Integrative Bioscience Nexus

Biologica Aid

Article

Epitope-Based Peptide Vaccine Design against Coronavirus using Immuno-informatics Approaches

Sabbir Ahmed[1,2], Md. Zakariya Alam Bappi[1,2], Shaila Akter Tumpa[1,2], Shatabdi Islam Shanta[1,2], Kaniz Fatema[1,2], Tasfia Wahed Annanya[1,2], Nilufa Akhter Banu[1,2], Md. Rezuanul Islam[1,2], Mohammad Abu Hena Mostofa Jamal[1,2*]

[1] Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Islamic University, Kushtia-7003, Bangladesh
[2] Laboratory of Medical and Environmental Biotechnology, Faculty of Biological Sciences, Islamic University, Kushtia-7003, Bangladesh
* Correspondence: jamalbtg@gmail.com

Abstract

The rapid emergence and global spread of coronavirus infections have emphasized the need for effective and rapidly deployable vaccine strategies. Epitope-based peptide vaccines have gained considerable attention due to their safety, specificity, and ability to induce targeted immune responses. In this study, an immunoinformatics-driven approach was employed to identify potential T-cell and B-cell epitopes for the development of a peptide vaccine against coronavirus. Viral protein sequences were analyzed to predict antigenic, immunogenic, conserved, and non-toxic epitopes capable of eliciting both cellular and humoral immune responses. Six highly antigenic CD8+ T-cell epitopes (AYAKLGLSI, YAKLGLSII, AKDKAYAKL, VEFKTGKLL, FAVTLKVEF, and LLAAKDKAY) were identified as promising vaccine candidates. All selected epitopes showed 100% sequence conservancy and demonstrated favorable interactions with major histocompatibility complex (MHC) molecules. Population coverage analysis indicated that these epitopes could potentially provide immune coverage for 71.63% of the global population. Furthermore, molecular docking studies confirmed strong binding affinity between the predicted epitopes and their respective immune receptors. Two conserved and non-toxic linear B-cell epitopes, YAKLGLSII and FAVTLKVEF, were also identified as potential candidates for antibody-mediated immunity. The findings of this study provide promising peptide vaccine candidates and highlight the usefulness of immunoinformatics approaches in accelerating coronavirus vaccine development through computational vaccine design and rational epitope selection.

Keywords: Coronavirus, immunoinformatics, peptide vaccine design, epitope prediction

Academic Editor: Dr. Shabana Bibi

Received: 10 May 2026
Revised: 27 June 2026
Accepted: 29 June 2026
Published: 09 July 2026

1. Introduction

Coronaviruses are part of the Coronaviridae family's Corona virinae subfamily, which are enveloped viruses of the class Nidovirales with a large plus-strand RNA genome. The genomic RNA size is 27–32 kb, coated and polyadenylated. Coronaviruses have been identified in rodents, rats, chickens, turkeys, swine, pigs, cats, rabbits, horses, animals, and humans [1], [2]. This virus causes many acute and chronic diseases of the respiratory, enteric, and central nervous system (CNS) of humans and other animal species [3]. In Wuhan, Hubei Province, China, the Human coronavirus (HCoV) was first identified, which usually works as an agent of upper

respiratory tract infections and also causes atypical pneumonia [4]. The latest COVID-19 outbreak has caused approximately 96,081 confirmed cases in China, with another 233,302,582 reported cases worldwide during the time period of December 2020 to September 2021. So far, about 4,773,253 have died, and 210,073,102 have recovered. WHO's projected mortality rate is 3.4 percent [5]. Cases are now identified across Asia, Europe, and the globe. Most human coronavirus infections are fairly mild. SARS-CoV and MERS-CoV are deadlier than any other coronavirus. The SARS-CoV epidemic resulted in a total of 916 deaths in more than 30 countries out of over 8000 confirmed cases [6]. Wuhan seafood market pneumonia virus, also labeled as Novel Coronavirus (2019-nCoV), is a descendant of both the devastating SARS and the MERS [7].

For the last 20 years, many viral epidemics have been reported, including the severe respiratory coronavirus syndrome (SARS-CoV) from 2002 to 2003 and H1N1 in 2009. In 2012, a viral respiratory disorder named "Middle East Respiratory Syndrome (MERS)" was first identified in Saudi Arabia [8]. Research performed by the China CDC and local CDCs shows that the incubation period of the coronavirus is almost 3 to 7 days, and up to 2 weeks and 12.5 days are required as the longest span from infection to symptoms. This outbreak expanded almost every seven days [9]. COVID-19 has been described as the sixth international public health emergency by the WHO. It is transmitted from human to human through droplets or direct touch, with an average incubation period of 6.4 days and a clear reproduction cycle of 2.24–3.58 days. At the end of 2019, in China, the severe respiratory syndrome of coronavirus 2 (SARS-CoV-2); previously identified as 2019 novel coronavirus or 2019-nCoV, or (COVID-19), triggered a massive global epidemic disease which becomes a serious public health issue. Patients with SARS-CoV-2 pneumonia (novel coronavirus pneumonia or Wuhan pneumonia) have the most severe symptoms, followed by coughing [10]. No effective therapeutic agent or vaccine for curing coronavirus infections is presently available on the market. Therefore, the development of an appropriate coronavirus therapy is a study goal [10]. There are currently few tools available to avoid coronavirus infections. The treatment is symptomatic, and oxygen therapy is the main medical method for seriously infected patients. Most deaths occurred in people older than 50 years of age [8]. Considering the ever-increasing transmission of this virus infection, the production of vaccines or antivirals for the infections of HCoVs is important [4]. The development of an epitope-based vaccine is our aim in this research, which could cause a substantial immune response, and to predict inhibitors that can bind to potential drug target sites using various immunoinformatics and simulation methods to dock them [11]. We tried to classify large immunogenic epitopes of the coronavirus proteins and predict a vaccine. The research will further improve future laboratory-based attempts to establish the successful treatment of coronavirus infection and prevention [12].

In our research, we analyzed 271 COVID-19 proteomes to classify the most antigenic protein as well as its strongest immunogenic and antigenic T-cell epitopes, along with B-cell epitopes, using immunoinformatics techniques to create a peptide vaccine based on an epitope. We also projected the 3D structure of the most antigenic polyprotein and inhibitor-binding sites using different bioinformatics tools to evaluate docking simulations. Overall, this research is intended to assist future laboratory efforts in the production of successful COVID-19 infection prevention vaccines [13]. This research offers detailed information about the potential epitope of the candidate vaccine. This prediction can be used by wet-lab researchers to verify this result.

2. Materials and Methods

2.1. Retrieval of the desired sequence

From the National Center for Biotechnology Information (NCBI), the FASTA-formatted coronavirus amino acid sequences have been recovered as a UniProtKB database, which also helps to retrieve the FASTA-formatted sequence from both GenBank and UniProtKB. In this analysis, the non-structural proteins (NSPs) were omitted from the 271 coronavirus proteome. Therefore, a total of 23 epitopes available were chosen primarily for the prediction of antigenicity [13], [14].

2.2. Antigenic protein identification

VaxiJen v2.0 (http://www.ddg-pharmfac.net / vaxijen / VaxiJen.html)) is a defensive antigen as well as a subunit vaccine for the prediction system, which was used to measure the most effective antigenic protein. Here, we used the server's default parameters to evaluate the antigenic protein [4], [15]. All the antigenic proteins were then filtered in Excel, with their respective scores. A particular antigenic protein with the highest antigenicity scores has been identified for further evaluation [16], [17].

2.3. Identification of T-cell epitope

T-cell epitopes are usually immunodominant peptide fragments that can conjure different immune responses, which are essential to design peptide-based vaccines. The NetCTL 1.2 server was used for the prediction of cytotoxic T-lymphocyte

(CTL) epitopes, which result from the protein sequence due to the value of T-cell epitopes. In any given protein, human CTL epitopes can be predicted by the NetCTL 1.2 server (https:// www.cbs.dtu.dk/services/NetCTL/) and this server can also predict the CD8+ T-cell epitopes for 12 supertypes such as A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62 [18]. We've described all 12 supertypes with epitopes. In this analysis, the epitope recognition threshold value was fixed at 0.75 and this has a sensitivity and specificity of respectively 0.54 and 0.993. As default parameters were used the weight based on the utmost cumulative score, at first, 23 epitopes were selected. And finally, the selection was completed following the epitopes prediction of antigenicity and immunogenicity through the VaxiJen v2.0 server and IEDB server, respectively. The epitopes were explored using the steadied matrix-based method (SMM) in the IEDB analytics tool (http:/tools.iedb.org/mhci/) to classify both frequently and non-frequently occurring MHC-I-binding alleles. The peptide 9.0 amino acid length and the IC_{50} value below 200 were chosen as parameters for the identification of MHC-I-binding alleles. Here, peptides with IC_{50} values < 50nM are known as high affinity, intermediate affinity < 500nM, and low affinity < 5000nM; so, higher affinity is indicated by the lower IC_{50} value. IEDB is a resourceful server that has also been used to predict the processing performance using SMM. The SMM-align method was used to determine the best candidate epitope's MHC-II-binding alleles [13], [19].

2.4. Prediction of an epitope conservancy

Using the IEDB analytical resource epitope tool (http:/tools.iedb.org/conservancy/), the epitope conservancy was predicted for specific peptides [16].

2.5. Prediction of population coverage

To analyze the population coverage of the predicted epitopes, the IEDB population coverage estimation method evaluating resources selected for a single epitope. For each consistent epitope, the allelic frequency of the interacting HLA alleles was used here to estimate the population coverage [4], [20].

2.6. Identification of the B cell epitope

Predicting hypothetically immunogenic epitopes in a given protein sequence can condense the wet laboratory effort successfully, which is required to detect the epitopes needed for vaccine design and immunodiagnosis. Identifying the potential antigen was the objective of the B cell epitope prediction that would interrelate with B lymphocytes and induce a humoral reaction. IEDB tools were used to identify B cell antigenicity, including Kolaskar and Tongaonkar antigenicity, Emini surface accessibility prediction, Karplus and Schulz flexibility prediction, and Bepipred linear epitope prediction analysis. The Chou and Fasman beta-turn prediction tool 36 is required for antigenic sections of a protein that belongs to the beta-turn regions [21], [22].

2.7. Allergenicity and toxicity extrapolation

Allergenicity had been predicted by AllerTOP v. 2.0 (http:/www.ddgpharmfac.net / AllerTOP/) and AllergenFP 1.0 (http:/ www.ddg-pharmfac.net/AllergenFP/). AllergenFP 1.0 was promoted by using a novel, descriptor-dependent, alignment-focused fingerprint process. AllergenFP 1.0 will distinguish allergens from non-allergic compounds with 87.9 percent validity. Moreover, AllerTOP v. 2.0 is a powerful and efficient complementary solution based on the method of classifying allergens and non-allergenic neighbors using k-nearest (kNN). Comparing various allergen prediction servers, AllerTOP v. 2.0 is the highest performing 88.7 percent accuracy platform surveyed by AllergenFP 1.0 (87.9 percent). The web server ToxinPred (http:/www.imtech.res.in/raghava/toxinpred/) enumerated the toxicity of the peptides. By using various peptide properties, the methods of machine learning, and a quantitative matrix, this technique was promoted [13], [23].

2.8. Three-dimensional structure design of the best epitope

According to the measurement of the conservation of epitopes, composite ranking, antigenicity, immunogenicity, non-toxicity, and abundance of alleles, the target epitopes YAKLGLSII and FAVTLKVEF were identified and the prediction database for the PEP-FOLD peptide structure (http:/bioserv.rpbs.univ-paris-diderot.fr / services / PEP-FOLD/) was sent to the Ressource Parisienne en Bioinformatique Mobyle portal. For the exploration of the link with MHC-I and MHC-II-binding alleles, we used the best model [13], [24], [25].

2.9. Docking simulation study

The research on molecular docking simulation was carried out using AutoDock Tools and AutoDock Vina software to learn the relationship between binding alleles and predicted epitopes. At this point, the crystal layout of the protein molecule HLA-C*12:03 was obtained in PDB format from the RCSB Protein Data Bank (PDB). The expected crystal layout was therefore in a complex shape with the protein and an epitope. But to simplify the dynamic layout, the Discovery Studio (version 16.1.0.15350) was used. The protein molecule and the peptide (epitope) were then translated to PDBQT format

using AutoDock Tools after the separation. The space box core was set at 4.496, −12.428, and −35.544Å in the X, Y, and Z axes respectively for the detection of binding energy at the binding groove of HLA-C*12:03 with an epitope; The scale for the X, Y, and Z dimensions was set at 30, 24, and 18Å, respectively, and these analyzes were performed at a spacing parameter of 0.964Å. With AutoDock Vina the docking simulation was eventually done [26], [27]. We regarded the HLA molecule as a receptor and our preferred epitope as a ligand blindly, meaning that the epitope would bind to the best position possible and not to the predetermined spot [28].

3. Results

3.1. Selection of the highest antigenic protein

The inquiry for Coronavirus structural and nonstructural proteins resulted in a total of 208 hits in the ViPR database. VaxiJen v2.0 server evaluated all the structural proteins, which predicted an overall score for each protein sequence and denote the protein sequence as antigen or nonantigen. The protein sequence showed the highest antigenic score 0.8257 among all the structural proteins with the GenBank id ADY39741.1. The protein itself is the corona polyprotein containing 73 amino acid residues, and this polyprotein was used for further analyses.

3.2. T-cell epitope identification

The NetCTL 1.2 webserver was used for the prediction of CD8+ T-cell epitopes, and we considered the combined score for the epitope selection. In this study, we unraveled 23 epitopes in total, which achieved the selected threshold value of 0.75 (**Table 1**). Furthermore, the predicted T-cell epitopes were evaluated by the VaxiJen server. This repetition results facilitated get a good rate of T-cell epitope prediction. Here, among 23 primarily selected T-cell epitopes. The results revealed that all epitopes with the highest combined score could not be the best epitopes. Among 23 epitopes, we selected 6 epitopes that have positive antigenicity and immunogenicity scores, so that these highly antigenic epitopes can interact with the MHC alleles with high affinity and can create an effective immune response (**Table 2**). Finally, we identified 2 epitopes containing a 100% conservancy score, since the conservancy score is a vital issue during the development of an effective vaccine. Epitope conservancy among the different strains of the selected protein can provide effective immunization. So, the higher conservancy of an epitope can ensure a better target for the improvement of vaccine design. The selected epitopes, conservancy score, and their position are shown in **Table 1**. In this study, 6 epitope conservancy analysis results revealed that the epitopes AYAKLGLSI, YAKLGLSII, AKDKAYAKL, VEFKTGKLL, FAVTLKVEF, LLAAKDKAY are 100.0% conserved. The immunogenicity score of the epitope can be a good criterion for the selection of the best epitope. A higher score describes a greater probability of eliciting an effective immune response. Herein, among the 6 epitopes, the epitopes FAVTLKVEF (-0.068) and YAKLGLSII (-0.12046) are less immunogenic than the epitopes AKDKAYAKL (-0.26816), VEFKTGKLL (-0.29778), and AYAKLGLSI (-0.28922) (**Table 1**).

Table 1. Most potential 6 T-cell epitopes with interacting MHC-I alleles, epitope conservancy score, allergenicity, and toxicity.

Epitope	AYAKLGLI	YAKLGLSII	AKDKAYAKL	VEFKTGKLL	FAVTLKVEF	LLAAKDKAY
Antigenicity	1.3232	1.3415	0.8805	0.866	1.7021	1.261
Toxicity	Non	Non	Non	Non	Non	Non
Allergenicity	Non	Non	Non	Non	Non	Non
MHC-1	HLA-C*14:02(4.3) HLA-C*12:03(77) HLA-C*03:03(58) HLA-A*24:02(0.3)	HLA-C*12:03(0.8) HLA-C*03:03(29) HLA-B*15:02(8.3) HLA-C*14:02(36)	HLA-C*12:03(20) HLA-B*15:02(5) HLA-C*08:02(1.1) HLA-C*05:01(13)	HLA-C*12:03(52) HLA-B*15:02(4.1) HLA-B*40:01(0.7) HLA-C*03:03(58) HLA-B*40:02(1.2)	HLA-C*03:03(1.5) HLA-B*35:01(0.5) HLA-C*12:03(39) HLA-B*15:02(5.5) HLA-B*53:01(0.4) HLA-B*46:01(0.2)	HLA-C*03:03(7.5) HLA-C*14:02(12) HLA-B*15:01(0.5) HLA-C*12:03(92) HLA-A*29:02(1.2) HLA-B*35:01(2.4)

The selected 6 T-cell epitopes were found to be recognized by the significant MHC class-I molecule such as HLA-A, HLA-B, and HLA-C according to the IEDB MHC class 1-binding analysis resource. We selected humans as MHC source species and the SMM method for the prediction of a distinct set of MHC HLA alleles for humans. This tool gives an output result for the HLA-binding affinity of the epitopes in the IC_{50} nM unit. A lower IC_{50} value indicates a higher binding affinity of the epitopes with the MHC class I molecule. So, in this study, due to confirm higher affinity, we chose IC_{50} values less than 200nM ($IC_{50} < 200$) (**Table 2**).

Effective immune response not only depends on the successful recognition of epitopes by HLA molecules with significant affinity but also depends on the antigenicity and immunogenicity score. So, the epitopes that were recognized by the considerable number of HLA alleles and contained the highest immunogenicity, antigenicity value, and non-toxic to humans were considered as the potential epitope to induce a strong immune response. So, for docking purposes, we selected two 100.0% conserved 9-mer epitopes "YAKLGLSII" and "FAVTLKVEF". Here, the epitope "YAKLGLSII" showed the highest affinity for 4 MHC-1 molecules, including HLA-C*12:03(0.8) whereas the epitope "FAVTLKVEF" revealed the affinity with HLA-C*03:03(1.5), so, 7 MHC-I alleles HLA-C*12:03, HLA-C*03:03, HLA-B*15:02 were found as common in both "YAKLGLSII" and "FAVTLKVEF".

Table 2. MHC-II alleles for the desired epitope.

Core sequence	MHC-II alleles	Peptide	IC_{50} value
YAKLGLSII	HLA-DRB4*01:01	DKAYAKLGLSIIEEV	203
	HLA-DRB1*01:01	DKAYAKLGLSIIEEV	44
	HLA-DRB1*01:01	KAYAKLGLSIIEEVN	44
	HLA-DRB1*01:01	AKDKAYAKLGLSIIE	47
	HLA-DRB1*01:01	KDKAYAKLGLSIIEE	47
	HLA-DRB1*09:01	DKAYAKLGLSIIEEV	501
	HLA-DRB1*04:05	DKAYAKLGLSIIEEV	610
	HLA-DRB1*01:01	AYAKLGLSIIEEVNS	128
	HLA-DRB1*01:01	YAKLGLSIIEEVNSH	134
	HLA-DRB4*01:01	AKDKAYAKLGLSIIE	777
	HLA-DRB4*01:01	KDKAYAKLGLSIIEE	779
	HLA-DRB1*04:05	KDKAYAKLGLSIIEE	847
	HLA-DRB1*04:05	AKDKAYAKLGLSIIE	877
	HLA-DRB1*07:01	DKAYAKLGLSIIEEV	646
	HLA-DRB1*11:01	AKDKAYAKLGLSIIE	839
FAVTLKVEF	HLA-DRB1*11:01	AYFAVTLKVEFKTGK	151
	HLA-DRB1*11:01	CAYFAVTLKVEFKTG	157
	HLA-DRB1*11:01	DCAYFAVTLKVEFKT	159
	HLA-DRB1*11:01	LDCAYFAVTLKVEFK	161
	HLA-DRB5*01:01	AYFAVTLKVEFKTGK	165
	HLA-DRB1*07:01	DCAYFAVTLKVEFKT	188
	HLA-DRB1*07:01	AYFAVTLKVEFKTGK	269
	HLA-DRB1*07:01	CAYFAVTLKVEFKTG	282
	HLA-DRB1*11:01	FAVTLKVEFKTGKLL	388
	HLA-DRB1*11:01	YFAVTLKVEFKTGKL	397
	HLA-DRB1*04:05	LDCAYFAVTLKVEFK	685
	HLA-DRB1*04:05	DCAYFAVTLKVEFKT	740
	HLA-DRB1*04:05	CAYFAVTLKVEFKTG	740
	HLA-DRB1*03:01	AYFAVTLKVEFKTGK	1840
	HLA-DRB1*03:01	DCAYFAVTLKVEFKT	1895
	HLA-DRB1*07:01	YFAVTLKVEFKTGKL	610
	HLA-DRB5*01:01	LDCAYFAVTLKVEFK	677
	HLA-DRB1*04:05	AYFAVTLKVEFKTGK	996
	HLA-DRB5*01:01	CAYFAVTLKVEFKTG	820
	HLA-DRB1*09:01	DCAYFAVTLKVEFKT	1051
	HLA-DRB1*09:01	LDCAYFAVTLKVEFK	1053
	HLA-DRB1*09:01	AYFAVTLKVEFKTGK	1247
	HLA-DRB1*09:01	CAYFAVTLKVEFKTG	1260

The T-cell epitopes with interacting MHC-I alleles & MHC-II alleles for the desired epitope are summarized in Table 1 and Table 2.

However, we selected HLA-B*15:02 for docking purposes due to the availability of this allele in the PDB database. After that, we further used the epitopes "YAKLGLSII" and "FAVTLKVEF" for the prediction of MHC-II alleles and their respective peptide or CD4+ T-cell epitope. The results showed that the epitope "YAKLGLSII" and respective 15-mer peptides "DKAYAKLGLSIIEEV" and "KAYAKLGLSIIEEVN" have a low affinity with the predicted MHC-II alleles (**Table 2**). Besides, the epitope "FAVTLKVEF" was found as the core sequence of 31 15-mer peptides or CD4+ T cell epitopes in the range of the IC_{50} value 1–2000 nM. Among these 21 peptides, nine peptides revealed a strong affinity (IC_{50} value 159 to 610 nM) with HLA-DR alleles, HLA-DRB1*11:01, HLA-DRB1*11:01, LA-DRB5*01:01, and HLA-DRB1*07:01 (**Table 2**). So, HLA-DR alleles could be the best binding grooves for the predicted peptides.

3.3. Analysis of population coverage and allergenicity

MHC HLA allele distribution differs among diverse geographic regions and ethnic groups around the world. Therefore, population coverage must be taken into consideration during the design of an effective vaccine. In this study, for the population coverage, identified MHC-I-binding alleles of 2 epitopes were considered. Significant population coverage was found for the selected 2 epitopes in different geographic regions of the world (**Figure 1**). These epitopes and their HLA-alleles cover 71.63 % of the world population. The highest population coverage was found in Italy (98.60%) which was closely followed by France, Germany, the United States, England, Finland, Saudi Arabia, and Brazil with population coverage of 97.90%, 97.94%, 97.83%, 97.73%, 96.14%, 94.73%, and 94.42%, respectively. The lowest population coverage was found in Peru Mestizo (1.99%). The Corona infection was first found in China, and several outbreaks have been recorded in this country. So, the population coverage prediction in China is essential for vaccine design. The cumulative population coverage in China is 87.05%.

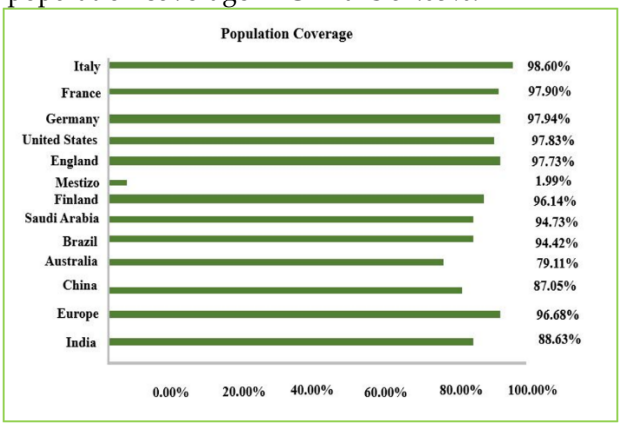

Figure 1. Population coverage of epitopes. The epitopes with different HLA-binding specificities increase population coverage, as several HLA types are expressed at radically different frequencies in various ethnicities.

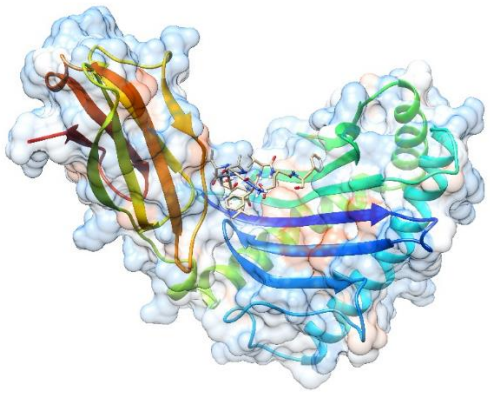

Figure 2. Docking simulation between HLA-B*15:02 and FAVTLKVEF

Before vaccine design, allergenicity prediction is a key step, because most vaccines transfer the immune response to allergic reaction by initiating immunoglobulin E and Type II T-helper cells. In that case, we predicted the allergenicity of the selected epitope through the AllergenFP1.0 server. The server identified the epitopes AYAKLGLSI, YAKLGLSII, AKDKAYAKL, VEFKTGKLL, FAVTLKVEF, LLAAKDKAY, and 11 more epitopes as non-allergens and the rest of the other epitopes as potential allergens to humans (**Table 1**). However, the results were found to be the same for the rest of the other epitopes in both AllergenFP 1.0 and AllerTOP v. 2.0 servers (data not shown). The predicted three-dimensional structures of the T-cell epitopes were deposited to the Protein Model Database (PMDB). As a result, the database provided a unique accession number for each of the submissions of the epitope. The structure of the epitopes could be identified by using the accession numbers PM0081248 and PM0081251 for T-cell epitopes YAKLGLSII and FAVTLKVEF, respectively, in the PMDB database.

3.4. Docking simulation study

The molecular docking simulation study anticipated nine (9) potential binding affinities between HLA-B*15:02 and epitope YAKLGLSII. At the same time, it was also found in the interaction between HLA-B*15:02 and epitope FAVTLKVEF. The finest model was selected by higher binding affinity. The best interacting model of epitope YAKLGLSII and the HLA molecule contained a binding affinity of −8.0 kcal/mol. On the other hand, the epitope FAVTLKVEF and HLA molecules contained a binding affinity of −8.4 kcal/mol. The epitope position in the cartoon and surface structure and the binding

interactions is shown in Figure 3. The epitope and the HLA molecule interacted through hydrogen bonds, and we found nine hydrogen bonds after the docking simulation between HLA-B*15:02 and the epitope FAVTLKVEF.

In this study, the amino acid residues, namely, Tyr-7, Tyr-9, Tyr-99, Tyr-159, Trp-147, Thr-69, and Thr-73, acted as the hydrogen bond donors or acceptors with the epitope residues (**Figure 2**). Herein, eight hydrogen bonds, 89 no bonded contacts, and the interacting residues, Thr-143, Ser-77, Gln-155, Tyr-9, Asn-70, and Tyr-59, were found for HLA-B*15:02 and epitope YAKLGLSII. The cartoon and surface structures were prepared and visualized by the PyMOL molecular graphics system. In the control section, the docking between HLA-B*15:02 and the influenza NP418 epitope (LPFERATVM) from the 2009 strain was performed. The best affinity or binding energy between these two molecules was found −8.3 kcal/mol, which is very close to the sample. The amino acid residues, namely, Tyr-7, Tyr-9, Tyr-99, Tyr-159, Tyr-171, Ser-77, and Gln-155, are involved in the interaction with the epitope residues (**Figure 2**). So, the results revealed that Tyr-7, Tyr-9, Tyr-99, and Tyr-159 are conserved residues in the docking simulation of the sample (FAVTLKVEF) and control and ensuring the accuracy of the binding affinity. Besides, Tyr-9, Ser-77, and Gln-155 residues were found as common residues in the docking simulation of the sample YAKLGLSII (binding affinity of −8.4 kcal/mol) and control which also ensures the accuracy and the acceptance of the docking simulation. The epitope (peptide) binding to the MHC class II HLA-DR molecule is important for the proper functioning of a vaccine candidate. So, we predicted the alleles for our selected epitope and performed a docking study between those molecules. Herein, we found the best binding affinity of −7.9 kcal/mol for the docking study among the predicted nine binding affinity results. This result indicates a good affinity between HLA-DR and the epitope YAKLGLSII (**Figure 3**).

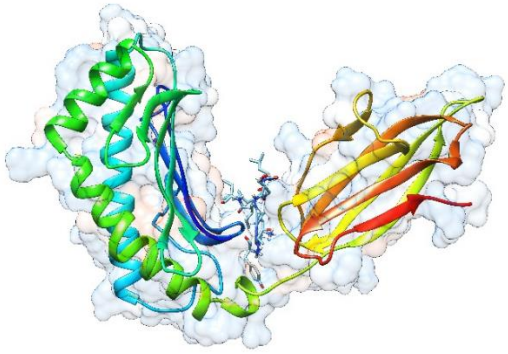

Figure 3. Docking simulation between HLA-B*15:02 and YAKLGLSII

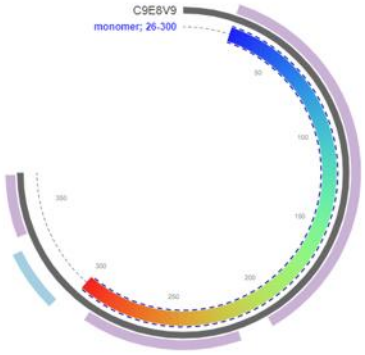

Figure 4. Protein structure

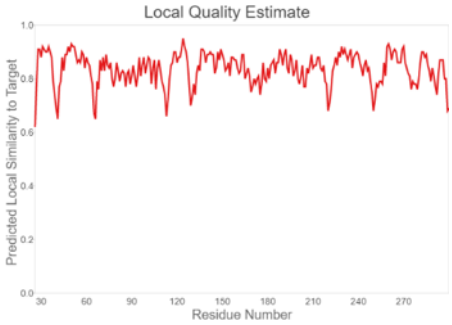

Figure 5. Local Quality Estimation

During the analysis of the interaction between HLA-DR and YAKLGLSII, we found five hydrogen bonds and 125 nonbonded contacts in the combined structure. However, the amino acid residues, namely, Ser-53, Gln-9, Asn-62 (2 times), and His-13, are involved in the interaction with the epitope residues.

3.5. Analysis of linear and conformational B-cell epitopes

A B-cell epitope is a precise portion of an antigen recognized by either a specific B-cell receptor or the elicited antibody in a humoral response. The B-cell epitopes are divided into two main categories: continuous or linear B-cell epitopes and discontinuous or conformational B-cell epitopes. It has been found that most of the B-cell epitopes are conformational epitopes, and the amount of this epitope is more than 90%. The prediction of B-cell epitopes from the antigenic protein is the major step of epitope-based vaccine design. Thus, to get essential B-cell epitope candidates in the polyprotein of the corona, we performed an *in silico* analysis through the web servers BepiPred-2.0, Chou & Fasman Beta-Turn, Emini Surface Accessibility, Karplus & Schulz Flexibility, Kolaskar & Tongaonkar, Parker Hydrophilicity prediction results.

3.6. SWISS-MODEL results

SWISS-MODEL (http://swissmodel.expasy.org) is a server for automated comparative modeling of three-dimensional (3D) protein structures. It pioneered the field of automated modeling starting in 1993 and is the most widely used free web-based automated modeling facility today. SWISS-MODEL is a web-based integrated service dedicated to protein structure homology modeling. Here, we predict structure, local quality estimation, and quality comparison in **Figures 4, 5,** and **6**. It guides the user in building protein homology models at different levels of complexity. ERRAT is used to analyze the statistics of non-bonded interactions between different atom types. Analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus the position of a 9-residue sliding window, calculated by comparison with statistics from highly refined structures (**Figure 7**).

3.7. Ramachandran plot results

A Ramachandran plot (also known as a Rama plot, a Ramachandran diagram or a [φ,ψ] plot), originally developed in 1963 by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan,[1] is a way to visualize energetically allowed regions for backbone dihedral angles ψ against φ of amino acid residues in protein structure. The Ramachandran plot is a plot of the torsional angles - phi (φ) and psi (ψ) - of the residues (amino acids) contained in a peptide. The figure on the left illustrates the definition of the φ and ψ backbone dihedral angles [2] (called φ and φ' by Ramachandran). The ω angle at the peptide bond is normally 180° since the partial-double-bond character keeps the peptide planar. A Ramachandran plot can be used in two somewhat different ways (**Figure 8**). One is to show in theory which values, or conformations, of the ψ and φ angles, are possible for an amino-acid residue in a protein (as at top right). A second is to show the empirical distribution of data points observed in a single structure (as at right, here) in usage for structure validation, or else in a database of many structures. Either case is usually shown against outlines for the theoretically favored regions.

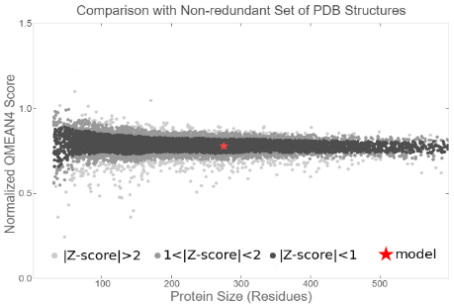

Figure 6. Quality Comparison

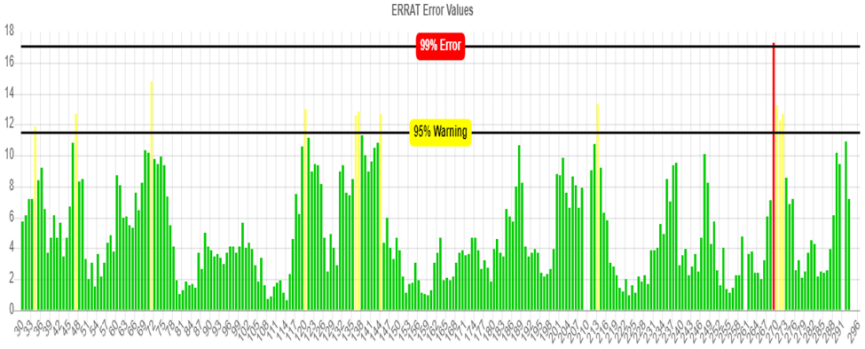

Figure 7. ERRAT Score

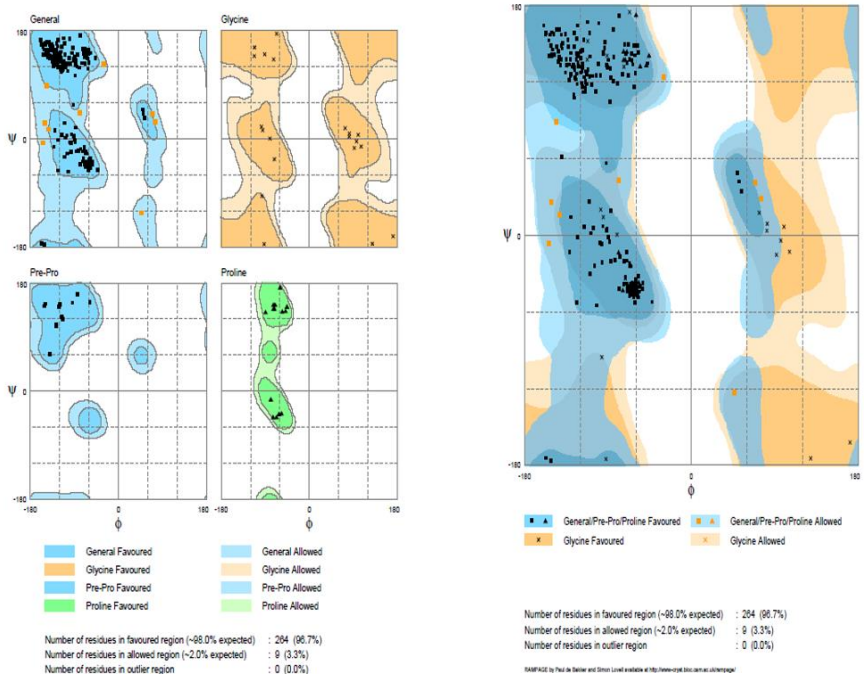

Figure 8. Ramachandran plot diagram visualizes energetically allowed regions for backbone dihedral angles ψ against φ of amino acid residues in protein structure.

4. Discussion

In recent years, many diseases have emerged due to the occurrence of several outbreaks through the different types of newer viruses. So, vaccine development against these emerging diseases within a short time is very crucial to protect people from the rising viral attacks. Vaccines are pharmacological products that can provide the finest cost-benefit ratio in the prevention or treatment of diseases. However, an effective vaccine development and production are costly and can take years to be completed. So, researchers have tried for many years to minimize the cost and time for the development of

vaccines. At this time, there are different strategies available for the design and development of effective and safe next-generation vaccines based on bioinformatics approaches. The next-generation sequencing and progressive genomics and proteomics technologies have brought about a great change in computational immunology. However, the advancement of newer immunoinformatics tools has made a broader way in developing vaccine or vaccine candidates through the satisfactory understanding of the immune response of the human body against an organism within a short time After recovery from an acute viral or microbial infection, a person generally achieves long-term or even lifelong immunity to the same viral disease or pathogen In recent trends, the surface glycoprotein (GP) is considered as the initial focus for the development of epitope-based peptide vaccine, as they are involved in the interaction between cell receptor and virus particle, thus playing a significant role in the pathogenesis of the disease. To our knowledge, the subunit or epitope/peptide-based vaccine approaches have not been used yet for the prediction of a potential vaccine against this neglected emerging Coronavirus.

However, the virus-like particle assay has been developed for this OROV that could be used as a tool to develop a potential recombinant vaccine. The population coverage information of the epitopes and their respective HLA alleles are crucial to being the potential vaccine candidates. The computational prediction of this HLA allele is much easier and cost-effective than the experimental one. Therefore, in this study, different types of immunoinformatics and molecular simulation tools were used to find the potential vaccine candidates present in the Corona polyprotein. We identified the potential T-cell epitopes from the most antigenic polyprotein, as they play a key role in the creation of a defensive immune response against different pathogenic infections. Various successful studies have been performed for the epitope-based peptide vaccine design against West Nile virus, Zika virus, dengue virus, Chikungunya virus, Rift Valley fever virus, shigellosis, and so on. Primarily, we identified 23 epitopes in total using 12 supertypes (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62) and selected 6 epitopes based on the highest combined score. Finally, we identified 2 epitopes by maintaining some critical criteria such as antigenicity, immunogenicity, and most importantly, conservancy scores of the epitopes (**Table 1**). We gave priority to the higher conservancy score of the epitopes because higher conservancy ensures extensive protection across numerous strains or even species. In this study, the results of the 6 epitope conservancy analyses revealed that the epitopes AYAKLGLSI, YAKLGLSII, AKDKAYAKL, VEFKTGKLL, FAVTLKVEF, LLAAKDKAY are 100.0% conserved. These epitopes could be considered as potential vaccine candidates for vaccine design against the Coronavirus. The immunogenicity score of the epitope can be a good criterion for the selection of the best epitope. The higher score describes a greater probability of eliciting an effective immune response. Herein, among the 6 epitopes, the epitopes FAVTLKVEF (-0.068) and YAKLGLSII (-0.12046) are less immunogenic than the epitopes AKDKAYAKL (-0.26816), VEFKTGKLL (-0.29778), and AYAKLGLSI (-0.28922) (**Table 1**). However, the epitope YAKLGLSII could be the most important in comparison with the rest of the other epitopes, because it is not allergenic (as predicted by both AllergenFP 1.0 and AllerTOP v. 2.0 servers) and nontoxic to the host. At present, researchers are using different immunoinformatics approaches for the prediction of high-potential CD4+ T-cell epitopes and CD8+ T-cell epitopes that can bind to different MHC-I and MHC-II molecules. So, the analysis revealed that immunoinformatics or in silico approaches could make a great contribution to vaccine lead identification for experimental study and/or the design of an effective vaccine. To get more population coverage, the T-cell epitope should bind with more MHC HLA alleles. So, we took the selected 2 epitopes and their respective MHC-I-binding HLA alleles for significant population coverage. The results revealed that the selected epitopes and their alleles have ideal population coverage in different geographic regions around the world. The highest population coverage was found in Italy (98.60%), which was closely followed by France, Germany, the United States, England, Finland, Saudi Arabia, and Brazil with population coverage of 97.90%, 97.94%, 97.83%, 97.73%, 96.14%, 94.73%, and 94.42%, respectively. The lowest population coverage was found in Peru Mestizo (1.99%). The 3D structure of the epitopes YAKLGLSII and FAVTLKVEF were designed for docking purposes and compared with a control (HLA-B*15:02 and influenza NP418 epitope) to measure the docking efficiency with a specific HLA allele HLA-B*15:02. The docking results showed almost similar binding affinity or energy for both the control (−8.3 kcal/mol) and the sample (−8.4 kcal/mol and −8.0 kcal/mol for the epitopes “YAKLGLSII” and “FAVTLKVEF,” respectively) which designates the satisfactory accuracy of the predicted epitopes to form an interaction with the MHC-I allele. However, another docking was performed between the MHC-II allele and the selected epitope “YAKLGLSII” which also exhibits satisfactory binding affinity. So, the results indicate the epitope “YAKLGLSII” as a novel vaccine candidate for coronavirus. The evaluated model was submitted to the PMDB database for further research in the future. To support future drug discovery and drug design, ligand-binding pockets and hydrophobicity analyses were performed. Peptide half-life prediction is one of the key challenges in successful peptide vaccine design, as its bioavailability depends on the extent of the half-life. So it is vital to identify the optimal half-

life of the peptide for ensuring its optimal function post-vaccination. Recently, a database called "PEPlife" has been developed as a repository of the half-life of peptides; however, it is customary to test and measure the peptide half-life to ensure optimal protection post vaccination. In general, protein-based vaccines have limited immunogenicity and reactogenicity compared to live-attenuated and inactivated vaccines, although protein-based vaccines have greater advantages concerning safety and cost-effectiveness.

5. Conclusion

This study provided valuable insights for the identification of vaccine candidates and ligand-binding pockets against the neglected and emerging Coronavirus infection through immunoinformatic as well as Bioinformatics approaches. Based on this strategy, the T-cell and B-cell epitopes in the polyprotein of coronavirus were mapped and selected as putative coronavirus vaccine targets. The predicted epitopes exhibited T-cell and B-cell selectivity, higher conservancy, no allergenicity, nontoxicity, higher population coverage, and significant interaction with the MHC class I allele with good affinity. Nonetheless, these findings will provide primary data for molecular docking and epitope-based peptide vaccine research in the future.

Acknowledgements: We would like to thank our department, Md. Chayan Ali and other team members sincerely for giving financial and logistic support.

Author contributions: Sabbir Ahmed and Md. Zakariya Alam Bappi contributed equally to this work.

Funding: This research did not receive any funds.

Data availability: Data is included in the manuscript.

Conflicts of interest: The authors declare no conflict of interest.

References

[1] van der H. L *et al.*, "Identification of a new human coronavirus," *Nat. Med.*, vol. 10, no. 4, pp. 368–373, Apr. 2004, doi: 10.1038/NM1024.

[2] J. Cui, F. Li, and Z.-L. Shi, "Origin and evolution of pathogenic coronaviruses," *Nat. Rev. Microbiol. 2018 173*, vol. 17, no. 3, pp. 181–192, Dec. 2018, doi: 10.1038/s41579-018-0118-9.

[3] S. R. Weiss and S. Navas-Martin, "Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus," *Microbiol. Mol. Biol. Rev.*, vol. 69, no. 4, pp. 635–664, Dec. 2005, doi: 10.1128/MMBR.69.4.635-664.2005.

[4] A. R. Oany, A.-A. Emran, and T. P. Jyoti, "Design of an epitope-based peptide vaccine against spike protein of humancoronavirus: an in silico approach," *Drug Des. Devel. Ther.*, vol. 8, p. 1139, Aug. 2014, doi: 10.2147/DDDT.S67861.

[5] "COVID Update from the Coronavirus - Worldometer," 2021. https://www.worldometers.info/coronavirus/ (accessed Sep. 29, 2021).

[6] R. Sharmin and A. B. M. M. K. Islam, "A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design," *BMC Bioinforma. 2014 151*, vol. 15, no. 1, pp. 1–10, May 2014, doi: 10.1186/1471-2105-15-161.

[7] F. Bojin, O. Gavriliuc, M.-B. Margineanu, and V. Paunescu, "Design of an Epitope-Based Synthetic Long Peptide Vaccine to Counteract the Novel China Coronavirus (2019-nCoV)." Preprints, Feb. 08, 2020, Accessed: Sep. 29, 2021. [Online]. Available: https://www.preprints.org/manuscript/202002.0102/v1.

[8] M. Cascella, M. Rajnik, A. Aleem, S. C. Dulebohn, and R. Di Napoli, "Features, Evaluation, and Treatment of Coronavirus (COVID-19)," *StatPearls*, Jul. 2021, Accessed: Sep. 29, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK554776/.

[9] M. Rajnik, M. Cascella, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, "Features, Evaluation, and Treatment of Coronavirus (COVID-19)." Mar. 01, 2021, Accessed: Sep. 29, 2021. [Online]. Available: https://apps.dtic.mil/sti/citations/AD1127230.

[10] L. CC, S. TP, K. WC, T. HJ, and H. PR, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges," *Int. J. Antimicrob. Agents*, vol. 55, no. 3, Mar. 2020, doi: 10.1016/J.IJANTIMICAG.2020.105924.

[11] S. Bibi *et al.*, "In silico analysis of epitope-based vaccine candidate against tuberculosis using reverse vaccinology," *Sci. Reports 2021 111*, vol. 11, no. 1, pp. 1–16, Jan. 2021, doi: 10.1038/s41598-020-80899-6.
[12] K. MA, H. MU, R.-U.-Z. SM, and M. MN, "Epitope-based peptide vaccine design and target site depiction against Ebola viruses: an immunoinformatics study," *Scand. J. Immunol.*, vol. 82, no. 1, pp. 25–34, Jul. 2015, doi: 10.1111/SJI.12302.
[13] U. K. Adhikari, M. Tayebi, and M. Mizanur Rahman, "Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging oropouche virus," *J. Immunol. Res.*, vol. 2018, 2018, doi: 10.1155/2018/6718083.
[14] O. H. Ahmed, A. A. Mohammed, S. Obi, A. Albagi, K. Ali, and A. Mohammed, "Immunoinformatic Approach for Epitope-Based Peptide Vaccine against Lagos Rabies Virus Glycoprotein G Human Genetics View project Vaccine design View project," *Immunome Res*, vol. 13, p. 3, 2017, doi: 10.4172/1745-7580.1000137.
[15] R. Islam, M. S. Sakib, and A. Zaman, "A computational assay to design an epitope-based peptide vaccine against chikungunya virus," *http://dx.doi.org/10.2217/fvl.12.95*, vol. 7, no. 10, pp. 1029–1042, Oct. 2012, doi: 10.2217/FVL.12.95.
[16] M. A. Hasan, M. Hossain, and M. J. Alam, "A computational assay to design an epitope-based peptide vaccine against saint louis encephalitis virus," *Bioinform. Biol. Insights*, vol. 7, pp. 347–355, Nov. 2013, doi: 10.4137/BBI.S13402.
[17] A. Alam, S. Ali, S. Ahamad, M. Z. Malik, and R. Ishrat, "From ZikV genome to vaccine: in silico approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein," *Immunology*, vol. 149, no. 4, pp. 386–399, Dec. 2016, doi: 10.1111/IMM.12656.
[18] M. Kamthania, S. Srivastava, M. Desai, A. Jain, A. Shrivastav, and D. K. Sharma, "Immunoinformatics Approach to Design T-cell Epitope-Based Vaccine Against Hendra Virus," *Int. J. Pept. Res. Ther. 2019 254*, vol. 25, no. 4, pp. 1627–1637, Jan. 2019, doi: 10.1007/S10989-018-09805-Z.
[19] H. Zahroh, A. Ma'rup, U. Sumo Friend Tambunan, and A. Parikesit, "Immunoinformatics Approach in Designing Epitope-based Vaccine Against Meningitis-inducing Bacteria (Streptococcus pneumoniae, Neisseria meningitidis, and Haemophilus influenzae Type b)," *Drug Target Insights*, vol. 2016, p. 10, 2016, doi: 10.4137/DTI.s38458.
[20] P. Oyarzun and B. Kobe, "Computer-aided design of T-cell epitope-based vaccines: addressing population coverage," *Int. J. Immunogenet.*, vol. 42, no. 5, pp. 313–321, Oct. 2015, doi: 10.1111/IJI.12214.
[21] J. L. Sanchez-Trincado, M. Gomez-Perosanz, and P. A. Reche, "Fundamentals and Methods for T- and B-Cell Epitope Prediction," *J. Immunol. Res.*, vol. 2017, 2017, doi: 10.1155/2017/2680160.
[22] Y. EL-Manzalawy and V. Honavar, "Recent advances in B-cell epitope prediction methods," *Immunome Res. 2010 62*, vol. 6, no. 2, pp. 1–9, Nov. 2010, doi: 10.1186/1745-7580-6-S2-S2.
[23] Y. Jin, R. E. Goodman, A. O. Tetteh, M. Lu, and L. Tripathi, "Bioinformatics analysis to assess potential risks of allergenicity and toxicity of HRAP and PFLP proteins in genetically modified bananas resistant to Xanthomonas wilt disease," *Food Chem. Toxicol.*, vol. 109, pp. 81–89, Nov. 2017, doi: 10.1016/J.FCT.2017.08.024.
[24] M. Moballegh Naseri, S. Shams, M. Moballegh Naseri, and B. Bakhshi, "In silico analysis of epitope-based CadF vaccine design against Campylobacter jejuni," *BMC Res. Notes 2020 131*, vol. 13, no. 1, pp. 1–6, Nov. 2020, doi: 10.1186/S13104-020-05364-Z.
[25] A. S. Kolaskar and U. Kulkarni-Kale, "Prediction of Three-Dimensional Structure and Mapping of Conformational Epitopes of Envelope Glycoprotein of Japanese Encephalitis Virus," *Virology*, vol. 261, no. 1, pp. 31–42, Aug. 1999, doi: 10.1006/VIRO.1999.9859.
[26] A. Radwan and G. M. Mahrous, "Docking studies and molecular dynamics simulations of the binding characteristics of waldiomycin and its methyl ester analog to Staphylococcus aureus histidine kinase," *PLoS One*, vol. 15, no. 6, p. e0234215, Jun. 2020, doi: 10.1371/JOURNAL.PONE.0234215.
[27] U. Raj, H. Kumar, and P. K. Varadwaj, "Molecular docking and dynamics simulation study of flavonoids as BET bromodomain inhibitors," *http://dx.doi.org/10.1080/07391102.2016.1217276*, vol. 35, no. 11, pp. 2351–2362, Aug. 2016, doi: 10.1080/07391102.2016.1217276.
[28] M. C. Ali, S. I. Jahan, M. S. Khatun, R. Das, M. M. Rahman, and R. Dash, "In Silico Epitope-Based Peptide Vaccine Design against Invasive Non Typhoidal Salmonella (iNTS) Through Immunoinformatic Approaches," Feb. 2020, doi: 10.20944/PREPRINTS202002.0383.V1.