Send Orders for Reprints to reprints@benthamscience.net

Current Bioinformatics, **xxxx,** *xx,* **x-x**

RESEARCH ARTICLE

Novel Design for Multi-Epitope Vaccines of COVID-19 and Critical *In-Silico* Assessment Steps

Tian Lan[1,3,4,5], Shuquan Su[1,3], Pengyao Ping[1,3] and Jinyan Li[2,6,*]

[1]*School of Computer Science, FEIT, University of Technology Sydney, Ultimo NSW 2007, Aus tralia;* [2]*Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China;* [3]*Data Science Institute, FEIT, University of Technology Sydney, Ultimo, NSW 2007, Australia;* [4]*Natexl, Ultimo, NSW 2007, Australia;* [5]*School of Medical Sciences, FMH, The University of Sydney, Sydney, NSW 2006, Australia;* [6]*Shenzhen University of Advanced Technology, Shenzhen 518055, China*

ARTICLE HISTORY

Received: January 10, 2024
Revised: May 27, 2024
Accepted: June 25, 2024

DOI:
10.2174/0115748936303461240827061629

Abstract: ***Introduction:*** The coronavirus disease COVID-19, caused by the SARS-CoV-2 virus, was a global pandemic that happened in March of 2020. The virus was mutated into several widely-spread strains such as Alpha, Beta, Gamma, Delta, and Omicron, and is continuing its unpredictable mutation.

Method: Multi-Epitope Vaccine (MEV) is one type of recombinant vaccine with its sequence containing multiple epitopes and is considered as an effective way to fight against the infectious disease. Previous in-silico approaches to MEV construction have been constrained by their inability to predict molecular conformation structures accurately, consequently leading to inaccurate property evaluations. In this work, we designed a novel MEV for the future prevention of COVID-19 or similar diseases. We set strict thresholds to screen for epitope candidates in order to construct highly effective MEV and use the latest ColabFold (a modified version of AlphaFold2) to predict accurate tertiary structures of the MEV.

Results: We especially studied epitopes from the main proteins of SARS-CoV-2 (*i.e*., the envelope, membrane, nucleo-, and spike proteins) that can provoke immunity response of B-cells, helper T-cells (Th), and cytotoxic T-cells (CTL), then we combined them through amino acid linkers to construct the MEV. We evaluated the vaccine in terms of its physicochemical properties, population coverage, safety for use, secondary and tertiary structure, docking immunity response, and immu nity response eliciting capability.

Conclusion: These *in-silico* assessments demonstrate that our proposed vaccine can elicit effective immune responses and it is safe to use with a high population coverage.

Keywords: Multi-epitope, vaccines, of COVID-19, and critical, in-silico, helper T-cells.

1. INTRODUCTION

COVID-19, caused by the infection of severe acute respiratory syndrome coronavirus 2, or SARS-CoV-2, was declared a global pandemic in March 2020 by the World Health Organization (WHO). The SARS-CoV-2 virus, from the Coronaviridae family, is a Baltimore class IV virus with an enveloped positive-sense single-stranded RNA genome ((+)ssRNA), homologous to other highly infectious virus, *e.g*., SARS-CoV-1 and MERS-CoV [1-3]. SARS-CoV-2 has a 29,881bp- length reference genome and a whole 9860-amino-acid protein sequence length. There are, in total, 4 structural viral proteins, namely the envelope, membrane, nucleo-, and spike proteins. Similar to SARS-CoV-1, the binding interactions between the peripheral Spike protein (S) of SARS-CoV-2 and the trans-membrane Angiotensin-converting enzyme 2 (ACE2) protein of human host cells are generally understood as the major mechanism allowing the viral genome enter the host cells [4, 5]. In fact, the Spike-ACE2 binding leads to the formation of endocytic vesicles with Transmembrane protease serine 2 (TMPRSS2), cleaving the S1 domain of Spike protein; Virions enter host cells *via* an endocytic pathway, and they replicate intracellularly in a rapid pace eventually leading to pathogenic infections[6-8]. After being released into the host cells, the viral RNA replicates itself to synthesize virus proteins [6-8]. SARS-CoV-2 also has immune evasion capacity through different mechanisms, *e.g*., suppressing the type I and III interferons (IFNs) signaling pathway by various viral Non-Structural Proteins (NSPs) and other encoding Open Reading Frames (ORFs) [9, 10]. Due to these characteristics, the virus has

*Address correspondence to this author at the Shenzhen University of Advanced Technology, Shenzhen 518055, China; E-mail: jinyan.li@siat.ac.cn

1574-8936/xx $65.00+.00

spread widely, and this has been demonstrated by the recently developed computational model [11, 12].

In the fight against the desperate spread of the COVID-19 pandemic, governments, scientists, researchers, and vaccine companies spared no effort in developing various categories of vaccines against SARS-CoV-2 [13, 14], where live attenuated vaccines [15, 16], adenovirus vaccines [17, 18], mRNA vaccines [19-21], and recombinant protein vaccines [22] are most widely used during the pandemic. Although the composition of different vaccine categories may be different, they basically share similar functionality related to immune-cell epitopes. That is, they all target immune-cell epitopes to elicit an immune response. This is achieved by presenting antigenic components to the immune system, which then recognizes and mounts a defense against the virus [23-28]. For example, live attenuated vaccines use a weakened form of the virus to stimulate an immune response without causing disease. Adenovirus vaccines leverage a modified virus to deliver the genetic material into cells, prompting them to produce viral proteins and stimulate immunity. mRNA vaccines use a piece of the virus's genetic code to instruct cells to produce a protein that triggers an immune response. Recombinant protein vaccines utilize engineered viral proteins to stimulate the immune system directly.

However, the virus genome sequences are not immutable, and contrarily, they have a significantly high mutation rate, mostly caused by encoded polymerase with low fidelity [29-31]. Thus, viruses of- ten exist inside hosts as quasi-species, which is a population of viruses within a single infected host with genetic diversity [32]. Moreover, infectious viruses have always been evolving under selective pressure from the host immune responses [33, 34]. Thus, recombinant protein vaccines containing limited epitope species are questioned regarding their potential deficiency in facing virus quasi-species and virus diversity evolution, which is not perfect as an ideal vaccine design, especially in an urgent situation such as a pandemic due to potential antibody evasion[35]. This brings the idea of multi-epitope vaccines, or MEV, to overcome this difficulty.

MEV is a type of recombinant vaccine with its sequence containing multiple epitopes, which are short peptides from specific antigens for the host's immune system to recognize and elicit immune response such as antibody production. These epitope sequences in the MEV are linked by specifically designed amino acid linkers. An MEV initiates the immune response by merging the epitopes, which can be simultaneously identified by B-cell, helper T-cell (Th), and cytotoxic T-cell (CTL)[36]. Compared to traditional vaccines, MEV has been found to be a promising way to fight against viral infections [37] for several reasons, such as no need for microbial culture, its lower development cost, safer usage (no real pathogen required), and stronger immunogenicity [38-41].

Recent research on MEVs has shown promising developments, revealing their potential to elicit robust immune responses by targeting multiple regions of a pathogen [42-44.] These vaccines offer enhanced efficacy and broader protection compared to traditional vaccines, as they are designed to efficiently induce immunity, which is crucial for long-term protection against viruses [45, 46]. Studies have demonstrated the effectiveness of MEVs in generating comprehensive immune responses and addressing a wide range of viral threats. Additionally, significant advancements in bioinformatics tools have allowed for the precise prediction and selection of epitopes that ensure high population coverage while minimizing potential adverse reactions [37, 47]. However, a notable challenge remains for those *in silico* frameworks icluding the inability to accurately predict the conformation of the constructed MEV, which has led to inaccuracies in the *in-silico* evaluation of these vaccines. This issue highlights the need for continued research and refinement in the design and evaluation processes to fully realize the potential of MEVs in combating infectious diseases.

In this work, we construct a Multi-Epitope Vaccine (MEV) using strictly screened epitopes. These epitopes are selected based on various metrics, including percentile rank, conservation, anti-genicity, ability to induce interferon gamma, allergenicity, and toxicity. Our approach employs *in-silico* immunoinformatics methods to ensure the vaccine's effectiveness, safety, broad population coverage, and immunogenicity. We further use the latest ColabFold, which is a modified AlphaFold2-Multimer algorithm interrogating the MMseqs2 dataset [48] to predict the MEV's ac- curate tertiary structure, supported by the high ERRAT score. On top of that, the properties of the MEV are comprehensively evaluated. *In silico* simulation results confirm the effectiveness of the developed vaccine, revealing its remarkable ability to provoke long-lasting immunity through the activation and proliferation of both T and B cells while also triggering a vigorous production of antibodies.

2. MATERIALS AND METHODS

The flowchart in Fig. (**1**) describes a pipeline of steps for the construction and evaluation of our MEV. We began with the prediction of the antigenicity of SARS-CoV-2's Envelope (E), Membrane (M), Nucleo- (N), and Spike (S) proteins. We then predicted the associated MHC-I, MHC-II, and B-cell epitopes for each of the E, M, N, and S virus proteins. We then evaluated and score-ranked these predicted epitopes in terms of their antigenicity, allergenicity, toxicity, and interferon gamma-inducing capability. The MEV was then constructed using the top-ranked epitopes, and its physio-chemical properties were evaluated. We then predicted, refined, and validated the construct's conformation structure. Docking of the construct was then performed with the TLR3 and HLA alleles. The immune response of the MEV construct was also simulated for the final evaluation of the effect.

2.1. Epitope Selection

Four main COVID-19 virus protein sequences, including envelope, membrane, nucleo-, and spike proteins, wereretrieved from the Uniprot Swiss database under the accession numbers P0DTC4, P0DTC5, P0DTC9, and

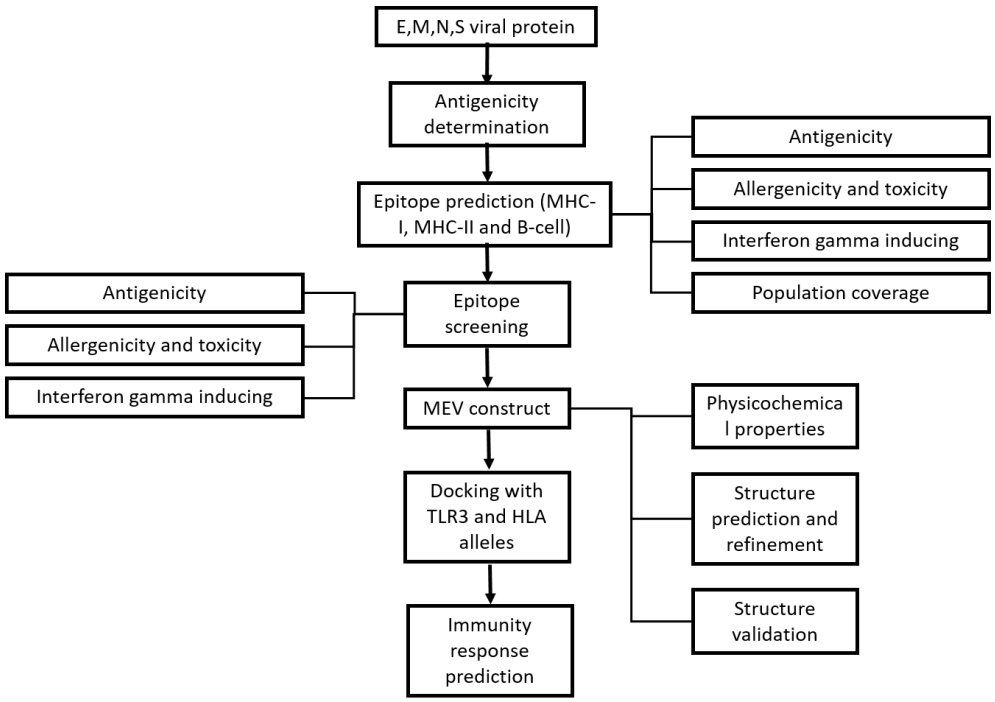


Fig. (1). Flowchart of MEV design and evaluation.

P0DTC2. Antigenicity were detected by the Vaxigen 2 server under the virus model. B-cell and T-cell epitopes werepredicted separately based on the virus's envelope, membrane, nucleo-, and spike proteins. B-cell epitopes werepredicted using ABCpred for 16- mer sequences, where the threshold was set at 0.5. T-cell epitopes werepredicted using IEDB MHC-i and MHC-ii separately. Both predictions used the reference allele from the IEDB prediction tool. For each different MHC allele type and different peptide length, we filtered out the predicted epitopes for MHC-i (at percentile ratio > 0.01) and for MHC-ii (at percentile ratio > 0.1). Conservatory analysis was performed using the IEDB conservatory analysis tool for all T-cell epitopes. Epitopes with all 100 percent conservatories werereserved.

Antigenicity for the remaining epitopes was determined using vaxigen 2. For CD4 cell epitopes, the capability of inducing the Interferon-gamma waspredicted *via* IFNepitope, with the motif and SVM hybrid model. Epitopes determined to be able to induce interferon were reserved. Population coverage for MHC-I and MHC-II epitopes wasanalyzed using the IEDB Population Coverage tool. Toxins werepredicted *via* Toxinpred, and allergenicity was predicted by Allercatpro. Only epitopes that weredetermined as non-toxic and nonallergenic were kept for analysis afterward.

2.2. Vaccine Construction, Structure Prediction, Refinement, and Evaluation

2.2.1. Vaccine Construction and Properties

The epitopes were chosen among the remaining epitopes sat isfying that they must have a low percentile rank, a wide coverage of different immune cells, and from all four virus proteins. As an adjuvant, beta-defensin-3 was added at the start of the vaccine after the EAAAK linker to boost immune responses and enhance the effectiveness. AYY, GPGPG, and KK linkers wereused to join within MHC-I epitopes, MHC-II epitopes, and B-cell epitopes respectively. A 6xHis-tag tail wasadded at the end of the construct. The links wereadded in order to genera,te a sequence with minimized junctional immunogenicity. The physicochemical properties of the construct was assayed *via* ProtParam at Expasy.

2.2.2. Interferon-γ Induction and Allergenicity Prediction

The interferon-*γ* inducing capability of the vaccine construct was predicted by IFNepitope, where the length was set as 9-mer amino acids scanning over the entire construct and SVM was chosen as the prediction method. The AllerCatPro server was used to predict the allergenicity of the vaccine.

2.2.3. Structure Prediction, Refinement, and Evaluation

The secondary structure of the vaccine was predicted *via* PSIPRED 4.0, while Colabfold v1.5.2 was used to predict the tertiary structure. MMseqs2 model (UniRef+ Environmental) was used for multiple sequence alignment (MSA), and 5 models were set to generate the result. The model with the highest IDDT was selected. GalaxyRefine2 performed structure refinement on the predicted tertiary structure from Colabfold, where all parameters were set by default.

Prosa, ERRAT, and PROCHECK were then used to examine the quality of the predicted tertiary protein structure. All parameters were set as default.

2.2.4. Molecular Docking of Vaccine Construct

Molecular docking of the vaccine construct was performed *via* HDOCK. Chain A of the Protein Data Bank

(PDB) entity 1ZIW was chosen as a receptor model of the human toll-like receptor 3(TLR3). For docking with the HLA allele, 2 HLA alleles (HLA-A*32:01 and HLA-DRB*07:01) were input as the model of the receptors. The protein sequencing data were obtained from an allele search tool under the ID 6at5.1.A and 3pdo.1.B respectively. SWISS-MODEL was used to model their tertiary structures. The docking models for each of the three docking constructs (TLR3, HLA-A, HLA-DRB) with the lowest docking score were set as the final models.

2.2.5. Codon Adaption and In-silico Cloning

Codon adaption analysis was conducted by the JCAT server. JCAT took the sequence of the vaccine construct as input data, where the host organism was set as e-coli k12, and the restriction enzymes wereset as BamHI and EcoRI. The *in-silico* cloning was performed by the Snapgene software, where Plasmid pET-28(+) was used. Sequences GAATTC and GGATCC, which are the cleavage sites of the restriction enzymes EcoRI and BamHI, were added at the start and end of the vaccine construct, respectively. The same restriction enzymes were used to cut plasmid to integrate the sequences of the vaccine at positions 192-1078.

2.2.6. Immunity Simulation

The immune response of the vaccine was simulated by the C-immsim server. The total number of simulation steps wasset as 1050, with each step being 8 hours (350 days). Three injections occured at steps 1, 84, and 168 (0 days, 28 days, and 56 days). The other parameters wereset by default.

3. RESULTS AND DISCUSSION

3.1. Identification of Various Categories of Epitopes

3.3.1. Antigenicity test for the Four Main Proteins of SARS-CoV-2

The amino acid sequences of the four proteins of SARS-CoV-2 (the envelop, membrane, nucleoprotein, and spike protein) were tested to understand whether they have high levels of antigenicity as antigens. The test was carried out using Vaxigen 2 [49]. The corresponding overall protective antigen prediction scores were3cm025, 0.5102, 0.5059, and 0.4646, respectively. The threshold was set as 0.4 for the virus model.

3.3.2. Prediction of HLA-I, HLA-II, and B-cell Epitopes

We predicted three types of epitope sites at each of the antigen sequences of SARS-CoV-2. We used IEDB mhc-i [50, 51] to predict T-cell HLA-I epitopes and mhc-ii [52-54] to predict T-cell HLA-II epitopes. The B-cell epitope sites were predicted through ABCpred [55]. We identified a total of 98 HLA-I and HLA-II epitopes and a total of 203 B-cell epitopes, which all have a high percentile rank. The selection criteria are detailed in the Methods section.

3.3.3. Epitope Filtration and Docking Analysis

All those epitopes were found to be 100% conserved by the IEDB conservatory analysis tool [56] and determined as antigens *via* the Vaxigen 2 server, and all those CD4 cell epitopes were determined to have the capability of inducing the Interferon γ by IFNepi- tope server, were remained for further consideration to construct the vaccine. In total, 56 T-cell epitopes (13 MHC-I epitopes and 43 MHC-I epitopes) and 13 B-cell epitopes were held in this consideration.

To find out the interaction level between predicted epitopes and the receptor (HLA-a, HLA-drb, and TLR3), we performed docking analysis on two selected epitopes (NVSLVKPS- FYYVSRVK, and RVKNLNSSR). Fig. (**2**) shows that epitope residues fit well with HLA-a, HLA-drb, and TLR3 receptors, which are also shown by the low docking score calculated by the hdock server (-221.83, -180.42, -243.25, -144.00, -192.76, -141.54 for NVSLVKPSFYYVSRVK and RVKNLNSSR docking to HLA-a, HLA-drab, and TLR3, respectively). It is noted that NVSLVKPS- FYYVSRVK shows a slightly stronger fitting level with HLA-a, HLA-drab, and TLR3 compared to RVKNLNSSR. We concluded the reason was that its longer sequence leads to more surface exposure chances to the receptors.

3.3.4. Population coverage Analysis

We conducted the population coverage analysis on the remaining MHC-I and MHC-II epitopes through an IEDB population coverage tool [57][in order to identify those epitopes that are able to cover wide populations in the world. It has been found that these MHC-I and MHC-II epitopes have respectively 86.53% and 92.92% coverage of the world population. Their average hit values (*i.e.*, the average numbers of epitope hits / HLA combinations recognized by the population) are 2.51 and 6.6 for the MHC-I and MHC-II epitopes, respectively. Their PC90 values (*i.e.*, the minimum numbers of epitope hits / HLA combinations recognized by 90% of the population) are 0.74 and 3.63, respectively (Fig. **3**).

3.3.5. Allergenicity and Toxicity Prediction

Allergenicity and toxicity of these epitopes were predicted by Allercatpro [58] and Toxinpred [59]. All epitopes were found to be of non-allergenicity and non-toxicity.

3.4. Vaccine Construction and Structure Prediction

3.4.1. Vaccine Construct Design

In our design, the MEV was required to include epitopes from each of the virus proteins (envelope, membrane, nucleoprotein, spike), and each of the antigens should contribute epitopes corresponding to different types of immune cells (B-cell, HLA-I cell and HLA- II cell). We excluded those epitopes with relatively high percentile rank in each epitope category. Based on these selection criteria, the vaccine was designed as demonstrated in Fig. (**4a**). This vaccine contains 293 amino acids, consisting of beta-defensin-3 as adjuvants, 7 HLA-I epitopes joined by AYY linkers, 4 HLA-II epitopes joined by GPGPG linkers, 4 B-cell epitopes joined by KK linkers, an EAAAK linker at the beginning of the sequence and a 6xHis-tag tail at the end. Among these epitopes, three

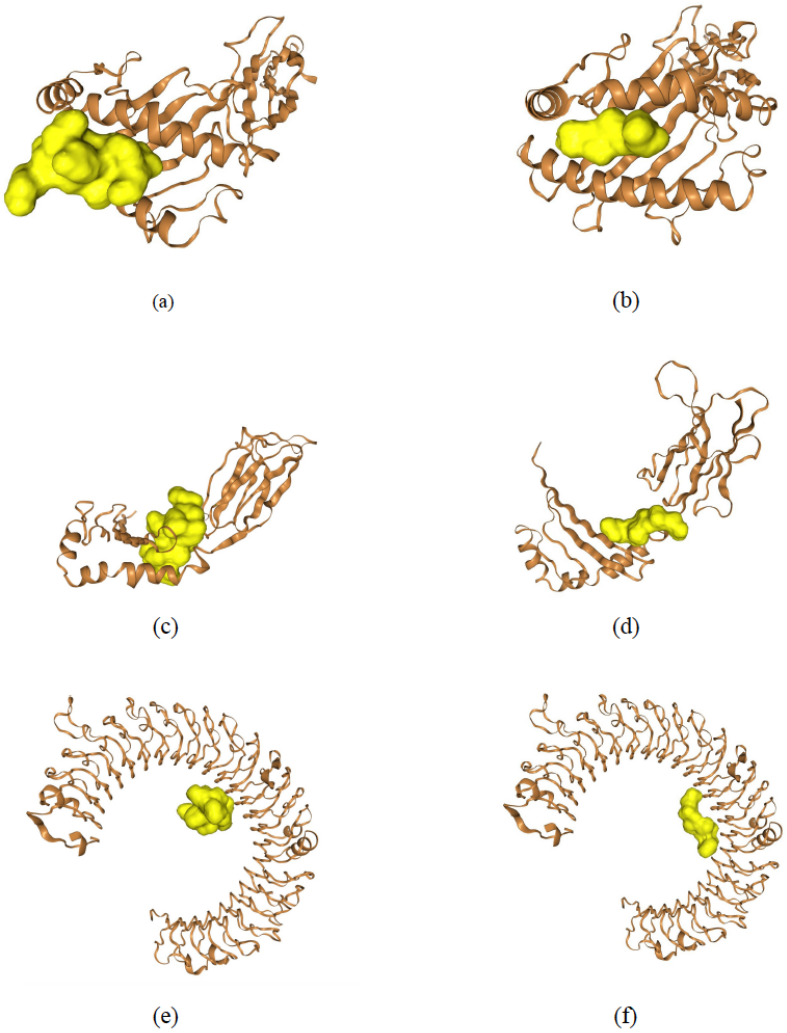
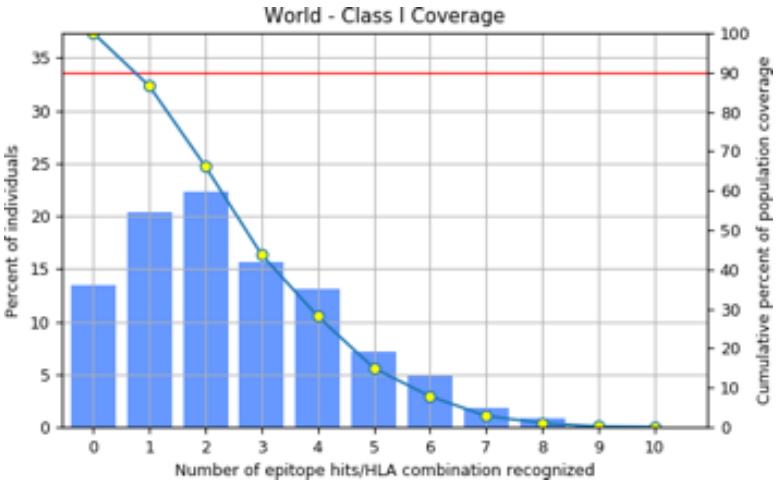
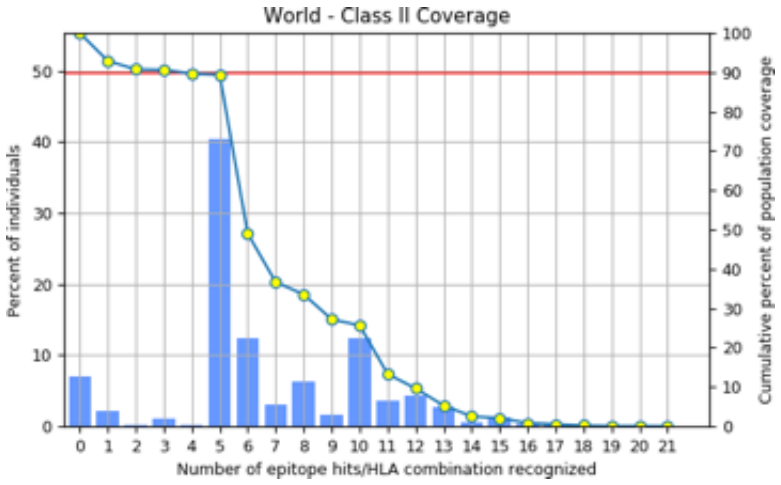


Fig. (2). Selected epitopes docking analysis. **(a)** NVSLVKPSFYYVSRVK in complex with HLA- a **(b)** RVKNLNSSR in complex with HLA-a. **(c)** NVSLVKPSFYYVSRVK in complex with HLA- drb MHC-I epitope. **(d)** RVKNLNSSR in complex with HLA-drb. **(e)** NVSLVKPSFYYVSRVK in complex with TLR3 MHC-I epitope. **(f)** RVKNLNSSR in complex with TLR3. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).



(a)

(Fig. 3) Contd….



(b)

Fig. (3). Population coverage analysis. **(a)** MHC-I epitope population coverage. **(b)** MHC-II epitope population coverage. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

are from the envelop protein, two from the membrane protein, one from the nucleo-protein protein, and 9 from the spike protein. The linkers serve the purpose of maintaining a stable epitope structure and avoiding the formation of novel epitopes.

3.4.2. Vaccine Properties and Secondary Structure Prediction

We used the Expasy server [60] to examine physicochemical properties of the vaccine construct. The chemical formula of our construct was $C1478H2298N396O383S10$, with 293 amino acids having a molecular weight 32063.53KDa. The score of 92.18 aliphatic index implies that the construct is thermally stable, and its instability index

27.16 again indicates that the construct is a stable protein. A grand average of hydropathicity (GRAVY) score of 0.058 classifies the construct as hydrophobicity.

The IFNepitope server [61] determined a 47.5% positive interferon epitope ratio of the con- struct, and the AllerCatPro server found the construct to be non-allergenic.

The secondary structure of the vaccine construct was predicted by the PSIPRED server [62], as displayed in Fig. (**4b**), where the alpha helix and beta sheets are in different colours.

3.4.3. Tertiary Structure Prediction and Refinement

Colabfold [48, 63] was used to predict a tertiary structure for the vaccine construct. The model with the highest local distance difference test (IDDT) score was selected. The refined model 3 from the GalaxyRefine2 server [64] was chosen, which has a Root-Mean-Square Deviation (RMSD) of 3.574, molProbity of 1.099, clash score of 1, poor rotamers of 0, Rama favored score of 95.5 and a GALAXY energy of -5193.01, indicating a relatively accurate prediction structure (see Fig. **4c**).

3.4.4. Structure Validation for the MEV

The predicted structure was tested by Prosa [65], ERRAT [66], and PROCHECK [67]. Prosa returned a z-score of -2.36 (see Supplementary Figs. **1-5**), which is in the range of the scores for a protein of such size (around 300 amino acids). The ERRAT server returned an overall quality score of 95 (see Supplementary Fig. **4**), which is a high score indicating the structure with non-bonded interactions between different atom types. PROCHECK returned no errors for the vaccine construct, and some of the validation results are shown imSupplementary Figs. (**1-5**); Supplementary Figs. (**1-3**) display a Ramachandran plot and Chi1-Chi2 plot for different types of the residuals and a plot summarizing the statistic parameters of the main chain structure, respectively. Most residues are found at favourable conformations in terms of Psi-Phi and Chi1-Chi2 torsion angles, and all of the statistics for the main chain of the construct are within a reasonable range, indicating the high quality of the tertiary vaccine synthesis.

3.4.5. Molecular Docking Tests for the Vaccine Construct

The docking of the vaccine construct was performed by the HDOCK server [68]. A Human toll-like receptor 3 (TLR3), one HLA-A allele (HLA-A*32:01), and one HLA-DRB allele (HLA-DRB*07:01) were set to evaluate the vaccine's immunogenicity. All three docking complexes have a low docking score of -333.16,

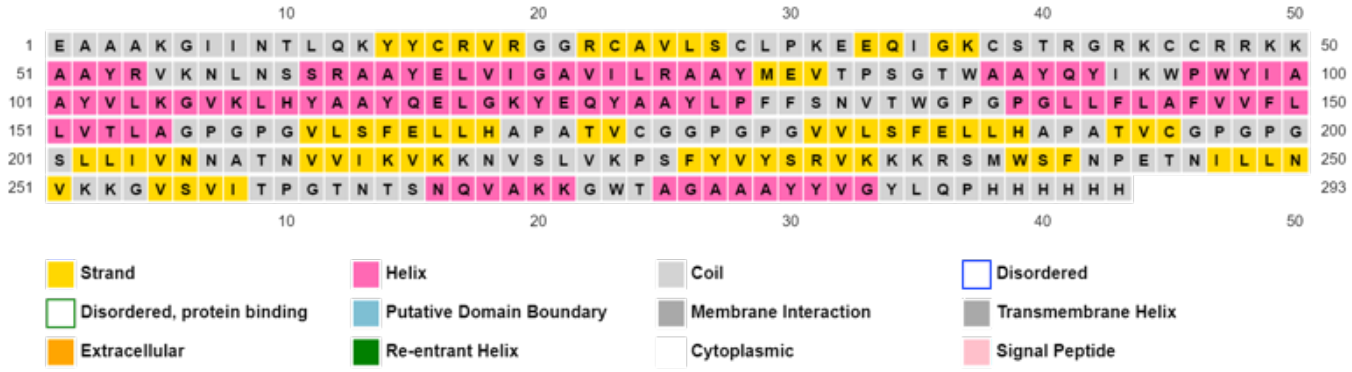
-289.97 or -349.13 corresponding to the model TLR3-vaccine, HLA-A-vaccine, or HLA-DRB- vaccine calculated by the HDOCK server, respectively, indicating stable bindings between the ligand and the receptor (Fig. **5**).

3.4.6. Codon Adaption and In-silico Cloning

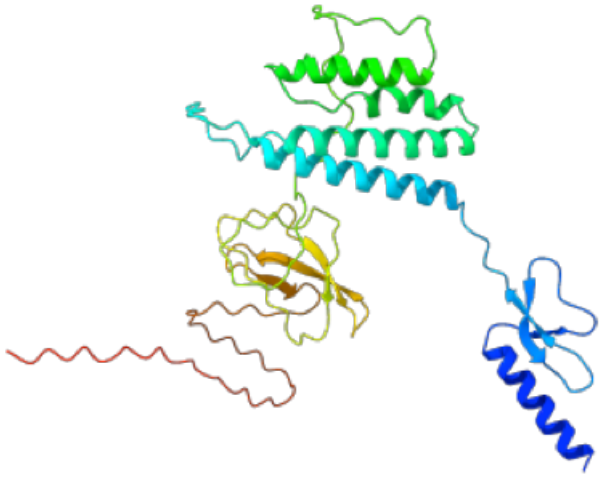
Gene expression of a vaccine is largely affected by the codon sequence and composition. Appropriate codon

EAAAKGIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKAAYR
VKNLNSSRAAYELVIGAVILRAAYMEVTPSGTWAAYQYIKWPWYIAAYVLKG
VKLHYAAYQELGKYEQYAAYLPFFSNVTWGPGPGLLFLAFVVFLLVTLAGPGP
GVLSFELLHAPATVCGGPGPGVVLSFELLHAPATVCGPGPGSLLIVNNATNVVI
KVKKNVSLVKPSFYVYSRVKKKRSMWSFNPETNILLNVKKGVSVITPGTNTSN
QVAKKGWTAGAAAYYVGYLQPHHHHHH

(a)

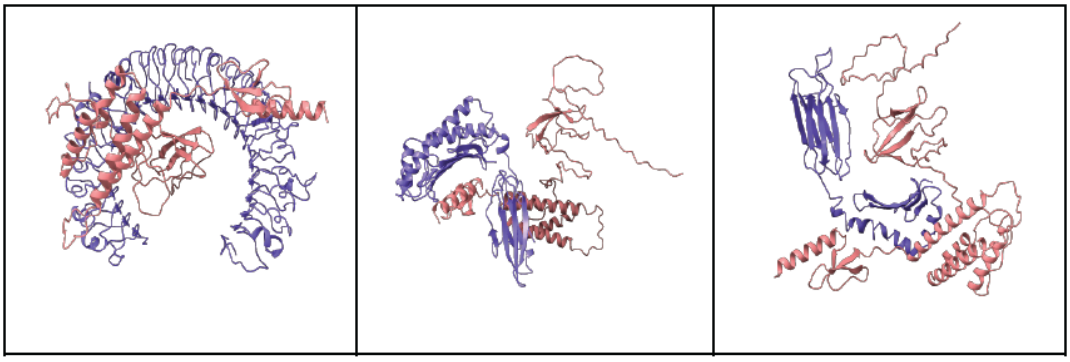


(b)



(c)

Fig. (4). Multi-epitope vaccine construct **(a)** Vaccine sequence where the colors gray, yellow, brown, and black represent adjuvant, HLA-I epitope, HLA-II epitope, and B-cell epitope respectively. The colour red represents any linker between two epitopes. **(b)** Predicted secondary structure of our vaccine construct. **(c)** Predicted tertiary structure. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

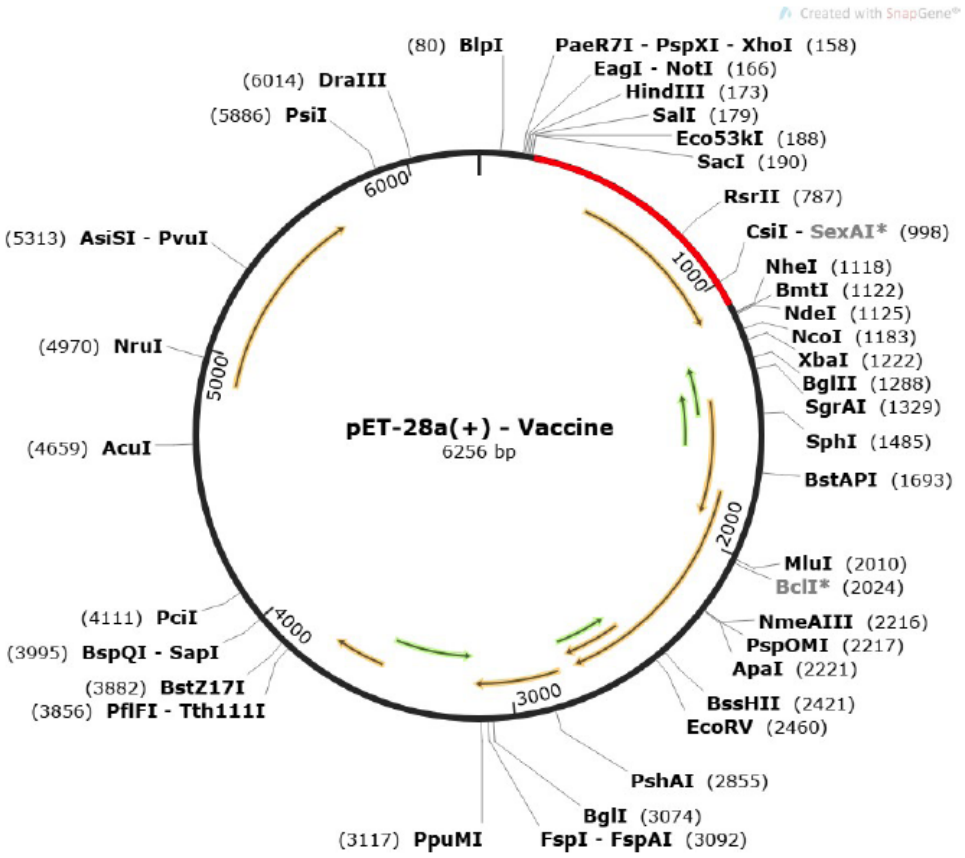


(a) (b) (c)

Fig. (5). Vaccine construct docking with TLR3 **(a)**, HLA-A **(b)** and HLA-DRB **(c)**, where the vaccine as the ligand is coloured purple and the receptor is coloured red. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

```
GAAGCTGCTGCTAAAGGTATCATCAACACCCTGCAGAAATACTACTGCCG     50
TGTTCGTGGTGGTCGTTGCGCTGTTCTGTCTTGCCTGCCGAAAGAAGAAC    100
AGATCGGTAAATGCTCTACCCGTGGTCGTAAATGCTGCCGTCGTAAAAAA    150
GCTGCTTACCGTGTTAAAAACCTGAACTCTTCTCGTGCTGCTTACGAACT    200
GGTTATCGGTGCTGTTATCCTGCGTGCTGCTTACATGGAAGTTACCCCGT    250
CTGGTACCTGGGCTGCTTACCAGTACATCAAATGGCCGTGGTACATCGCT    300
GCTTACGTTCTGAAAGGTGTTAAACTGCACTACGCTGCTTACCAGGAACT    350
GGGTAAATACGAACAGTACGCTGCTTACCTGCCGTTCTTCTCTAACGTTA    400
CCTGGGGTCCGGGTCCGGGTCTGCTGTTCCTGGCTTTCGTTGTTTTCCTG    450
CTGGTTACCCTGGCTGGTCCGGGTCCGGGTGTTCTGTCTTTCGAACTGCT    500
GCACGCTCCGGCTACCGTTTGCGGTGGTCCGGGTCCGGGTGTTGTTCTGT    550
CTTTCGAACTGCTGCACGCTCCGGCTACCGTTTGCGGTCCGGGTCCGGGT    600
TCTCTGCTGATCGTTAACAACGCTACCAACGTTGTTATCAAAGTTAAAAA    650
AAACGTTTCTCTGGTTAAACCGTCTTTCTACGTTTACTCTCGTGTTAAAA    700
AAAAACGTTCTATGTGGTCTTTCAACCCGGAAACCAACATCCTGCTGAAC    750
GTTAAAAAAGGTGTTTCTGTTATCACCCCGGGTACCAACACCTCTAACCA    800
GGTTGCTAAAAAAGGTTGGACCGCTGGTGCTGCTGCTTACTACGTTGGTT    850
ACCTGCAGCCGCACCACCACCACCACCAC
```

(a)



(b)

Fig. (6). Codon adaption and in-silico cloning for our vaccine construct. **(a)** Codon optimization result. **(b)** In-silico cloning result (the vaccine construct is coloured red). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

orientation may result in a high level of gene expression compared to others, although they produce the same peptide. The JCAT server [69] helped perform codon adaption for the vaccine construct in Escherichia coli (strain k12), where the cleavage sites of restriction enzymes BamHI and EcoRI wereavoided. The result of our MEV is shown in Fig. (**6**). The total codon of our vaccine was 879-base with a CAI-Value 1.0. It indicated a high level of expression capability. The GC-Content of the codon was 50.967, which is similar to that of Escherichia coli strain K12 (50.734). The optimized codon was then inserted into the pET-28(+) plasmid *via* the Snapgene software (shown in red color between the sites EcoRI and BamHI.1).

3.4.7. Immunity Simulation

The immunity prediction results are shown in Fig. (**7**) as derived by the C- immsim server [70]. The x-axis represents the time of injection, and the y-axis stands for the amount of

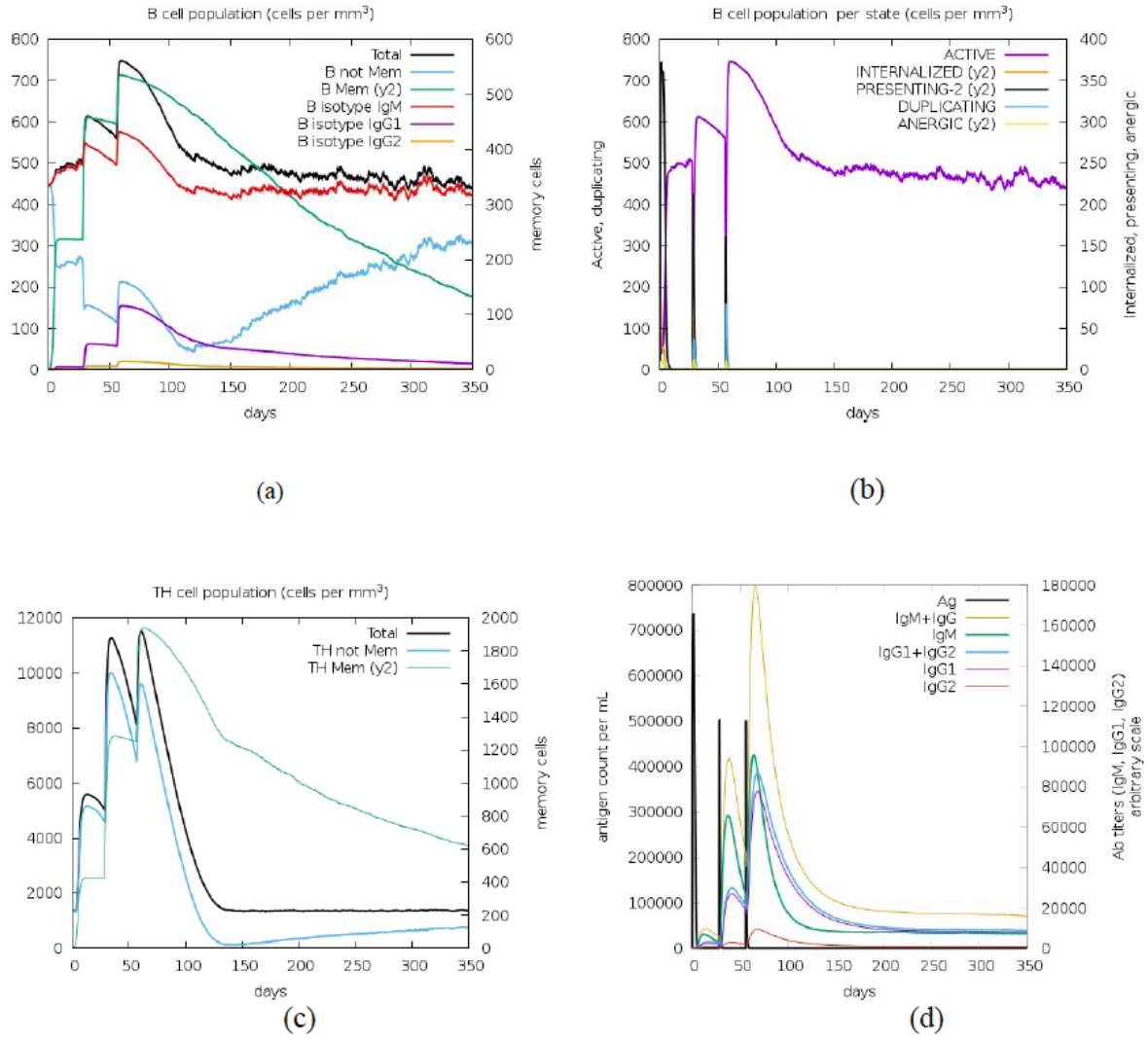


Fig. (7). Simulation for vaccine immunity response. **(a)** B-cell population plot. **(b)** B-cell popu- lation plot in terms of cell states. **(c)** Helper T-cell population plot. **(d)** Antibody type plot. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

the specific cell/antibody/antigen types. Immune cells (T and B cells) and the antibody wereinduced and enriched sharply at the injection time of the vaccine. Active B-cells maintained at a high level over the 360 days after the injection. Antibodies (IgG1, IgG2, and IgGM) wereinduced and remained in the body after the injection, with IgG1 generated at a very high level after the third dose. The findings establish the efficacy of the formulated vaccine, elucidating its remarkable capacity to induce enduring immunity by stimulating the activation and proliferation of both T and B cells, in addition to eliciting a robust production of antibodies.

CONCLUSION

We designed a novel multi-epitope vaccine according to the sequences from the four main proteins of the SARS-CoV-2 virus (envelope, membrane, nucleo-, and spike protein). We examined some key properties of the four proteins to ensure they all are antigens. Then, we predicted their T- and B-cell epitopes. We filtered the epitopes based on percentile rank, conservatory, antigenic- ity, interferon-γ inducing capability, allergenicity, and toxicity so that the remaining epitopes are relatively effective for stimulating immune response as well as safe for use. We also performed a population coverage analysis to ensure wide coverage of our designed vaccine. To understand more about the vaccine properties, secondary and tertiary structures were predicted *via* PSIPRED and ColabFold. The structures were validated by Prosa, ERRAT and PROCHECK. Molecular docking was then performed for the vaccine with human TLR3 and HlA alleles. The stable docking construct indicates its effectiveness. The corresponding gene, after codon optimization, was successfully cloned into the plasmid vector (pET-28a (+)). Then the C-immsim server was used for simulating the immune response of the vaccine, which revealed the capability of the MEV to elicit high-level primary, secondary, and tertiary immune responses. A limitation of this study is that all construction and analysis for the MEV are currently based

on in-silico re- sults. To achieve more accurate analyses, future wet-lab experiments regarding the properties of MEV may need to be conducted.

AUTHORS' CONTRIBUTION

It is hereby acknowledged that all authors have accepted responsibility for the manuscript's content and consented to its submission. They have meticulously reviewed all results and unanimously approved the final version of the manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

COVID-19 virus amino acid sequences (envelope, membrane, nucleo- and spike proteins) are retrieved from the Uniprot Swiss database under the accession numbers P0DTC4, P0DTC5, P0DTC9, and P0DTC2. Human toll-like receptor 3(TLR3) is retrieved from Protein Data Bank (PDB) entity 1ZIW. HLA alleles' (HLA-A*32:01 and HLA-DRB*07:01) amino acid sequences are obtained from the IPD-IMGT/HLA Allele Query Tool under the ID 6at5.1.A and 3pdo.1.B.

FUNDING

This work was supported by the Australia Research Council Discovery Project DP180100120, and the National Innovation Fellow Program of the MOST of China (J.L., grant no. E327130001).

CONFLICT OF INTERESTS

The authors declare that they have no competing financial interests.

ACKNOWLEDGMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material, along with the published article, is available on the publisher's website.

REFERENCES

[1] Lu R, Zhao X, Li J, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. Lancet 2020; 395(10224): 565-74. http://dx.doi.org/10.1016/S0140-6736(20)30251-8 PMID: 32007145

[2] Zhu N, Zhang D, Wang W, *et al.* A novel coronavirus from patients with pneumonia in china, 2019. N Engl J Med 2020; 382(8): 727-33. http://dx.doi.org/10.1056/NEJMoa2001017 PMID: 31978945

[3] Chan JFW, Kok KH, Zhu Z, *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect 2020; 9(1): 221-36. http://dx.doi.org/10.1080/22221751.2020.1719902 PMID: 31987001

[4] Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol 2021; 19(3): 141-54. http://dx.doi.org/10.1038/s41579-020-00459-7 PMID: 33024307

[5] Zipeto D, Palmeira JF, Argañaraz GA, Argañaraz ER. Ace2/adam17/tmprss2 interplay may be the main risk factor for covid-19. Front Immunol 2020; 11: 576745. http://dx.doi.org/10.3389/fimmu.2020.576745 PMID: 33117379

[6] Scudellari M. How the coronavirus infects cells — and why Delta is so dangerous. Nature 2021; 595(7869): 640-4. http://dx.doi.org/10.1038/d41586-021-02039-y PMID: 34321669

[7] Fehr AR, Perlman S. Coronaviruses: An overview of their replication and pathogenesis. Methods Mol Biol 2015; 1282: 1-23. http://dx.doi.org/10.1007/978-1-4939-2438-7_1 PMID: 25720466

[8] Huang Y, Yang C, Xu X, Xu W, Liu S. Structural and functional properties of SARS-CoV-2 spike protein: Potential antivirus drug development for COVID-19. Acta Pharmacol Sin 2020; 41(9): 1141-9. http://dx.doi.org/10.1038/s41401-020-0485-4 PMID: 32747721

[9] Rashid F, Xie Z, Suleman M, Shah A, Khan S, Luo S. Roles and functions of SARS-CoV-2 proteins in host immune evasion. Front Immunol 2022; 13: 940756. http://dx.doi.org/10.3389/fimmu.2022.940756 PMID: 36003396

[10] Xia H, Cao Z, Xie X, *et al.* Evasion of type i interferon by sars-cov-2. Cell Rep 2020; 33(1): 108234. http://dx.doi.org/10.1016/j.celrep.2020.108234 PMID: 32979938

[11] Tutsoy O, Tanrikulu MY. Priority and age specific vaccination algorithm for the pandemic diseases: A comprehensive parametric prediction model. BMC Med Inform Decis Mak 2022; 22(1): 4. http://dx.doi.org/10.1186/s12911-021-01720-6 PMID: 34991566

[12] Tutsoy O. Graph theory based large-scale machine learning with multi-dimensional con- strained optimization approaches for exact epidemiological modelling of pandemic diseases. IEEE Trans Pattern Anal Mach Intell 2023; 45(8): 9836-45. http://dx.doi.org/10.1109/TPAMI.2023.3256421 PMID: 37028303

[13] Matrajt L, Eaton J, Leung T, Brown ER. Vaccine optimization for COVID-19: Who to vaccinate first? Sci Adv 2021; 7(6): eabf1374. http://dx.doi.org/10.1126/sciadv.abf1374 PMID: 33536223

[14] Callaway E. Omicron likely to weaken COVID vaccine protection. Nature 2021; 600(7889): 367-8. http://dx.doi.org/10.1038/d41586-021-03672-3 PMID: 34880488

[15] Chen J, Wang P, Yuan L, *et al.* A live attenuated virus-based intranasal COVID-19 vaccine provides rapid, prolonged, and broad protection against SARS-CoV-2. Sci Bull (Beijing) 2022; 67(13): 1372-87. http://dx.doi.org/10.1016/j.scib.2022.05.018 PMID: 35637645

[16] Li M, Wang H, Tian L, *et al.* COVID-19 vaccine development: Milestones, lessons and prospects. Signal Transduct Target Ther 2022; 7(1): 146. http://dx.doi.org/10.1038/s41392-022-00996-y PMID: 35504917

[17] Hasanpourghadi M, Novikov M, Ertl HCJ. Covid-19 vaccines based on adenovirus vectors. Trends Biochem Sci 2021; 46(5): 429-30. http://dx.doi.org/10.1016/j.tibs.2021.03.002 PMID: 33810926

[18] Imai M, Iwatsuki-Horimoto K, Hatta M, *et al.* Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development. Proc Natl Acad Sci USA 2020; 117(28): 16587-95. http://dx.doi.org/10.1073/pnas.2009799117 PMID: 32571934

[19] Fang E, Liu X, Li M, *et al.* Advances in COVID-19 mRNA vaccine development. Signal Transduct Target Ther 2022; 7(1): 94. http://dx.doi.org/10.1038/s41392-022-00950-y PMID: 35322018

[20] Verbeke R, Lentacker I, De Smedt SC, Dewitte H. The dawn of mRNA vaccines: The COVID-19 case. J Control Release 2021; 333: 511-20.

http://dx.doi.org/10.1016/j.jconrel.2021.03.043 PMID: 33798667
[21] Szabó GT, Mahiny AJ, Vlatkovic I. COVID-19 mRNA vaccines: Platforms and current developments. Mol Ther 2022; 30(5): 1850-68. http://dx.doi.org/10.1016/j.ymthe.2022.02.016 PMID: 35189345
[22] Yadav T, Srivastava N, Mishra G, *et al.* Recombinant vaccines for COVID-19. Hum Vaccin Immunother 2020; 16(12): 2905-12. http://dx.doi.org/10.1080/21645515.2020.1820808 PMID: 33232211
[23] Gebre MS, Brito LA, Tostanoski LH, Edwards DK, Carfi A, Barouch DH. Novel approaches for vaccine development. Cell 2021; 184(6): 1589-603. http://dx.doi.org/10.1016/j.cell.2021.02.030 PMID: 33740454
[24] Brisse M, Vrba SM, Kirk N, Liang Y, Ly H. Emerging concepts and technologies in vaccine development. Front Immunol 2020; 11: 583077. http://dx.doi.org/10.3389/fimmu.2020.583077 PMID: 33101309
[25] Clem A. Fundamentals of vaccine immunology. J Glob Infect Dis 2011; 3(1): 73-8. http://dx.doi.org/10.4103/0974-777X.77299 PMID: 21572612
[26] Ulmer JB, Valley U, Rappuoli R. Vaccine manufacturing: Challenges and solutions. Nat Biotechnol 2006; 24(11): 1377-83. http://dx.doi.org/10.1038/nbt1261 PMID: 17093488
[27] Excler JL, Saville M, Berkley S, Kim JH. Vaccine development for emerging infectious diseases. Nat Med 2021; 27(4): 591-600. http://dx.doi.org/10.1038/s41591-021-01301-0 PMID: 33846611
[28] Singh M, O'Hagan D. Advances in vaccine adjuvants. Nat Biotechnol 1999; 17(11): 1075-81. http://dx.doi.org/10.1038/15058 PMID: 10545912
[29] Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cell Mol Life Sci 2016; 73(23): 4433-48. http://dx.doi.org/10.1007/s00018-016-2299-6 PMID: 27392606
[30] Domingo E, García-Crespo C, Lobo-Vega R, Perales C. Mutation rates, mutation fre- quencies, and proofreading-repair activities in RNA virus genetics. Viruses 2021; 13(9): 1882. http://dx.doi.org/10.3390/v13091882 PMID: 34578463
[31] Cai X, Lan T, Ping P, Oliver B, Li J. Intra-host co-existing strains of sars-cov-2 reference genome uncovered by exhaustive computational search. Viruses 2023; 15(5): 1065. http://dx.doi.org/10.3390/v15051065 PMID: 37243151
[32] Tay JH, Porter AF, Wirth W, Duchene S. The emergence of sars-cov-2 variants of concern is driven by acceleration of the substitution rate. Mol Biol Evol 2022; 39(2): msac013. http://dx.doi.org/10.1093/molbev/msac013 PMID: 35038741
[33] Koonin EV, Dolja VV, Krupovic M. The logic of virus evolution. Cell Host Microbe 2022; 30(7): 917-29. http://dx.doi.org/10.1016/j.chom.2022.06.008 PMID: 35834963
[34] Lan T, Su S, Ping P, *et al.* Generating mutants of monotone affinity towards stronger protein complexes through adversarial learning. Nat Mach Intell 2024; 6(3): 315-25. http://dx.doi.org/10.1038/s42256-024-00803-z
[35] Kimura I. Virological characteristics of the SARS-CoV-2 Omicron BA.2 subvariants, including BA.4 and BA.5. Cell 2022; 185(21): 3992-4007.e16.
[36] Negahdaripour M, Golkar N, Hajighahramani N, Kianpour S, Nezafat N, Ghasemi Y. Harnessing self-assembled peptide nanoparticles in epitope vaccine design. Biotechnol Adv 2017; 35(5): 575-96. http://dx.doi.org/10.1016/j.biotechadv.2017.05.002 PMID: 28522213
[37] Zhang L. Multi-epitope vaccines: A promising strategy against tumors and viral infections. Cell Mol Immunol 2018; 15(2): 182-4. http://dx.doi.org/10.1038/cmi.2017.92 PMID: 28890542
[38] Naz A, Shahid F, Butt TT, Awan FM, Ali A, Malik A. Designing multi-epitope vaccines to combat emerging coronavirus disease 2019 (covid-19) by employing immuno-informatics approach. Front Immunol 2020; 11: 1663. http://dx.doi.org/10.3389/fimmu.2020.01663 PMID: 32754160
[39] Oyarzun P, Ellis JJ, Gonzalez-Galarza FF, *et al.* A bioinformatics tool for epitope-based vaccine design that accounts for human ethnic diversity: Application to emerging infectious diseases. Vaccine 2015; 33(10): 1267-73. http://dx.doi.org/10.1016/j.vaccine.2015.01.040 PMID: 25629524
[40] Zhou WY, Shi Y, Wu C, *et al.* Therapeutic efficacy of a multi-epitope vaccine against Helicobacter pylori infection in BALB/c mice model. Vaccine 2009; 27(36): 5013-9. http://dx.doi.org/10.1016/j.vaccine.2009.05.009 PMID: 19446591
[41] Samad A. Designing a multi-epitope vaccine against sars-cov-2: An immunoinformatics approach. J Biomol Struct Dyn 2020; •••: 1-17. PMID: 32677533
[42] Alshabrmi FM, Alatawi EA. Subtractive proteomics-guided vaccine targets identification and designing of multi-epitopes vaccine for immune response instigation against Burkholderia pseudomallei. Int J Biol Macromol 2024; 270(Pt 1): 132105. http://dx.doi.org/10.1016/j.ijbiomac.2024.132105 PMID: 38710251
[43] Aarthy M, Pandiyan GN, Paramasivan R, Kumar A, Gupta B. Identification and prioritisation of potential vaccine candidates using subtractive proteomics and designing of a multi-epitope vaccine against Wuchereria bancrofti. Sci Rep 2024; 14(1): 1970. http://dx.doi.org/10.1038/s41598-024-52457-x PMID: 38263422
[44] Suhrbier A. Multi-epitope DNA vaccines. Immunol Cell Biol 1997; 75(4): 402-8. http://dx.doi.org/10.1038/icb.1997.63 PMID: 9315485
[45] Tarrahimofrad H, Rahimnahal S, Zamani J, Jahangirian E, Aminzadeh S. Designing a multi-epitope vaccine to provoke the robust immune response against influenza A H7N9. Sci Rep 2021; 11(1): 24485. http://dx.doi.org/10.1038/s41598-021-03932-2 PMID: 34966175
[46] Naveed M, Sheraz M, Amin A, *et al.* Designing a novel peptide-based multi-epitope vaccine to evoke a robust immune response against pathogenic multidrug-resistant providencia heimbachae. Vaccines (Basel) 2022; 10(8): 1300. http://dx.doi.org/10.3390/vaccines10081300 PMID: 36016188
[47] Enayatkhani M, Hasaniazad M, Faezi S, *et al.* Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: An *in silico* study. J Biomol Struct Dyn 2021; 39(8): 2857-72. http://dx.doi.org/10.1080/07391102.2020.1756411 PMID: 32295479
[48] Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. Nat Methods 2022; 19(6): 679-82. http://dx.doi.org/10.1038/s41592-022-01488-1 PMID: 35637307
[49] Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007; 8(1): 4. http://dx.doi.org/10.1186/1471-2105-8-4 PMID: 17207271
[50] Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. Bioinformatics 2008; 24(11): 1397-8. http://dx.doi.org/10.1093/bioinformatics/btn128 PMID: 18413329
[51] Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: Accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res 2008; 36(Suppl. 2): W509-12. http://dx.doi.org/10.1093/nar/gkn202 PMID: 18463140
[52] Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 2020; 48(W1): W449-54. http://dx.doi.org/10.1093/nar/gkaa379 PMID: 32406916
[53] Sturniolo T, Bono E, Ding J, *et al.* Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol 1999; 17(6): 555-61. http://dx.doi.org/10.1038/9858 PMID: 10385319
[54] Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLOS Comput Biol 2008; 4(4): e1000048. http://dx.doi.org/10.1371/journal.pcbi.1000048 PMID: 18389056
[55] Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 2006; 65(1): 40-8. http://dx.doi.org/10.1002/prot.21078 PMID: 16894596
[56] Bui HH, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics 2007; 8(1): 361. http://dx.doi.org/10.1186/1471-2105-8-361 PMID: 17897458
[57] Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 2006; 7(1): 153. http://dx.doi.org/10.1186/1471-2105-7-153 PMID: 16545123

[58] Nguyen MN, Krutz NL, Limviphuvadh V, Lopata AL, Gerberick GF, Maurer-Stroh S. AllerCatPro 2.0: A web server for predicting protein allergenicity potential. Nucleic Acids Res 2022; 50(W1): W36-43. http://dx.doi.org/10.1093/nar/gkac446

[59] Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. In silico approach for predicting toxicity of peptides and proteins. PLoS One 2013; 8(9): e73957. http://dx.doi.org/10.1371/journal.pone.0073957 PMID: 24058508

[60] Artimo P, Jonnalagedda M, Arnold K, *et al.* ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 2012; 40: W597-603. PMID: 22661580

[61] Dhanda SK, Vir P, Raghava GPS. Designing of interferon-gamma inducing MHC class-II binders. Biol Direct 2013; 8(1): 30. http://dx.doi.org/10.1186/1745-6150-8-30 PMID: 24304645

[62] McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000; 16(4): 404-5. http://dx.doi.org/10.1093/bioinformatics/16.4.404 PMID: 10869041

[63] Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. Nature 2021; 596(7873): 583-9. http://dx.doi.org/10.1038/s41586-021-03819-2 PMID: 34265844

[64] Lee GR, Won J, Heo L, Seok C. GalaxyRefine2: Simultaneous refinement of inaccurate local regions and overall protein structure. Nucleic Acids Res 2019; 47(W1): W451-5. http://dx.doi.org/10.1093/nar/gkz288 PMID: 31001635

[65] Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007; 35(Web Server): W407-10. http://dx.doi.org/10.1093/nar/gkm290 PMID: 17517781

[66] Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 1993; 2(9): 1511-9. http://dx.doi.org/10.1002/pro.5560020916 PMID: 8401235

[67] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Cryst 1993; 26(2): 283-91. http://dx.doi.org/10.1107/S0021889892009944

[68] Yan Y, Tao H, He J, Huang SY. The HDOCK server for integrated protein-protein docking. Nat Protoc 2020; 15(5): 1829-52. http://dx.doi.org/10.1038/s41596-020-0312-x PMID: 32269383

[69] Grote A, Hiller K, Scheer M, *et al.* JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 2005; 33: W526-31. http://dx.doi.org/10.1093/nar/gki376 PMID: 15980527

[70] Castiglione F, Duca K, Jarrah A, Laubenbacher R, Hochberg D, Thorley-Lawson D. Simulating epstein-barr virus infection with C-ImmSim. Bioinformatics 2007; 23(11): 1371-7. http://dx.doi.org/10.1093/bioinformatics/btm044 PMID: 17341499