLSFELL-AAY-KIADYNYKL-AAY-CVIAGTITD-AAY-TLDSKTQSL-AAY-LTTLDSKTQ

Newly Constructed MHC Class II peptide with linkers:
FRVQPTLIV-GPGPG-VVFLHVTYV-GPGPG-FAMQMAYRF-GPGPG-LVKNKCVNF-GPGPG-FNGVNCTEV

Multi-epitope design:
VVVLSFELLAAYKIADYNYKLAAYCVIAGTITTDAAY TLDSKTQSLAAYLTTLDSKTQGPGPGFRVQPPTLIV GPGPGVVFLHVTYVGPGPGFAMQMAYRFGPGPG LVKNKCVNFGPGPGFNGVNCTEV

TABLE 2. List of linear B-cell epitopes sequence predicted by BepiPred 2.0

Residues no.	Peptide sequence	Nucleotide sequence	Score
826	ADAGFIKQYGDCLGDIAARD	GCCGACGCCGGCTTCATCAAGCAGTACGGCGACTGCCTGGGCGACATCGCCGCCAGGGAC	0.90
651	IGAEHVNNSYECDIPIGAGI	ATCGGCGCCGAGCACGTGAACAACAGCTACGAGTGCGACATCCCCATCGGCGCCGGCATC	0.90
393	TNVYADSFVIRGDEVRQIAP	ACCAACGTGTACGCCGACAGCTTCGTGATCAGGGGCGACGAGGTGAGGCAGATCGCCCCC	0.90
143	VYYHKNNKSWMESEFRVYSS	GTGTACTACCACAAGAACAACAAGAGCTGGATGGAGAGCGAGTTCAGGGTGTACAGCAGC	0.90
68	IHVSGTNGTKRFDNPVLPFN	ATCCACGTGAGCGGCACCAACGGCACCAAGAGGTTCGACAACCCCGTGCTGCCCTTCAAC	0.89
614	GVNCTEVPVAIHADQLTPTW	GGCGTGAACTGCACCGAGGTGCCCGTGGCCATCCACGCCGACCAGCTGACCCCCACCTGG	0.89
247	SYLTPGDSSSGWTAGAAAYY	AGCTACCTGACCCCCGGCGACAGCAGCAGCGGCTGGACCGCCGGCGCCGCCGCCTACTAC	0.89
974	LNDILSRLDPPEAEVQIDRL	CTGAACGACATCCTGAGCAGGCTGGACCCCCCCGAGGCCGAGGTGCAGATCGACAGGCTG	0.88
964	SSNFGAISSVLNDILSRLDP	AGCAGCAACTTCGGCGCCATCAGCAGCGTGCTGAACGACATCCTGAGCAGGCTGGACCCC	0.88
206	KHTPINLVRDLPQGFSALEP	AAGCACACCCCCATCAACCTGGTGAGGGACCTGCCCCAGGGCTTCAGCGCCCTGGAGCCC	0.88
126	VVIKVCEFQFCNDPFLGVYY	GTGGTGATCAAGGTGTGCGAGTTCCAGTTCTGCAACGACCCCTTCCTGGGCGTGTACTAC	0.88
337	PFGEVFNATRFASVYAWNRK	CCCTTCGGCGAGGTGTTCAACGCCACCAGGTTCGCCAGCGTGTACGCCTGGAACAGGAAG	0.87
1130	VNNTVYDPLQPELDSFKEEL	GTGAACAACACCGTGTACGACCCCCTGCAGCCCGAGCTGGACAGCTTCAAGGAGGAGCTG	0.86
505	YQPYRVVVLSFELLHAPATV	TACCAGCCCTACAGGGTGGTGGTGCTGAGCTTCGAGCTGCTGCACGCCCCCGCCACCGTG	0.85
44	RSSVLHSTQDLFLPFFSNVT	AGGAGCAGCGTGCTGCACAGCACCCAGGACCTGTTCCTGCCCTTCTTCAGCAACGTGACC	0.85

TABLE 3. List of epitopes encountering the MHC class I and class II alleles, residues number, VaxiJen antigenic score and IC_{50} value

Residues no.	Peptide Sequence	Real score	VaxiJen2.0 antigenic score	MHCPred HLA A*1101 IC_{50} Value (nM)
MHC Class I				
306	VVVLSFELL	2.400	1.0909	64.54
1382	KIADYNYKL	2.000	1.6639	53.20
1397	CVIAGTITD	1.200	1.3029	43.21
282	TLDSKTQSL	0.800	1.0685	59.54
280	LTTLDSKTQ	0.240	1.0841	73.21
MHC Class II				
1174	FRVQPTLIV	2.2000	1.4344	14.09
1587	VVFLHVTYV	2.0900	1.5122	88.30
1867	FAMQMAYRF	1.6000	1.0278	19.40
1726	LVKNKCVNF	1.4400	1.3263	10.99
98	FNGVNCTEV	1.3400	1.0939	39.08

Selected peptides based on pre-defined criteria: VaxiJen score >1.0 and the total number of different binding alleles must be ≥5 and the binding affinity against HLA A*1101 and DRB1 *0101 for MHC class I and II, respectively (IC_{50} < 100 nM)

CODON OPTIMIZATION

The sequence was optimized with inputs of *E. coli* as the expression host. The optimized sequence result is shown in Figure 2. The increase in codon adaptation index (CAI) indicates the level of gene expression, given that translational selection is made to optimize gene sequences according to their levels of expression (Suzuki & Morton 2016). Meanwhile, the GC content was reduced from the initial content. The GC content was optimized because an optimum GC content matching *E. coli's* GC content is vital to correlate higher optimum temperature and a broader tolerance range for a species (Musto et al. 2006). RE, NdeI, and BamHI were added at the 5' and 3' end of the designed gene to enable the synthetic gene to be cloned into the vector, pET 28 for future study purposes. The NdeI was added at 5' and BamHI at 3' end of the designed gene sequence to produce a sticky end during cloning in pET 28 system. The NdeI was added at 5' end because it contains a start codon, which confirms the gene translation process in the designed gene sequence (Figure 3).

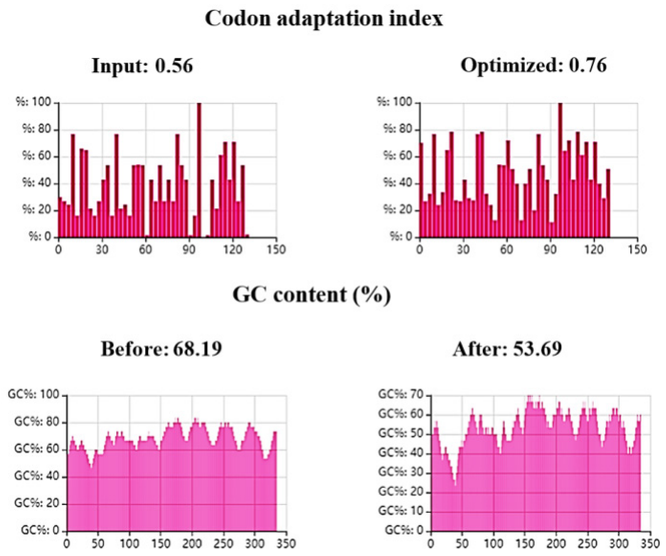

FIGURE 2. The codon optimization components of synthetic gene construct by using ExpOptimizer

5'CATATGGTTGTCGTTCTGTCCTTTGAACTGCTGGCCGCCTATAAAATCGCTGATTAC
AACTACAAACTGGCTGCTTATTGTGTTATCGCAGGCACGATCACCGACGCAGCTTACAC
CCTGGACAGCAAGACCCAATCCCTGGCGGCTTACCTGACGACCCTGGATTCTAAGACCC
AGGGTCCGGGCCCTGGTTTCCGTGTACAGCCGACCCTGATCGTGGGTCCGGGTCCGGGT
GTAGTCTTCCTGCACGTTACTTACGTTGGTCCGGGCCCGGGTTTCGCAATGCAGATGGC
ATACCGTTTCGGTCCGGGCCCGGGGCTGGTTAAAAACAAATGCGTGAACTTCGGCCCGG
GCCCGGGTTTCAACGGCGTCAACTGCACCGAAGTAGGATCC3'

FIGURE 3. Finalized synthetic gene design. The multi-epitope gene design with the addition of restriction enzyme NdeI (yellow) at 5' and BamHI (green) at 3' end in nucleotide form

EVALUATION OF ANTIGENICITY, ALLERGENICITY, SOLUBILITY, AND PHYSICOCHEMICAL PROPERTIES

The *in silico* prediction ensured higher stability, expression score and also glycosylation of proteins after the gene was modified. The glycosylation provided some benefits to our vaccine design, which specifically targets and enhance biodistribution by increasing the receptor-binding affinity in tissues to ACE-receptor. The glycosylation was intended to increase metabolic stability and lower the clearance rate when administer to animal models and humans. It also protects the peptide from oxidation and stabilizes the physical properties of peptides (Di 2015). From the physicochemical evaluation of constructed synthetic gene (Table 4), the gene construct was predicted as a probable antigen (0. 9626 by VaxiJen 2.0) that is soluble in water, which showed the design was an efficient, precise, robust and long-lasting immune response. The construct was also found as probable non-allergen, indicating that when administered to the human body as a vaccine, the synthetic gene would not evoke any allergic reaction. A pI of 8.48 indicates that the gene construct is basic in nature and most stable at this pH range using ProtParam analysis (Saha et al. 2021). The molecular weight of 13.69 kiloDalton (kDa) which is less than 110 kDa is considered as a good vaccine candidate (Naz et al. 2015). On the other hand, the aliphatic index showed that the gene construct is potentially thermostable at different temperatures. This is an essential factor to increase the thermostability of globular proteins which are occupied by aliphatic side chains such as alanine, isoleucine, valine, and leucine. The constructed synthetic gene formulation is predicted to be stable after expression, as the instability index is 11.01, which is less than 40.00. The Grand average of hydropathicity (GRAVY) value was computed to be positive (0.271). This shows that the synthetic gene is hydrophobic in nature and increases of thermostability of globular proteins (Gurung 2020). The hydrophobic synthetic gene might face problems resolving in gel in terms of electrophoresis and tend to solubilize (Rehman et al. 2020).

TABLE 4. The evaluation of antigenicity, allergenicity, solubility, and physicochemical properties

Features	Assessment
Antigenicity	0. 9626 (Probable antigen)
Allergenicity	Probable non-allergen (AllerTOP v. 2.0)
	Probable non-allergen (AllergenFP v.1.0)
Solubility	0.606 (Soluble)
Number of amino acids	131
Molecular weight (Dalton)	13693.84
Theoretical Isoelectric point (pI)	8.48
Total number of atoms	1929
Formula	C627H962N156O176S6
Estimated half-life	3.5 h (mammalian reticulocytes, *in vitro*).
	10 min (yeast, *in vivo*).
	>10 h (*Escherichia coli, in vivo*).
Instability index	11.01 (Stable)
Aliphatic index	84.81
Grand average of hydropathicity (GRAVY)	0.271

TERTIARY STRUCTURE PREDICTION, REFINEMENT, VALIDATION OF PROTEIN AND CONFORMATIONAL B-CELL EPITOPE PREDICTION

A total of 33 residues out of 131 amino acids of synthetic gene construct were predicted as discontinuous B-cell epitopes at the threshold score of -3.70 by using DiscoTope. The generated refined tertiary structure of synthetic gene construct was found valid for identification of conformational epitopes and suitable for molecular docking experiments (Figure 4(a)). The ProSA Z-score (-5.01) falls within the range characteristics of the native protein structure which indicates the overall quality score of the model protein by using computational X-ray and Nuclear magnetic resonance (NMR) analysis (Figure 4(b)). The PROCHECK Ramachandran plot was used to determine the energetically allowed and disallowed psi (ψ) and phi (Φ) dihedral angles of amino acids, measured based on the van der Waal radius of the side chain. The synthetic gene design shows 73% of most favored regions, 20% of additional allowed areas and 1% disallowed areas, respectively (Figure 4(c)). The disallowed regions showed a negligible number of steric clashes between the side-chain atoms and main-chain atoms. A good quality model would be expected to have over 90% in the most favored regions and has less than 1.5% disallowed areas, based on an analysis of 118 structures of resolution of at least 2.0 Angstroms (Laskowski et al. 2006). From these results, it could be concluded that this synthetic gene design has a small range of drawbacks in terms of torsional angles of amino acids. The predicted conformational B-cell epitope sequences were shown in Figure 4(d).

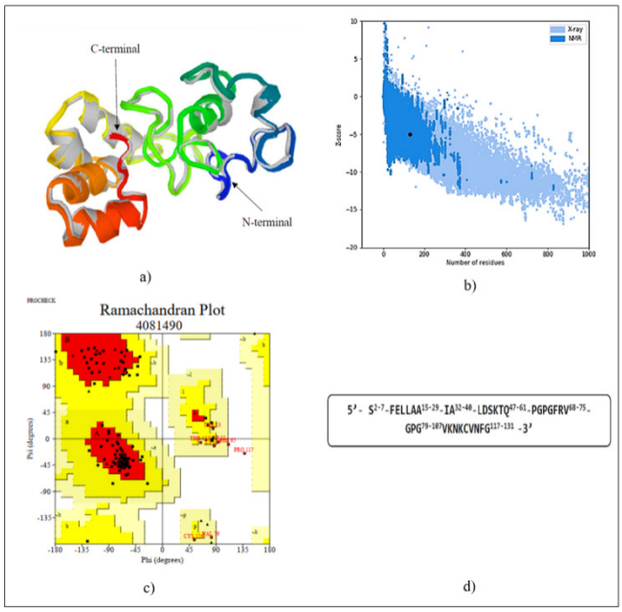

FIGURE 4. a) 3D modelling of the refined tertiary structure of designed synthetic gene of COVID-19 variant D614G.b) ProSA Z-score showing the overall model quality with predicted Z-score value, -5.01. c) The Ramachandran plot generated using PROCHECK for the multi-epitope synthetic gene construct of COVID-19 variant D614G. The areas showing different colours where red represents, most favoured regions, yellow represents additional allowed regions, and light yellow represent disallowed regions, respectively. d) The predicted conformational B-cell epitope sequences

ASSEMBLY PCR AND SEQUENCING

The expected band at 393 bp of the constructed gene was successfully amplified by assembly PCR (Figure 5). Analysis of the amplified product's sequence using Clustal Omega software showed that the synthetic gene sequence was aligned to the D614G wildtype sequence. The Water EMBOSS analysis showed 95.9% similarity, 348/363 sequence matched, and 0.8% gaps (Figure 6). These results indicated that the synthetic gene was successfully produced, validated by the sequencing.

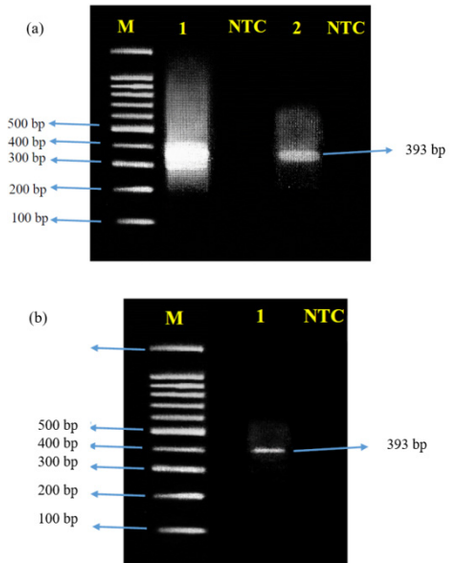

FIGURE 5. Assembly PCR. (a) Lane 1 represents the assembled gene (in the first PCR) and lane 2 represents the amplified gene (in the second PCR). A clear 393 bp synthetic gene band with slight smearing close to 400 bp can be seen by 1.5% (w/v) agarose gel electrophoresis, (b) purified synthetic gekne (M=100 100 bp DNA ladder, NTC= non template control)

```
  1 AGGGGGGCGNCGCCTATAAATCGCTGATTACAACTACAAACTGGCTGCTT    50
    ||||||||| ||||||||||||||||||||||||||||||||||||||||
  1 AGGGGGGCGNCGCCTATAAATCGCTGATTACAACTACAAACTGGCTGCTT    50

 51 ATTGTGTTATCGCAGGCACGATCACCGACGCAGCTTACACCCTGGACAGC    100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 ATTGTGTTATCGCAGGCACGATCACCGACGCAGCTTACACCCTGGACAGC    100

101 AAGACCCAATCCCTGGCGGCTTACCTGACGACCCTGGATTCTAAGACCCA    150
    |||||||||||||||||.|||||||||||||||||||||||||||||||
101 AAGACCCAATCCCTGGCAGCTTACCTGACGACCCTGGATTCTAAGACCCA    150

151 GGGTCCGGGCCCTGGTTTCCGTGTACAGCCGACCCTGATCGTGGGTCCGG    200
    ||||||||||||||||||||||||||||||||||||.|||||.|||||
151 GGGTCCGGGCCCTGGTTTCCGTGTACAGCCGACCCTGAACGTGGCTCCGG    200

201 GTCCGGGTGTAGTCTTCCTGCACGTTACTTACGTTGGTCCGGGCCCGGGT    250
    ||||||||||||||||||||||||||||||||||||.|||||||||||||
201 GTCCGGGTGTAGTCTTCCTGCACGTTACTTACGTTGATCCGGGCCCGGGT    250

251 TTCGCAATGCAGATGGCATACCGTTTCGGTCCGGGCCCGGGGGNNGTTAA    300
    |||||||||||||||||||||||||.|||||||||.|||||||  |.|||
251 TTCGCAATGCAGATGGCATACCGTATCGGTCCGGCCCCGGGGGNNGGTAA    300

301 AAACAAATGCGTGAACTTCGGCCCGGGCCCGCGCTCTAACGGGAGAAACT    350
    |||.|.|.|||||||||.|||||||.|||||||||||||||||||||||
301 AAAGAGAGGCGTGAACTTAGGCCCGGTCCCGCGCTCTAACGGGAGAAACT    350

351 GCACCGAAGTAGGATCCAGTGGAGTAGTGATTATCTTTTCTCTTCAANTA    400
    ||||||.||.|||||||.|||||||||||||||||||||||||||| ||
351 GCACCGGAGGAGGATCAAGTGGAGTAGTGATTATCTTTTCTCTTCAANTA    400

401 GGAGGTGCTTCATTAGTGGNGGTGATCATTATTTTTACTACAATTATTAA    450

401 --------------------------------------------------    400
```

← Wild-type

← D614G synthetic spike fragment gene sequence

FIGURE 6. Sanger Sequencing. Sequence alignment of the synthetic gene construct (lower arrow) in comparison to wildtype sequence (upper arrow) (% similarity= 95.9%, sequence matched= 348/363, % gaps= 0.8%)

Conclusion

In conclusion, the *in silico* model of the SARS-CoV-2 D614G variant was able to be constructed using the reverse vaccinology approach. The present immunoinformatic analysis identified 5 MHC class 1 and 5 MHC class II epitopes found with above threshold VaxiJen and binding affinity against HLA A*1101 and DRB1*0101 alleles, respectively, within the spike glycoprotein of SARS-CoV-2 D614G variant. D614G variant was chosen due to at the time of the study in August 2020, the D614G variant was the first mutated variant formed from the parent, Wuhan SARS-CoV-2. Now D614G became dominant, and present in all variants; Alpha, Beta, Gamma, Delta, and Omicron. We also have tested the regions that are not hotspots for mutation, but the epitopes resulted in very low binding affinity and elicited a very low immune response score from *in silico* prediction (VaxiJen score >1.0 and the total number of different binding alleles must be ≥5 and the binding affinity against HLA A*1101 (IC_{50}<100 nM)). Several studies have shown the effect of mutational hotspot in strengthen the RBD-ACE2 binding affinity *in silico* (Nelson-Sathi et al. 2022; Verma & Subbarao 2021). The current peptide construct had all parameters within the recommended range, thus, an ideal candidate for future vaccine construct. The first epitope-based subunit vaccine was successfully developed in 1985 against Cholera, and many are under development for Influenza, Malaria, Swine-flu, and Anthrax vaccines. Some of the multi-epitope vaccine against solid tumors have entered phase III clinical trial (Nain et al. 2020). In October 2021, WHO has endorsed the use of Malaria vaccine, RTS, S (Mosquirix), an epitope-focused recombinant vaccine, in children in Sub-Saharan Africa and related area with high *Plasmodium falciparum* malaria transmission (CDC 2022). Currently, Corbevax, a peptide-based COVID-19 vaccine has received an Emergency Use of Authorization (EUA) from the drug controller General of India (DCGI) in February 2022 for use in children with age ranged 12-18 years.

Acknowledgements

URICAS for providing research initiative platform, Biomedical Science Programme, School of Health Sciences, USM Health Campus, Institute for Research in Molecular Medicine (INFORMM), USM, Penang. The authors would like to acknowledge financial support from Universiti Sains Malaysia and the Ministry of Higher Education, Malaysia grant (PRGS: 203. PPSK.6740076).

References

Bhattacharya, M., Sharma, A.R., Patra, P., Ghosh, P., Sharma, G., Patra, B.C., Lee, S.S. & Chakraborty, C. 2020. Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-COV-2): Immunoinformatics approach. *Journal of Medical Virology* 92(6): 618-631.

CDC. 2022. *Malaria Vaccine Recommended for Broader Use by WHO: Best Thing Since Bed Nets*. https://www.cdc.gov/parasites/features/malaria_vaccine_ who.html

CDC. 2021. *SARS-CoV-2 Variant Classifications and Definitions*. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/ variant-info.html#Consequence

Cheng, J., Randall, A.Z., Sweredoski, M.J. & Baldi, P. 2005. SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Research* 33(Web Server issue): W72-W76.

Di, L. 2015. Strategic approaches to optimizing peptide ADME properties. *The AAPS Journal* 17(1): 134-143.

Dimitrov, I., Bangov, I., Flower, D.R. & Doytchinova, I. 2014. AllerTOP v.2 - A server for *in silico* prediction of allergens. *Journal of Molecular Modeling* 20(6): 2278.

Doytchinova, I.A. & Flower, D.R. 2007. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. *Vaccine* 25(5): 856-866.

Dutta, N.K., Mazumdar, K., Lee, B.H., Baek, M.W., Kim, D.J., Na, Y.R., Park, S.H., Lee, H.K., Kariwa, H. & Park, J.H. 2008. Search for potential target site of nucleocapsid gene for the design of an epitope-based SARS DNA vaccine. *Immunology Letters* 118(1): 65-71.

Galloway, S.E., Paul, P., MacCannell, D.R., Johansson, M.A., Brooks, J.T., MacNeil, A., Slayton, R.B., Tong, S., Silk, B.J., Armstrong, G.L. & Biggerstaff, M. 2021. Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020 - January 12, 2021. *MMWR Morb Mortal Wkly Rep* 70(3) 95-99.

Guan, P., Doytchinova, I.A., Zygouri, C. & Flower, D.R. 2003. MHCPred: A server for quantitative prediction of peptide-MHC binding. *Nucleic Acids Research* 31(13): 3621-3624.

Gurung, A.B. 2020. *In silico* structure modelling of SARS-CoV-2 Nsp13 helicase and Nsp14 and repurposing of FDA approved antiviral drugs as dual inhibitors. *Gene Reports* 21: 100860-100860.

Hebditch, M., Carballo-Amador, M.A., Charonis, S., Curtis, R. & Warwicker, J. 2017. Protein-sol: A web tool for predicting protein solubility from sequence. *Bioinformatics (Oxford, England)* 33(19): 3098-3100.

Heo, L., Park, H. & Seok, C. 2013. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Research* 41(Web Server issue): W384-W388.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, S., Herrler, G., Wu, N.H., Nitsche, A., Muller, M.A., Drosten, C. & Pöhlmann, S. 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181(2): 271-280.

Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon III, K.H., Leist, S.R., Schafer, A., Nakajima, N., Takahashi, K. & Lee, R.E. 2020. SARS-CoV-2 D614G variant exhibits efficient replication *ex vivo* and transmission *in vivo. Science (New York)* 370(6523): 1464-1468.

Khateeb, J., Li, Y. & Zhang, H. 2021. Emerging SARS-CoV-2 variants of concern and potential intervention approaches. *Critical Care* 25(1): 244.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B. & Hastie, K.M. 2020. Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 Virus. *Cell* 182(4): 812-827.

Koyama, T., Platt, D. & Parida, L. 2020. Variant analysis of SARS-CoV-2 genomes. *Bulletin of the World Health Organization* 98(7): 495-504.

Kringelum, J.V., Lundegaard, C., Lund, O. & Nielsen, M. 2012. Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLOS Computational Biology* 8(12): e1002829.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L. & Wang, X. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581(7807): 215-220.

Laskowski, R.A., MacArthur, M.W. & Thornton, J.M. 2006. PROCHECK: Validation of protein-structure coordinates. In *International Tables for Crystallography.* The Netherlands: Kluwer Academic Publishers.

Ma, J. 2020. *Coronavirus: China's first confirmed COVID-19 case traced back to November 17 South China Morning Post*. https://www.scmp.com/news/china/ society/ article/3074991/coronavirus-chinas-first-confirmed-covid-19-case-traced-back.

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F. & Bernardi, G. 2006. Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochemical and Biophysical Research Communications* 347(1): 1-3.

Nain, Z., Karim, M.M., Sen, M.K. & Adhikari, U.K. 2020. Structural basis and designing of peptide vaccine using PE-PGRS family protein of *Mycobacterium ulcerans* - An integrated vaccinomics approach. *Molecular Immunology* 120: 146-163.

Naz, A., Awan, F.M., Obaid, A., Muhammad, S.A., Paracha, R.Z., Ahmad, J. & Ali, A. 2015. Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: A reverse vaccinology based approach. *Infection, Genetics and Evolution* 32: 280-291.

Nelson-Sathi, S., Umasankar, P.K., Sreekumar, E., Nair, R.R., Joseph, I., Nori, S.R.C., Philip, J.S., Prasad, R., Navyasree, K.V., Ramesh, S. & Pillai, H. 2022. Mutational landscape and in silico structure models of SARS-CoV-2 spike receptor binding domain reveal key molecular determinants for virus-host interaction. *BMC Molecular and Cell Biology* 23(1): 1-12.

Patra, P., Mondal, N., Patra, B.C. & Bhattacharya, M. 2020. Epitope-based vaccine designing of nocardia asteroides targeting the virulence factor mce-family protein by immunoinformatics approach. *International Journal of Peptide Research and Therapeutics* 26(2): 1165-1176.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R. & Mirchandani, D. 2021. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592(7852): 116-121.

Rehman, S.U., Nadeem, A., Javed, M., Hassan, F-U., Luo, X., Khalid, R.B. & Liu, Q. 2020. Genomic identification, evolution and sequence analysis of the heat-shock protein gene family in buffalo. *Genes* 11(11): 1388.

Saha, R., Ghosh, P. & Burra, V.L.S. 2021. Designing a next generation multi-epitope based peptide vaccine candidate against SARS-CoV-2 using computational approaches. *3 Biotech* 11(2): 47.

Saha, S. & Raghava, G.P.S. 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics* 65(1): 40-48.

Sanchez-Trincado, J.L., Gomez-Perosanz, M. & Reche, P.A. 2017. Fundamentals and methods for T- and B-Cell epitope prediction. *Journal of Immunology Research* 2017: 2680160.

Schindewolf, C. & Menachery, V.D. 2019. Middle East respiratory syndrome vaccine candidates: Cautious optimism. *Viruses* 11(1): 74.

Singh, H. & Raghava, G.P.S. 2003. ProPred1: Prediction of promiscuous MHC Class-I binding sites. *Bioinformatics* 19(8): 1009-1014.

Suzuki, H. & Morton, B.R. 2016. Codon adaptation of plastid genes. *PLoS ONE* 11(5): e0154306.

Verma, J. & Subbarao, N. 2021. *Insilico* study on the effect of SARS-CoV-2 RBD hotspot mutants' interaction with ACE2 to understand the binding affinity and stability. *Virology* 561(1): 107-116.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T. & Veesler, D. 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 181(2): 281-292.

Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J-C., Williams, K.L., Appel, R.D. & Hochstrasser, D.F. 1999. In *Protein Identification and Analysis Tools in the ExPASy Server*. *Methods Mol. Biol.* 112: 531-552.

World Health Organization (WHO). 2021. *Draft Landscape and Tracker of COVID-19 Candidate Vaccines.* Accessed on 25 May 2021.

Xu, J.B., Zhao, S.Z., Teng, T.S., Abdalla, A.E., Zhu, W., Xie, L.X., Wang, Y.L. & Guo, X.Q. 2020. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses* 12(2): 244.

Zhang, J., Zeng, H., Gu, J., Li, H., Zheng, L. & Zou, Q. 2020. Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines* 8(2): 153.

Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Peng, H., Quinlan, B.D., Rangarajan, E.S., Pan, A., Vanderheiden, A., Sulthar, M.S. & Li, W. 2020. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications* 11(1): 6013.

*Corresponding author; email: shaharum1@usm.my