J Appl Biotechnol Rep. 2021 June;8(2):146-154

doi 10.30491/JABR.2020.253820.1303

Journal of
Applied Biotechnology
Reports

Original Article

Multi-epitope Based Peptide Vaccine Design Using Three Structural Proteins (S, E, and M) of SARS-CoV-2: An *In Silico* Approach

Arpita Singha Roy[1], Mahafujul Islam Quadery Tonmoy[1], Atqiya Fariha[1], Ithmam Hami[1], Ibrahim Khalil Afif[1], Md. Adnan Munim[1], Mohammad Rahanur Alam[2], Md. Shahadat Hossain[1*]

[1]Department of Biotechnology & Genetic Engineering, Noakhali Science and Technology University, Sonapur 3814, Noakhali, Bangladesh
[2]Department of Food Technology and Nutrition Science, Noakhali Science and Technology University, Sonapur 3814, Noakhali, Bangladesh

Corresponding Author: Md. Shahadat Hossain, MSc, Department of Biotechnology & Genetic Engineering, Noakhali Science and Technology University, Sonapur 3814, Noakhali, Bangladesh. Tel: +880-1757691222, Email: shahadat5.bge@nstu.edu.bd, reyad.geb.du@gmail.com

Received October 24, 2020; **Accepted** December 29, 2020; **Online Published** May 21, 2020

Abstract

Introduction: The ongoing global pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has jeopardized our health system and leaving everyone in disarray. Despite the diligent cumulative effort of academia, there is hardly any light in the end tunnel so far in developing efficient and sustainable treatment options to tackle this public health threat. Therefore, designing a suitable vaccine to overcome this hurdle calls for immediate attention. The current study aimed to design a multi-epitope based vaccine using immunoinformatics tools.

Materials and Methods: We approached the structural proteins: S, E, and M proteins of SARS-CoV-2 since they facilitate the infection of the virus into a host cell. By using different bioinformatics tools and servers, the multiple B-cell and T-cell epitopes were predicted potential for the required vaccine design. The phylogenetic analysis provides in-depth knowledge on ancestral molecular changes and the molecular evolutionary relationship of S, E, and M proteins.

Results: Based on the antigenicity and surface accessibility of the spike (S), envelope (E), and membrane (M) proteins, eight epitopes were selected by various B cell and T cell epitope prediction tools. Molecular docking was executed to interpret the binding interactions of these epitopes from where three potential epitopes WTAGAAAYY, YVYSRVKNL, and GTITVEELK were finalized with their noticeable higher binding affinity scores -9.1,-7.4, and -7.0 kcal/mol, respectively. It is noteworthy to mention that the targeted epitopes are believed to cover 91.09% of the population coverage worldwide.

Conclusions: In sum, we identified the three most potential epitopes at length, which might be turned to our purpose of designing the peptide-based vaccine against SARS-CoV-2.

Keywords: SARS-CoV-2, Structural Protein, Epitope, Antigenicity, Molecular Docking

Citation: Singha Roy A, Tonmoy MIQ, Fariha A, Hami I, Afif IK, Munim MA, et al. Multi-epitope Based Peptide Vaccine Design Using Three Structural Proteins (S, E, and M) of SARS-CoV-2: An In Silico Approach. J Appl Biotechnol Rep. 2021;8(2):146-154. doi:10.30491/JABR.2020.253820.1303.

Introduction

History suggests that humanity is always confronted by emerging viruses and viral infections in terms of livelihood and economic progress in a population. The current situation of the ongoing pandemic of the coronavirus disease 2019 (COVID-19) verily supports this history. The disease came to recognition by the World Health Organization (WHO) as a pandemic on March 11, 2020 and has caused a global emergency across 210 countries and territories worldwide and two international conveyances.[1] As of 9th of December 2020 at 20.09 (GMT +6), a confirmed report of 19.5 million active infected cases with 0.5% criticality and over 1.5 million deaths has been found (https://www.worldometers.info/coronavirus/). Initially, the disease originated back in December 2019 in Wuhan City, Hubei province, China, in the form of a cluster of pneumonia-like symptoms that quickly transcended the border to spread across the globe within a short span of time.[2]

The causative agent behind the infection was first designated as 2019 novel coronavirus (2019-nCoV) by the WHO.[3] The novel coronavirus was further renamed to severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) because of its genomic similarity of 79.5% and 96% at the nucleotide level, respectively, with SARS-CoV and bat coronavirus.[4, 5] Through phylogenetic analysis, SARS-CoV-2 has been categorized under the family Coronaviridae and order Nidoviralae and has shown an origin in the bat as a natural host.[3,6] The SARS-CoV-2 is similar to the SARS-CoV

and MERS-CoV of the same genus *Betacoronavirus* within the same family regarding infectivity in humans. The latter two viruses caused endemic situations in recent years and were the reasons for thousands of deaths across the world.[7-9] As for the structural similarity, the SARS-CoV-2 contains a positive single-stranded RNA as its genetic element having a genomic length of around 30 kilobases.[10] The encoded proteins by this genome are structural and non-structural, and the major structural proteins are S glycoprotein, M protein, E protein, and nucleocapsid (N) protein.[11,12] The exposed parts of these proteins contain domains necessary for infection into the host cells and account for antigenicity. The similarity between SARS-CoV and SARS-CoV-2 extends to the structure of spike glycoprotein, which may be exploited for a potential vaccine design because the S protein has two major subunits, namely S1 and S2, which assist during viral infection into the host cell. S1 subunit contains a receptor-binding domain (RBD) and N-terminal domain (NTD) where RBD binds to the receptor of the host cell known as angiotensin-converting enzyme-2 (ACE2).[11,13] The E protein also has its role in establishing a series of involvement from pathogenesis to the viral assembly by not only interacting with the host cell protein but also maintaining a defining connection with all the other structural proteins (M, S, N proteins) of the virus.[14] Moreover, the exposed portion of the M protein outside the membrane provides us an opportunity to design a suitable epitope-based subunit vaccine.

The scientific community is currently utilizing different therapeutic strategies to combat this dangerous spread of COVID-19, most of which are opted to develop novel drugs or vaccines. Chinese traditional drugs named *ShuFengJieDu* capsules and *LianHua Qing Wen* capsules had been applied to some of the preliminary cases before they were reported to be effective. However, no clinical trials to this date have ever proven them to be safe enough.[15] Some other drugs, such as remdesivir and chloroquine, had also been reported to be effective against COVID-19 through in-vitro trials. However, no authentic clinical trial has justified such a claim so far.[15,16] Besides, SARS-CoV-2 is dispersing too fast across almost all the countries of the world and is mutating in an unbelievable manner. Due to the high mutation rate in the genome of the SARS-CoV-2 single epitope will not suffice to provide for a successful vaccine; instead, multiple epitope-based vaccines may do the trick here.[17]

At the current stage of the pandemic, it will be highly insensible to develop a vaccine through a classical approach in-vitro that involves the identification, isolation, and culture of pathogenic viruses. This particular process will actually be too expensive and time-consuming, which is not desirable at all the given circumstances. A very sustainable way to overcome such hurdles would be to design a peptide vaccine by genome and proteome analysis of the virus using computational methods. Since the genome and proteome sequencing of SARS-CoV-2 has already been done, it is only rational that we take control measures by making the best out of computation based analysis to design therapeutic targets. In this study, we explored the S, E, and M proteins of SARS-CoV-2 by using different *in silico* tools and servers to predict B-Cell and T-Cell epitopes to eventually design an effective epitope-based vaccine. The predicted epitopes were analyzed further to check their antigenicity and surface accessibility. Epitope-allele interaction was investigated through molecular docking. A phylogenetic tree was constructed to identify the selected S, E, and M proteins' molecular evolutionary relationships. The study was concluded with the introduction of a properly designed vaccine from the most suitable epitopes.

Materials and Methods

Retrieval of Protein Sequences

The FASTA format of S, E, and M protein sequences of SARS-CoV-2 from various geographical areas: Australia, China, USA, Finland, India, Sweden, South Korea were retrieved from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/). Then, the BLAST program from NCBI was used to derive similar sequences against the proteins.

Phylogenetic Tree Construction and Analysis

The S, E, and M protein sequences of SARS-CoV-2, SARS-CoV, MERS, and common human coronavirus strains (229E, NL63, OC43, and HKU1) were targeted for this phylogenetic study. All the sequences of the proteins retrieved from the NCBI BLASTP result were aligned separately through the ClustalW algorithm by utilizing the MEGA (version 10.0.5).[18] All the required parameters for the alignment analysis were used as the default program in the software. The aligned sequences were then visualized with the Jalview (version 2.11.0)[19] to observe consensus and conserved sequences. The phylogenetic trees were built using the neighbor-joining tree function and default analysis preferences in MEGA.

Membrane Topology Analysis

Epitopes of a protein must be in the exposed regions to mount a sufficient immune response. The membrane topology of these proteins was analyzed using the TMHMM v2.0 server[20] and was later cross-referenced with the InterPro server.[21,22] The outer membrane regions of these proteins were selected for further analysis.

Antigenicity Prediction

A vaccine candidate must elicit a sufficient antigenic response as the antigenicity of epitopes plays a crucial role in provoking an adequate immune response. VaxiJen v2.0 server[23] calculates antigenicity depending on physicochemical properties of proteins with the threshold value 0.4 (for viral protein sequence).

B Cell Epitope Identification

B cell epitopes presented on the virus surface proteins are recognized by B lymphocytes to elicit an immune response. Based on an artificial neural network, the ABCpred v2.0 server[24-27] predicted the linear B cell epitopes. The epitopes were cross-referenced with the Immune Epitope Database (IEDB),[28] which uses amino acid scales and hidden Markov

models (HMMs) as a prediction method. Moreover, Kolaskar and Tongaonkar *antigenicity*, Parker hydrophilicity prediction tools from IEDB were explored to determine these selected epitopes' antigenicity and hydrophilicity properties.

T Cell Epitope Identification

T cell epitopes consist of a group of amino acids, presented by an antigen-presenting cell in the bound form with major histocompatibility (MHC) molecules to mount T cell-mediated immune response. Prediction of T cell epitopes was performed using the NetCTL tool, which utilizes MHC binding affinity, proteasomal processing, and TAP transport.[29-31] The tool predicts half-maximal inhibitory concentration (IC50) values of epitopes based on artificial neural network.[32,33] The lengths for epitopes were set at 9.0 and 15.0 for MHC I molecule and MHC II molecule, respectively.

Molecular Docking Analysis of HLA and Epitopes

The three-dimensional structures of targeted T-cell epitopes were modeled by the PEP-FOLD3 server,[34] a peptide structure predicting tool. The best 3D structure generated by this server was selected as a ligand for docking analysis. Molecular docking was performed by the virtual screening tool PyRx[35] through its Autodockvina[36,37] program in order to analyze the interactions among our proposed epitopes and different HLA molecules. The protein data bank provided most of the PDB files of the selected HLA molecules for docking study, whereas Phyre2 protein prediction server (Protein Homology/analogy Recognition Engine v2.0)[38] was used to generate the 3D models of other HLA molecules whose structures were not available in PDB. Discovery Studio (v4.5)[39] prepared these HLA molecules as macromolecules by removing water, non-polar hydrogen, and unnecessary molecules. The PPDB files were converted into PDBQT files, and the default grid box parameters were maintained with exhaustiveness value 8. The binding interactions of epitope- HLA molecules were visualized using UCSF Chimera 1.13rc.[40]

Population Coverage Prediction

Determination of population coverage for individuals is essential as epitopes may exhibit variation in their binding sites during interaction with different HLA alleles. The IEDB population coverage calculation tool[41] was utilized to determine the percentage of people expected to respond to a specific number of MHC-restricted epitopes worldwide.

Results

Sequence Retrieval

Respective amino acid sequences of S glycoprotein, E protein, and the M protein of SARS-CoV-2 were retrieved from NCBI. A total of 13 sequences of S protein (Figure S1), 13 sequences of M protein (Figure S2), and 6 sequences of E protein (Figure S3) were selected for the analysis. The length of the S, M, and E proteins were 1273, 220, and 75 amino acids, respectively.

Phylogenetic Tree Construction and Analysis

From the phylogenetic analysis of spike glycoprotein from the different strains of MERS, SARS-CoV, SARS-CoV-2, and common human coronaviruses (Human coronavirus strains: 229E, NL63, OC43, and HKU1) (Table S1-S3), we noticed that their ancestral and evolutionary relationship has been comprehensive and relatable (Figure S4). The analysis also showed that both human coronavirus OC43 and HKU1 are closely related to the MERS virus, whereas distantly related to the SARS and SARS-CoV-2. On the other hand, human coronavirus NL63 and 229E are closely related to each other. Phylogenetic analysis of the membrane proteins (Figure S5) of different strains of MERS, SARS-CoV, SARS-CoV-2 and envelop proteins (Figure S6) from diverse strains of MERS, SARS-CoV, SARS-CoV-2 along with common human coronaviruses strains (229E, NL63, OC43, and HKU1) (Table S1-S3) showed that SARS-CoV and SARS-CoV-2 are closely related to MERS virus whereas OC43 and HKU1 strains are closely related and also comparatively closer to the SARS-CoV, SARS-CoV-2 and MERS virus rather than the other closely related 229E and NL63 strains. Unlike the ancestral relationship in regards of the S and M proteins, the analysis showed a different result for the E protein. Actually, SARS-CoV and SARS-CoV-2 are closely related to MERS virus whereas human coronavirus 229E and NL63 are closely related and also comparatively closer to the SARS-CoV, SARS-CoV-2, and MERS virus rather than the other closely related strains of the human coronavirus OC43 and HKU1.

Prediction of Exposed Regions and Antigenicity

Viral infection has initially been facilitated by the outer domain of surface proteins, which bind with the receptor of the host cell. Interpro and TMHMM servers were used to predict the outer portions of each protein. The length of exposed regions (Non-cytoplasmic region) ranging from 1-1213 (Figure 1), (1-11, 35-75) and (1-19, 74-78) respectively for S, E, and M proteins were taken for our assessment. VaxiJen 2.0 server was used to predict antigenicity of all the respective proteins. The S protein region had an antigenic value of 0.4646, whereas values for the region of the E and M protein were 0.7282 and 0.5102, respectively.

Prediction of B-Cell Epitope

Primary sequences of S, E, and M protein were investigated

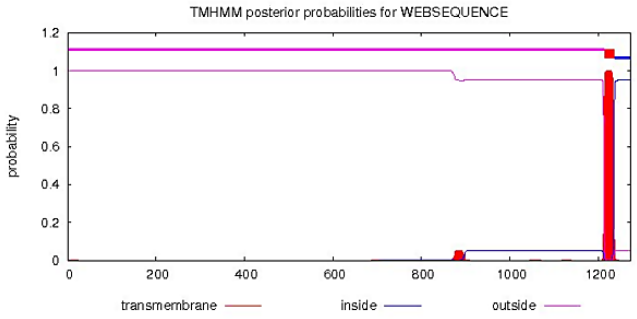

Figure 1. Membrane Topology Analysis. Exposed region of spike protein analyzed by TMHMM v2.0 server.

through ABCpred and IEDB server to predict B-cell epitopes. A total of 59, 5, and 19 B-cell epitopes for S, E, and M proteins were predicted individually by ABCpred. From all anticipated epitopes, only four epitopes, S (2 epitopes), E (1 epitope), and M (1 epitope) presented on the outside of S, E, and M proteins, were chosen (Table 1) with higher antigenicity scores. The Kolaskar and Tongaonkar antigenicity estimation tool (Figure 2) was exploited to predict the antigenicity score (Table 1), and the surface accessibility was determined by the TMHMM server.

Based on ABCpred, among these four identified epitopes of S protein, 'HRSYLTPGDSSSGWTA' and 'CFTNVYADSFVIRGDE' showed the most elevated antigenic score of 0.92 and 0.85, respectively. Moreover, epitope 'HRSYLTPGDSSSGWTA' is positioned within the NTD region, whereas 'CFTNVYADSFVIRGDE' is inside the RBD region. The E protein epitope 'NVSLVKPSFYVYSRVK' showed a 0.8 antigenic score. This epitope is situated at 48 position of E protein while at 7 position of M protein, the 'TITVEELKKLLE' revealed 0.75 antigenic score and was selected for M protein (Table 1). According to IEDB, the 'TITVEELKKLLE' indicated the most noteworthy antigenicity of 1.113, followed by the 'CFTNVYADSFVIRGDE,' 'TITVEELKKLLE' and 'HRSYLTPGDSSSGWTA' with the antigenicity of 1.048, 1.043, and 0.993, respectively. The ABCpred score and VaxiJen antigenicity results revealed the ability of all predicted peptides to expand barrier reactions within the host during SARS-CoV-2 infection as an extracellular part of the transmembrane protein. The Parker-hydrophilicity strategy was performed to discover the hydrophilicity of anticipated B-cell epitopes (Figure 3). With the hydrophilicity analysis, the 'HRSYLTPGDSSSGWTA' was found to have remarkable hydrophilicity with the value of 2.95, whereas the hydrophilicity values of 'CFTNVYADSFVIRGDE,' 'TITVEELKKLLE,' and 'NVSLVKPSFYVYSRVK' were 1.513, 0.492, and 0.45, respectively (Table 1).

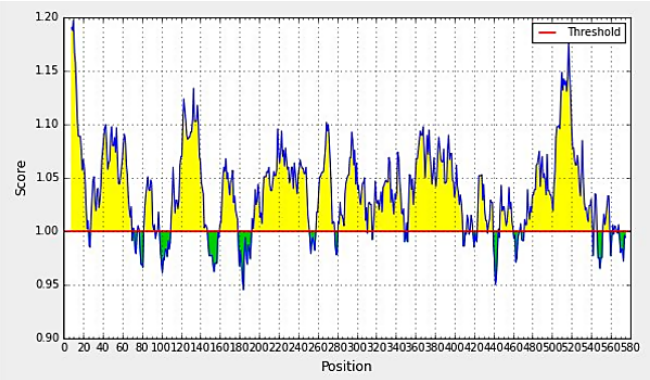

Figure 2. Antigenicity Prediction of the Epitope. Antigenicity determination of proposed epitope of S protein was performed by Kolaskar and Tongaonkar antigenicity prediction tool [Threshold value=1.00]. Yellow color indicates the antigenic nature of the residues.

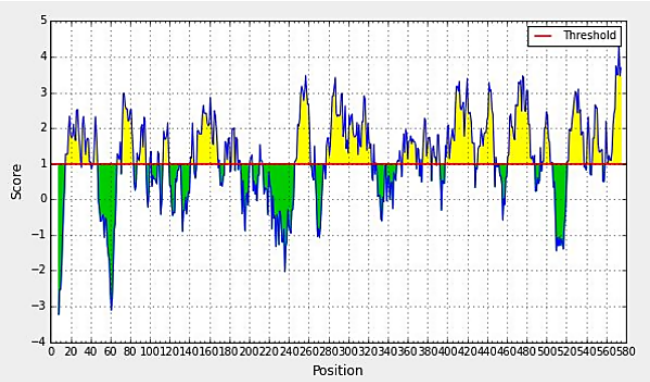

Figure 3. Hydrophilicity Prediction of the Epitope. Here, hydrophilicity of the anticipated B-cell epitope of S protein was predicted by Parker-hydrophilicity prediction tool with the threshold value 1.00 where hydrophilic nature of the residues was indicated by the yellow regions.

Prediction of T-Cell Epitope

Here, the IEDB server was utilized to evaluate the best T-cell epitopes from the chosen protein sequences of S, E, and M proteins. Moreover, antigenicity testing and screening of peptides were done with the help of the VaxiJen 2.0 server. In light of the high combinatorial score, the four best epitopes for MHC Class-I (Table 2) were chosen for additional investigation. For MHC class I assessment, NetMHCcons 1.1 Server is a consensus approach that combines the three best NetMHC, NetMHCpan, and PickPocket class strategies to provide the most accurate predictions. The MHC-I alleles for which the epitopes demonstrated higher affinity (IC50, 500 nM) were chosen (Table 2).

Between MHC class-I anticipated epitopes, a 9-mer epitope, 'KIADYNYKL' which is inside the RBD region of SARS-CoV-2 showed a higher antigenicity score of 1.6639 followed by the 'GTITVEELK,' 'YVYSRVKNL' and 'WTAGAAAYY' (within the NTD region of SARS-CoV-2) with an antigenicity score of 1.0976, 0.7020, and 0.5371, respectively. Based on the high combinatorial score for MHC Class-II, the four best epitopes (Table S4) were chosen for further analysis. The NetMHCII 2.3 server was used to anticipate MHC-II binding prediction with HLA-DR, HLA-DQ, and HLA-

Table 1. Predicted B-Cell Linear Epitopes With ABCpred Score, Antigenicity and Hydrophilicity Score

Protein	Epitope	Position	ABCpred Score	Antigenicity (IEDB)	Hydrophilicity (IEDB)
S	HRSYLTPGDSSSGWTA	245-260	0.92	0.993	2.95
	CFTNVYADSFVIRGDE	391-406	0.85	1.048	1.513
E	NVSLVKPSFYVYSRVK	48-63	0.8	1.113	0.45
M	TITVEELKKLLE	7-18	0.75	1.043	0.492

Table 2. Predicted Epitopes for CD8+ T-Cell Along With Their Interacting MHC Class I Alleles With Affinity <500 nM

Protein	Epitope	Position	Antigenicity (Vaxigenv2.0)	MHC Class I Allele With Total Score Having IC50 Values < 500 nM
S	WTAGAAAYY	258-266	Probable antigen (0.5371)	HLA-A*01:01(25.27), HLA-A*01:07(34.48, HLA-A*01:02(34.85), HLA-A*01:21(44.76), HLA-A*01:21(112.68), HLA-A*25:02(144.23), HLA-A*25:04(168.61), HLA-A*26:02(4.31), HLA-A*26:04(13.51), HLA-A*26:01(14.63), HLA-A*26:03(176.4), HLA-A*29:02(6.57), HLA-A*29:04(8,82), HLA-A*30:02(17.68), HLA-A*30:03(22.83), HLA-A*34:06(33.88), HLA-A*68:05(67.26), HLA-B*15:17(5.59), HLA-B*15:15(78.15), HLA-B*15:08(79.29), HLA-B*35:10(13.94), HLA-B*35:07(22.65), HLA-B*35:07(23.62), HLA-C*02:03(234.76), HLA-C*02:05(317.47), HLA-C*03:01(37.99), HLA-C*03:02(37.99), HLA-C*05:04(81.16), HLA-C*12:04(280), HLA-C*12:05(283.13).
	KIADYNYKL	417-425	Probable antigen (1.6639)	HLA-A*02:02(5.74), HLA-A*02:05(10.51), HLA-A*02:01(11.35), HLA-A*02:01(228.25), HLA-A*30:06(228.25), HLA-A*32:05(30), HLA-A*32:06(63.32), HLA-A*32:02(69), HLA-B*39:02(319.38), HLA-C*02:03(50), HLA-C*05:04(411.22), HLA-C*15:02(312.72), HLA-C*15:06(315.22).
E	YVYSRVKNL	57-65	Probable antigen(0.7020)	HLA-A*02:01(348.45), HLA-A*02:02(264.44), HLA-A*02:03(27.56), HLA-A*02:06(314.41), HLA-A*02:11(44.60), HLA-A*02:13(349.43), HLA-A*02:14(408.68), HLA-A*02:21(240.69), HLA-A*26:02(359.95), HLA-A*26:03(484.69), HLA-A*32:07(497.98), HLA-A*34:01(377.61), HLA-A*68:23(80.87), HLA-B*07:34(250.99), HLA-B*07:54(497.90), HLA-B*08:01(116.20), HLA-B*08:04(308.30), HLA-B*08:05(82.40), HLA-B*08:07(205.49), HLA-B*08:10(38.25), HLA-B*08:11(103.48), HLA-B*08:13(39.61), HLA-B*15:17(104.28), HLA-B*15:29(454.66), HLA-B*39:17(420.38), HLA-B*42:07(239.11), HLA-C*01:12(403.25), HLA-C*01:21(84.79), HLA-C*02:05(253.04), HLA-C*02:12(416.56), HLA-C*02:16(405.44), HLA-C*03:01(21.02), HLA-C*03:02(21.02), HLA-C*06:02(187.05), HLA-C*06:03(147.22), HLA-C*06:05(294.93), HLA-C*07:01(205.07), HLA-C*07:03(145.68), HLA-C*08:09(234.76), HLA-C*08:11(234.76), HLA-C*12:02(56.67), HLA-C*12:03(12.24), HLA-C*12:06(12.98), HLA-C*14:02(45.82), HLA-C*14:03(92.38), HLA-C*15:02(177.20), HLA-C*15:04(88.76), HLA-C*16:04(38.33), HLA-C*16:01(85.37).
M	GTITVEELK	6-14	Probable antigen(1.0976)	HLA-A*11:01(32.06), HLA-A*68:01(35.92), HLA-A*03:12(223.83), HLA-A*11:64(34.92), HLA-A*11:60(36.36), HLA-A*03:72(302), HLA-A*03:50(335.69), HLA-A*30:26(184.37), HLA-A*31:03(319.02), HLA-A*31:06(317.70), HLA-A*68:04(289.14), HLA-A*68:10(17.15).

DP MHC class II alleles. The MHC-II alleles for which the epitopes demonstrated higher affinity (IC50, 500nM) were selected (Table S4). The peptide 'GVLTESNKKFLPFQQ' which is within the RBD region of SARS-CoV-2 was viewed as increasingly antigenic for its higher antigenicity score of 0.8200, followed by the 'YFKIYSKHTPINLVR' (inside the NTD area of SARS-CoV-2), 'FYVYSRVKNLNSSRV', and 'MADSNGTITVEELKK' with an antigenicity score of 0.8197, 0.6103 and 0.4367, respectively.

Molecular Docking Analysis

Among these eight selected T cell epitopes, 'WTAGAAAYY' bound in the binding pocket of HLA-B*35:01 allele (PDB ID: 4PRN) showed the highest binding score of -9.1 kcal/mol (Figure 4). H-bond receptor surface of "WTAGAAAYY," indicating non-bond interactions is shown in the Figure 5. Eight T-cell epitope sequences with their individual docking score are shown in Table 3.

On the other hand, 'YVYSRVKNL' bounded with HLA-A02:03 (PDB ID: 3OX8) and 'GTITVEELK' bounded with HLA-A*11:01 (PDB ID: 6JOZ) with the binding affinity of -7.4 and -7 kcal/mol, respectively which exhibited the good binding interaction of these epitopes with HLA molecules. Nine hydrogen bonds formed between 'WTAGAAAYY' and HLA-B*35:01 allele where TRP1 formed one bond with SER116 and another bond with TRP147, THR2 formed one bond with ASN80 and double bonds with LYS146, ALA7 formed a single bond with TYR9, TYR8 formed another with ASN63 and TYR9 formed double bonds with THR69. Hydrogen bonds with their distance are shown in Figure 6.

Population Coverage Analysis

On the basis of the binding interaction of our targeted epitopes with their respective HLA alleles, we finalized three epitopes for vaccine design having the highest binding affinity. Population coverage for these three epitopes demonstrated the consequences of the host genetic variation on the binding specificity of these targeted epitopes to class I HLA alleles. Their cumulative population coverage around the world was 91.09%, which revealed that these epitopes could cover about 91% (Figure 7) of the population from different regions of the world.

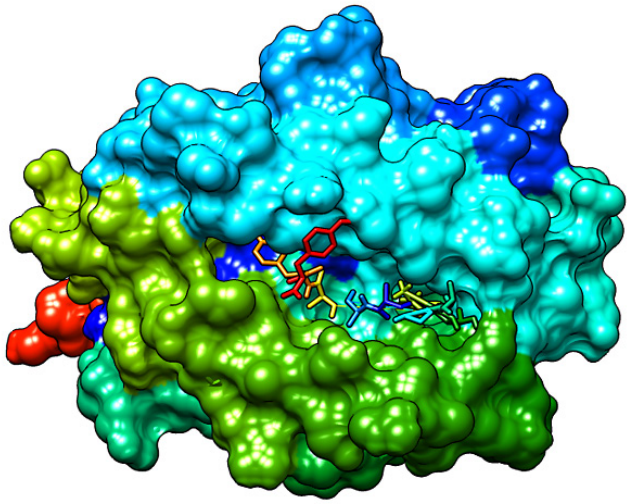

Figure 4. Molecular Docking Complex Analysis. Epitope "WTAGAAAYY" binds in the docking pocket of HLA-B*35:01 molecule. (PDB ID: 4PRN) with the binding affinity = -9.1 kcal/mol.

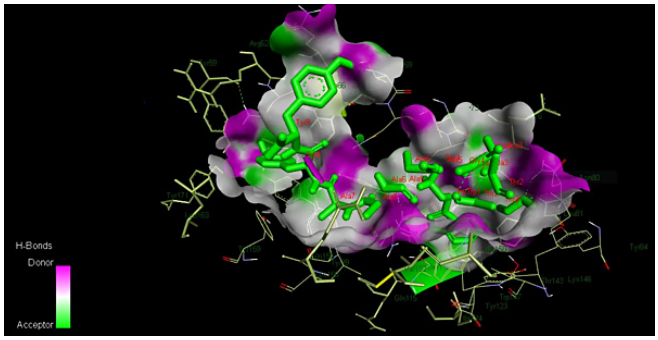

Figure 5. Non-bond Interactions Analysis. H-bond receptor surface of "WTAGAAAYY" representing non-bond interactions.

Discussion

One of the most prolific ways to reduce numerous viral diseases is to design novel vaccines.[42-45] In terms of SARS-CoV-2, ongoing research projects are being held to figure out the biological characteristics, genomics and proteomics, and pathophysiology of the virus worldwide.[46] Immunoinformatics tools have always been a medium to develop multi-epitope based vaccines for some of the dangerous viral diseases such as rhinovirus,[47] dengue virus,[48] chikungunya virus,[49] etc. Consistent with that idea, this study aimed at providing a peptide-based vaccine design against SARS-CoV-2 using different bioinformatic tools by exploring structural proteins of the virus, namely S glycoprotein, M protein, and E protein[50, 51] responsible for virulence mechanism and pathogenic pathways.

The multiple sequence alignment along with phylogenetic analysis of these structural (S, E, and M) protein sequences from diverse strains of MERS, SARS-CoV, SARS-CoV-2, and common human coronavirus (Human coronavirus strains: 229E, NL63, OC43, and HKU1) represented the ancestral molecular changes, evolutionary history, and the molecular evolutionary relationship. Among them, a phylogenetic tree of spike proteins highlighted a close relationship of OC43 and HKU1 with MERS, along with their distant relationship with SARS and SARS-CoV-2. While considering membrane proteins, OC43 and HKU1 were comparatively stronger in relation with SARS-CoV, SARS-CoV-2, and MERS rather than with 229E and NL63. A diversified result was also observed in the case of enveloping proteins where closely related strains 229E and NL63 were relatively closer to the SARS-CoV, SARS-CoV-2, and MERS rather than OC43 and HKU1.

Exposed portions of S, E, and M proteins of SARS-CoV-2 typically offer an extensive scope to induce antibody production against themselves.[52,53] Therefore, it is necessary to design a therapeutic target based on those exposed portions of the virus. Exploration of the exposed regions of our targeted proteins revealed that the S protein had a larger exposed region and was likely to be the best candidate for vaccine development rather than E and M proteins. Besides, a vaccine candidate must have antigenic features to maximize its reliability to be an agent of an immune response. To ensure the efficacy of the vaccine, the antigenicity of the targeted protein sequences must be evaluated as it is crucial for the design of a peptide-based vaccine.[54] The antigenicity analysis of S, E, and M proteins revealed the greater possibilities of these proteins to mount an immune response as a vaccine candidate.

Additionally, potential B cell epitopes play a significant role by providing protection against viral diseases.[55] Therefore, distinctive analysis techniques were utilized in this study for the prediction of a linear B cell epitope. Four potential linear B cell epitopes were selected with the best antigenicity

Table 3. Binding Affinity of T-Cell Epitopes and HLA Molecules

Protein Name	Epitope seq.	Allele	Docking Score (kcal/mol)
M Protein	GTITVEELK (MHC I)	HLA-A*11:01 (32.06)	-7
	MADSNGTITVEELKK (MHC II)	HLA-DQA1*01:02/DQB1*06:02 (215.20)	-5.5
E Protein	YVYSRVKNL (MHC I)	HLA-A02:03(27.56)	-7.4
	FYVYSRVKNLNSSRV (MHC II)	HLA-DRB1*04:01(12.50)	-6.2
S Protein	WTAGAAAYY (MHC I)	HLA-B*35:01(23.62)	-9.1
	KIADYNNKL (MHC I)	HLA-A*02:01(11.35)	-6.1
	YFKIYSKHTPINLVR (MHC II)	HLA-DRB1*04:01(185.8)	-6.4
	GVLTESNKKFLPFQQ (MHC II)	HLA-DRB1*03:01(188.1)	-4.6

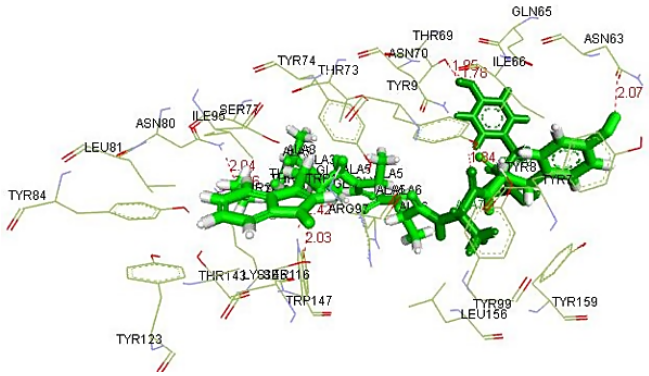

Figure 6. Analysis of Hydrogen Bond Formation Between Ligand (WTAGAAAYY) and Ligand Binding Site Atoms of the Receptor (HLA-B*35:01). In the docking complex, nine hydrogen bonds were formed between WTAGAAAYY' and HLA-B*35:01 allele. Among them, TRP1 formed one bond with SER116 (distance 2.02 Å) and another bond with TRP147 (distance 2.42 Å), THR2 formed one bond (distance 2.04Å) with ASN80 and double bonds (distance 2.39Å, 2.36Å) with LYS146, ALA7 formed single bond (distance 1.84 Å) with TYR9, TYR8 formed one bond (distance 2.07 Å) with ASN63, TYR9 formed double bonds (distance 1.78 Å,1.94 Å) with THR69.

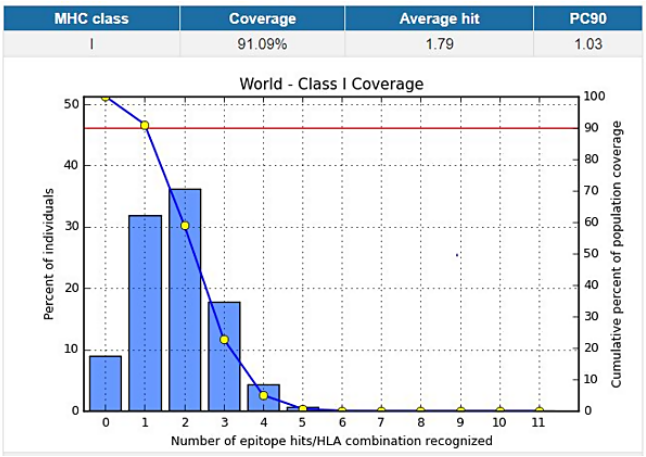

Figure 7. Population Coverage Analysis. Prediction of the population coverage for three potential vaccine candidates with MHC Class I HLA alleles around the world.

and hydrophilicity scores. The 'HRSYLTPGDSSSGWTA' situated in the NTD region exhibited the highest antigenic score of 0.92, and 'CFTNVYADSFVIRGDE' within the RBD region showed the second-highest antigenic score of 0.85, whereas NVSLVKPSFYVYSRVK and TITVEELKKLLE also showed antigenic properties. Their potentiality as vaccine candidates was also confirmed by calculating antigenicity and hydrophilicity through Kolaskar and Tongaonkar antigenicity and Parker-hydrophilicity analysis. The results of antigenicity and hydrophilicity suggest that all of these four selected linear B cell epitopes might have the potentiality to enhance humoral immunity as well as subsequent antibody neutralization.

B-cell epitope-based vaccines have been quite popular for a long time; however, vaccines based on T-cell epitopes are currently in the trend as the CD8+ T cells produce a long-term memory response in the host against the infected cell.[56] That is why the prediction of T-cell epitopes is an unquestionable requirement for a vaccine to have a preventive capacity. We predicted the best eight potential T cell epitopes from S, E, and M proteins. Among MHC class-I anticipated epitopes, higher antigenicity was observed for 'KIADYNYKL' situated in the RDB region, which was more responsible for binding with the host cell receptor than the other three epitopes within NTD region. The results concluded that along with the three other epitopes, 'KIADYNYKL' might significantly stimulate cellular immunity as they were represented to the cytotoxic T cells by the class-I MHC molecules. For MHC-II alleles, 'GVLTESNKKFLPFQQ' within the RBD region displayed higher antigenicity scores than the other three epitopes and displayed the helper T lymphocytes by MHC-II molecules, which might stimulate cellular and humoral immunity.

Moreover, molecular docking was performed to analyze the binding affinity of our targeted eight epitopes with different HLA molecules. Noticeable binding affinity score -9.1 kcal/mol was observed for the epitope 'WTAGAAAYY,' which was bound in the binding pocket of HLA-B*35:01 allele. Nine hydrogen bonds were formed between 'WTAGAAAYY' and HLA-B*35:01, which were observed in the docking complex. We also finalized epitopes' YVYSRVKNL', 'GTITVEELK' along with 'WTAGAAAYY' as ideal epitope-based vaccine candidates because of their notable docking scores.

Another essential prerequisite in epitope-based vaccine design is to determine population coverage as MHC polymorphism leads to the expression of different forms of HLA at considerably different rates among different ethnicities.[57,58] The three selected epitopes covered about 91.09% of population worldwide, signifying that these potential epitopes may be an effective universal vaccine in the future.

Most of the recent studies emphasized only the virus's spike protein, given its wide range of conservancy across the viral population. Nevertheless, our work was dedicated to specific epitopes and domains of all the mentioned proteins (S, M, and E proteins) for an *in-silico* design of a multi-epitope vaccine. Vaccine development against SARS-CoV-2 shed light on only the T-cell epitopes in some recent studies, whereas in this study, both eligible B and T cell epitopes of SARS-CoV-2 were explored for a maximized yield in this vaccine design approach.[59] All the ideal criteria of a potential epitope had been pulled off by the selected epitopes of S, M, and E proteins.

Conclusions

Different vaccine development strategies against SARS CoV2 are being opted along with ongoing trials on novel drug designs. In the current study, we suggested three epitopes from conserved regions of spike, membrane, and envelope proteins of SARS-CoV-2 as they possessed the abilities to produce an antigenic response. Their potentialities as vaccine candidates were assured by their antigenic properties and strong binding affinity for MHC molecules. This *in-silico* study requires experimental validation to get a cost-effective epitope-based peptide vaccine with higher efficacy against SARS-CoV-2.

Authors' Contributions

MSH conceived and designed this study. ASR, MIQT, AF, IBKA, IH, MAM performed the experiment and analyzed the data. MSH, MRA, ASR, IH wrote the manuscript. All authors reviewed the final manuscript.

Conflict of Interest Disclosures

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the Department of Biotechnology and Genetic Engineering, Noakhali Science and Technology University for providing the research facilities. The authors would also like to thank the Research Cell, Noakhali Science and Technology University and R&D, Ministry of Science and Technology for funding several research projects.

Supplementary Materials

Supplementary file 1 contains Figures S1-S6 and Table S1-S4 .

References

1. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed. 2020;91(1):157-160. doi:10.23750/abm.v91i1.9397.
2. Tang PF, Hou ZY, Wu XB, et al. Expert consensus on management principles of orthopedic emergency in the epidemic of coronavirus disease 2019. Chin Med J (Engl). 2020;133(9):1096-1098. doi:10.1097/cm9.0000000000000810.
3. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: evidence for virus evolution. J Med Virol. 2020;92(4):455-459. doi:10.1002/jmv.25688.
4. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7.
5. 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):727-733. doi:10.1056/NEJMoa2001017.
6. Hui DS, E IA, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis. 2020;91:264-266. doi:10.1016/j.ijid.2020.01.009.
7. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579(7798):265-269. doi:10.1038/s41586-020-2008-3.
8. Badawi A, Ryoo SG. Prevalence of comorbidities in the Middle East respiratory syndrome coronavirus (MERS-CoV): a systematic review and meta-analysis. Int J Infect Dis. 2016;49:129-133. doi:10.1016/j.ijid.2016.06.015.
9. Pallesen J, Wang N, Corbett KS, et al. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. Proc Natl Acad Sci U S A. 2017;114(35):E7348-E7357. doi:10.1073/pnas.1707304114.
10. 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):565-574. doi:10.1016/s0140-6736(20)30251-8.
11. Wang D, Yin Y, Hu C, et al. Clinical course and outcome of 107 patients infected with the novel coronavirus, SARS-CoV-2, discharged from two hospitals in Wuhan, China. Crit Care. 2020;24(1):188. doi:10.1186/s13054-020-02895-6.
12. Narayanan K, Maeda A, Maeda J, Makino S. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. J Virol. 2000;74(17):8127-8134. doi:10.1128/jvi.74.17.8127-8134.2000.
13. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020;181(2):271-280.e8. doi:10.1016/j.cell.2020.02.052.
14. Nieto-Torres JL, DeDiego ML, Verdiá-Báguena C, et al. Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. PLoS Pathog. 2014;10(5):e1004077. doi:10.1371/journal.ppat.1004077.
15. Yang Y, Islam MS, Wang J, Li Y, Chen X. Traditional Chinese medicine in the treatment of patients infected with 2019-new coronavirus (SARS-CoV-2): a review and perspective. Int J Biol Sci. 2020;16(10):1708-1717. doi:10.7150/ijbs.45538.
16. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. JAMA. 2020;323(18):1824-1836. doi:10.1001/jama.2020.6019.
17. Pachetti M, Marini B, Benedetti F, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med. 2020;18(1):179. doi:10.1186/s12967-020-02344-6.
18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547-1549. doi:10.1093/molbev/msy096.
19. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189-1191. doi:10.1093/bioinformatics/btp033.
20. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-580. doi:10.1006/jmbi.2000.4315.
21. Apweiler R, Attwood TK, Bairoch A, et al. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. 2001;29(1):37-40. doi:10.1093/nar/29.1.37.
22. Hunter S, Jones P, Mitchell A, et al. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 2012;40(Database issue):D306-312. doi:10.1093/nar/gkr948.
23. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8:4. doi:10.1186/1471-2105-8-4.
24. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 2017;45(W1):W24-W29. doi:10.1093/nar/gkx346.
25. Kolaskar AS, Tongaonkar PC. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett. 1990;276(1-2):172-174. doi:10.1016/0014-5793(90)80535-q.
26. Parker JM, Guo D, Hodges RS. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. 1986;25(19):5425-5432. doi:10.1021/bi00367a013.
27. Saha S, Raghava GP. Prediction of continuous B-cell epitopes

in an antigen using recurrent neural network. Proteins. 2006;65(1):40-48. doi:10.1002/prot.21078.
28. 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;8:361. doi:10.1186/1471-2105-8-361.
29. Fleri W, Paul S, Dhanda SK, et al. The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Front Immunol. 2017;8:278. doi:10.3389/fimmu.2017.00278.
30. Vita R, Overton JA, Greenbaum JA, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 2015;43(Database issue):D405-412. doi:10.1093/nar/gku938.
31. Zhang Q, Wang P, Kim Y, et al. Immune epitope database analysis resource (IEDB-AR). Nucleic Acids Res. 2008;36(Web Server issue):W513-518. doi:10.1093/nar/gkn254.
32. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal. 2000;22(5):717-727. doi:10.1016/s0731-7085(99)00272-1.
33. Bhasin M, Raghava GP. Prediction of CTL epitopes using QM, SVM and ANN techniques. Vaccine. 2004;22(23-24):3195-3204. doi:10.1016/j.vaccine.2004.02.005.
34. Lamiable A, Thévenet P, Rey J, Vavrusa M, Derreumaux P, Tufféry P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. 2016;44(W1):W449-454. doi:10.1093/nar/gkw329.
35. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. Methods Mol Biol. 2015;1263:243-250. doi:10.1007/978-1-4939-2269-7_19.
36. Sanner MF. Python: a programming language for software integration and development. J Mol Graph Model. 1999;17(1):57-61.
37. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-461. doi:10.1002/jcc.21334.
38. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009;4(3):363-371. doi:10.1038/nprot.2009.2.
39. Wang Q, He J, Wu D, Wang J, Yan J, Li H. Interaction of α-cyperone with human serum albumin: determination of the binding site by using Discovery Studio and via spectroscopic methods. J Lumin. 2015;164:81-85. doi:10.1016/j.jlumin.2015.03.025.
40. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-1612. doi:10.1002/jcc.20084.
41. 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;7:153. doi:10.1186/1471-2105-7-153.
42. De Groot AS, Rappuoli R. Genome-derived vaccines. Expert Rev Vaccines. 2004;3(1):59-76. doi:10.1586/14760584.3.1.59.
43. Fauci AS. Emerging and re-emerging infectious diseases: influenza as a prototype of the host-pathogen balancing act. Cell. 2006;124(4):665-670. doi:10.1016/j.cell.2006.02.010.
44. Korber B, LaBute M, Yusim K. Immunoinformatics comes of age. PLoS Comput Biol. 2006;2(6):e71. doi:10.1371/journal.pcbi.0020071.
45. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov. 2007;6(5):404-414. doi:10.1038/nrd2224.
46. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. 2020;12(3):254. doi:10.3390/v12030254.
47. Lapelosa M, Gallicchio E, Arnold GF, Arnold E, Levy RM. In silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes. J Mol Biol. 2009;385(2):675-691. doi:10.1016/j.jmb.2008.10.089.
48. Chakraborty S, Chakravorty R, Ahmed M, et al. A computational approach for identification of epitopes in dengue virus envelope protein: a step towards designing a universal dengue vaccine targeting endemic regions. In Silico Biol. 2010;10(5-6):235-246. doi:10.3233/isb-2010-0435.
49. Hasan MA, Hossain M, Alam MJ. A computational assay to design an epitope-based Peptide vaccine against saint louis encephalitis virus. Bioinform Biol Insights. 2013;7:347-355. doi:10.4137/bbi.s13402.
50. Frieman M, Heise M, Baric R. SARS coronavirus and innate immunity. Virus Res. 2008;133(1):101-112. doi:10.1016/j.virusres.2007.03.015.
51. Pang J, Wang MX, Ang IYH, et al. Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): a systematic review. J Clin Med. 2020;9(3):623. doi:10.3390/jcm9030623.
52. Gralinski LE, Menachery VD. Return of the coronavirus: 2019-nCoV. Viruses. 2020;12(2):135. doi:10.3390/v12020135.
53. Yong CY, Ong HK, Yeap SK, Ho KL, Tan WS. Recent advances in the vaccine development against Middle East respiratory syndrome-coronavirus. Front Microbiol. 2019;10:1781. doi:10.3389/fmicb.2019.01781.
54. Hossain MS, Reza HA, Hossain MS. Immunoinformatics approach identified two highly conserved B and T cell epitopes, LEASKRWAF and DSPLEASKRWAFRTG, for effective vaccine design against Ebola and Marburg viruses. J Adv Microbiol. 2019;17(1):1-16. doi:10.9734/jamb/2019/v17i130134.
55. Welsh RM, Selin LK, Szomolanyi-Tsuda E. Immunological memory to viral infections. Annu Rev Immunol. 2004;22:711-743. doi:10.1146/annurev.immunol.22.012703.104527.
56. Shrestha B, Diamond MS. Role of CD8+ T cells in control of West Nile virus infection. J Virol. 2004;78(15):8312-8321. doi:10.1128/jvi.78.15.8312-8321.2004.
57. Maenaka K, Jones EY. MHC superfamily structure and the immune system. Curr Opin Struct Biol. 1999;9(6):745-753. doi:10.1016/s0959-440x(99)00039-1.
58. Reche PA, Reinherz EL. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. J Mol Biol. 2003;331(3):623-641. doi:10.1016/s0022-2836(03)00750-2.
59. Sharmin R, Islam AB. A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design. BMC Bioinformatics. 2014;15:161. doi:10.1186/1471-2105-15-161.