Molecular Diversity (2025) 29:4273–4291
https://doi.org/10.1007/s11030-024-11033-1

ORIGINAL ARTICLE

Identification of potential antigenic proteins and epitopes for the development of a monkeypox virus vaccine: an in silico approach

Emre Aktaş[1] · Osman Uğur Sezerman[2] · Murat Özer[3] · Kevser Kübra Kırboğa[4] · Ahmet Efe Köseoğlu[5] · Nehir Özdemir Özgentürk[1]

Received: 22 August 2024 / Accepted: 25 October 2024 / Published online: 15 November 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Virus assembly, budding, or surface proteins play important roles such as viral attachment to cells, fusion, and entry into cells. The present study aimed to identify potential antigenic proteins and epitopes that could be used to develop a vaccine or diagnostic assay against the Monkeypox virus (MPXV) which may cause a potential epidemic. To do this, 39 MPXV proteins (including assembly, budding, and surface proteins) were analyzed using an in silico approach. Of these 39 proteins, the F5L virus protein was found to be the best vaccine candidate due to its signal peptide properties, negative GRAVY value, low transmembrane helix content, moderate aliphatic index, large molecular weight, long-estimated half-life, beta wrap motifs, and being stable, soluble, and containing non-allergic features. Moreover, the F5L protein exhibited alpha-helical secondary structures, making it a potential "structural antigen" recognized by antibodies. The other viral protein candidates were A9 and A43, but A9 lacked beta wrap motifs, while A43 had a positive GRAVY value and was insoluble. These two proteins were not as suitable candidates as the F5L protein. The KRVNISLTCL epitope from the F5L protein demonstrated the highest antigen score (2.4684) for MHC-I, while the GRFGYVPYVGYKCI epitope from the A9 protein exhibited the highest antigenicity (1.754) for MHC-II. Both epitopes met the criteria for high antigenicity, non-toxicity, solubility, non-allergenicity, and the presence of cleavage sites. Molecular docking and dynamics (MD) simulations further validated their potential, revealing stable and energetically favorable interactions with MHC molecules. The immunogenicity assessment showed that GRFGYVPYVGYKCI could strongly induce immune responses through both IFN-γ and IL-4 pathways, suggesting its capacity to provoke a balanced Th1 and Th2 response. In contrast, KRVNISLTCL exhibited limited immunostimulatory potential. Overall, these findings lay the groundwork for future vaccine development, indicating that F5L, particularly the GRFGYVPYVGYKCI epitope, may serve as an effective candidate for peptide-based vaccine design against MPXV.

Keywords Monkeypox virus (MPXV) · Peptide-based vaccine candidates · Monkeypox virus proteins · Bioinformatics analyses

✉ Emre Aktaş
emrea@yildiz.edu.tr

[1] Faculty of Art and Science, Molecular Biology and Genetics, Yıldız Technical University, Istanbul, Turkey

[2] School of Medicine, Department of Basic Sciences, Biostatistics and Medical Informatics, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[3] Department of Chemistry, Faculty of Science and Arts, University of Afyon Kocatepe, Afyonkarahisar, Turkey

[4] Faculty of Engineering, Bioengineering Department, Bilecik Seyh Edebali University, Bilecik 11100, Turkey

[5] Experimental Eye Research Institute, Ruhr-University Bochum, Bochum, Germany

Introduction

Monkeypox virus (MPXV), a zoonotic orthopoxvirus, was first identified in laboratory monkeys in 1958 and in humans in 1970 in the Democratic Republic of Congo (DRC) [1]. Unlike SARS-CoV-2, MPXV had been known to cause sporadic outbreaks in Central and West Africa for decades, but it gained global attention in 2022 due to a rapid increase in cases outside endemic regions (European Centre for Disease Prevention and Control:ECDC) [2, 3]. By 2022, MPXV had two well-defined clades: the West African (WA) and Central African/Congo Basin (CA) clades [1, 3]. The WA clade, which is associated with lower mortality, was responsible for

the global spread, while the CA clade, endemic to Central Africa, is known for higher virulence and mortality rates [4, 5]. The 2022 outbreak highlighted MPXV's capacity for human-to-human transmission, particularly through close physical contact, leading to widespread infection across previously unaffected countries. By October 25, 2022, 75,885 confirmed cases had been reported globally, including 74,994 cases from 101 non-endemic countries (World Health Organization (WHO, ECDC). This marked a significant shift from the historical pattern, where cases were mostly confined to African regions [5]. As of 2024, MPXV cases have reached more than 90 countries, with the virus showing a continued ability to infect through contact with infected individuals or contaminated materials (ECDC) [6]. The World Health Organization has urged for continued vigilance and monitoring of the virus (WHO).

MPXV is a double-stranded DNA virus with a genome size of approximately 197 kb, encoding around 190 proteins. These proteins play critical roles in viral replication, immune evasion, and host cell interaction, making them essential targets for research, especially in the development of vaccines and antiviral therapies [4]. As the global outbreak continues, understanding which clade is circulating is crucial for tailoring vaccine strategies, as vaccines must target the most prevalent strains to be effective. Research into the 39 core proteins of MPXV, especially in the context of the currently circulating WA clade, remains a priority for developing effective interventions [1, 4, 6]. The proteins we mentioned play crucial roles in the monkeypox virus's life cycle, particularly in viral entry, fusion, and propagation processes (https://viralzone.expasy.org/9959). For example, among these 39 key proteins, K2L, F13L, and A36 have been identified as vital for the virus's assembly, budding, and surface interactions. These proteins are essential for viral replication and immune evasion, making them prime targets for vaccine development [6]. MPXV's capacity for human-to-human transmission, particularly through close physical contact, leads to widespread infection across previously unaffected countries. The virus spends its life cycle in the cytoplasm of infected cells, and the incubation period is typically 6–13 days, though it can vary between 5–21 days [7–9]. It is reported that an increase in the number of deaths was observed, and 22 deaths were determined [10–12]. For persons infected by this virus, it is recommended to follow isolation and infection control procedures. Also, symptoms of people infected by this virus, including rash, fever, chills, headache, muscle aches, back pain, and fatigue that can progress to exhaustion, have been reported [13]. MPXV is at risk of being transmitted from various animal species (monkeys, squirrels, and mice) to humans by direct contact or by eating these animals [4, 14].

The basic idea for all types of vaccination is that they produce an effective immune response faster than the infecting virus. However, conventional vaccines are developed using a variety of biochemical assays and elicit antibodies in vaccinated individuals. These vaccines require costly, time-consuming, and potentially allergenic in vitro cultures of harmful microbes, which can pose serious safety concerns [15–19]. In contrast, the reverse vaccinology (RV) approach plays an essential role in the development of vaccines by using computational (in silico, bioinformatics) analyses. This approach has several advantages, including screening pathogen genomes, shortening the time of candidate peptide discovery, and reducing the cost of the entire vaccine development process [17, 20, 21]. A similar approach was applied for the development of a vaccine for COVID-19 by Can et al.

Before starting experiments in the laboratory, RV strategy-based bioinformatics analyses provide preliminary results, such as the prediction of physicochemical properties, solubility, antigenicity, allergenicity, toxicity, as well as specific T and B cell epitopes of vaccine candidate proteins of a pathogen [2, 22–25]. Also, the selected peptides underwent further in silico analyses to assess their immunogenic potential [26, 27]. Adopting an RV-based approach, this study aims to examine 39 MPXV proteins and predict potential peptide candidates for vaccine development against this virus. In the first step, we predicted the physico-chemical parameters (including molecular weight, theoretical PI, total number of negatively charged residues (Asp + Glu), total number of positively charged residues (Arg + Lys), the estimated half-life (hour), the instability index, aliphatic index, GRAVY score, secondary structure content, subcellular localization, transmembrane helices, antigenicity, and existence of signal peptides for these proteins. In the second step, we evaluated predicted epitopes derived from these proteins for their toxicity, hydrophobicity, allergenicity, solubility, antigenicity, and number of cleavage sites. Peptides with the best scores were selected for both molecular dynamics simulations and docking with receptors of MHC-I/II alleles.

In the first step, we predicted the physico-chemical parameters of the selected 39 MPXV proteins, including molecular weight, theoretical pI, the total number of negatively charged residues (Asp + Glu), the total number of positively charged residues (Arg + Lys), estimated half-life, instability index, aliphatic index, GRAVY score, secondary structure content, subcellular localization, transmembrane helices, antigenicity, and the existence of signal peptides. In the second step, predicted epitopes derived from these proteins were evaluated for their toxicity, hydrophobicity, allergenicity, solubility, antigenicity, and number of cleavage sites. Peptides with the best scores were selected for molecular dynamics simulations (MDs) and docking with receptors of MHC-I/II alleles. Following docking and molecular dynamics simulations, the selected peptides underwent further in silico analyses to assess their immunogenic potential. The IFNepitope server predicted the peptides' ability

to induce interferon-gamma (IFN-γ) production.Next, the IL4Pred tool, employing a support vector machine (SVM) algorithm, assessed the potential for interleukin-4 (IL-4) induction, with results presented as SVM scores. Finally, the CTLpred server predicted cytotoxic T lymphocyte (CTL) epitopes using an artificial neural network (ANN) model. These analyses aimed to identify peptides with strong immunostimulatory properties for potential use in vaccine development.

Material—methods

Selecting viral proteins

Full genome or FASTA format of K2L (AAL40486.1), F5L (AAL40512), F9L (AAL40516), F13L (AAL40503.1), E8 (AAL40563.1), E10R (AAL40565.1), I2L (AAL40521), I5L (AAL40524), G3 (AAL40530), G9 (AAL40536), L1 (AAL40543), L5 (AAL40547), J1 (AAL40459), H2 (AAL40550), D8 (AAL40469), D13 (AAL40474), A9 (AAL40577), A13 (AAL40581.1), A14 (AAL40582), A15 (AAL40583), A15.5, A16 (133aa: AAL40584.1), A16 (340aa), A16 (377aa), A36 (228aa: AAL40603.1), A36 (212aa), A17 (AAL40585), A21 (AAL40589), A26 (AAL40594), A27 (AAL40595), A28 (AAL40596), A30 (AAL40598), A33 (AAL40601), A34 (AAL40601), A38 (AAL40606), A43 (AAL40611), A56 (AAL40621.1), and B5 (AAL40624) MPXV proteins in the full genome or FASTA format were obtained from GenBank (https://www.genome.jp/dbget-bin/www_bget?genbank-vrl:AF380138) and NCBI Virus (National Center for Biotechnology Information) https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Protein&VirusLineage_ss=Monkeypox%20virus,%20taxid:10244&Completeness_s=complete&SourceDB_s=GenBank. These proteins were chosen because all selected proteins are either assembly, budding, or surface proteins of this virus (https://viralzone.expasy.org/9959). The full genomic sequences of all 39 viral proteins were aligned and edited by the MEGA (Molecular Evolutionary Genetics Analysis) tool [28].

Prediction of physico-chemical parameters and secondary structures

The 39 selected viral proteins were investigated by ExPASy ProtParam online server for the prediction of physico-chemical parameters such as molecular weight, theoretical pI, the total number of negatively charged residues (Asp + Glu), the total number of positively charged residues (Arg + Lys), the estimated half-life (hour), the instability index, aliphatic index, GRAVY [29]. This analysis is essential to assess the protein's stability, solubility, and potential immunogenicity for vaccine development [20–22]. The cutoff value for the instability index was set at 40, where proteins with values below 40 are considered stable. Additionally, the molecular weight units are provided in Daltons (Da) as per the standard output from the ExPASy ProtParam tool [29].

SolPro predicted solubility [30], as solubility is a key factor in determining the feasibility of protein expression and formulation in vaccine design [16, 20]. Garnier-Osguthorpe-Robson IV (GOR IV) online server was used for secondary structures of viral proteins [31], which helps identify structural features crucial for selecting effective epitopes in vaccine candidates [16, 20, 22, 31].

Prediction of antigenicity, allergenicity and toxicity

The antigenicity of all 39 selected viral proteins and predicted epitopes was analyzed by the Vaxijen v2.0 online server using a threshold value of 0.4 [32]. This step is crucial as antigenicity prediction helps identify potential candidates that can elicit a strong immune response, a key factor in effective vaccine design [17, 20]. The allergenicity of the predicted epitopes was examined by Algpred, based on the similarity of the known epitope with any protein region [33]. This is important to ensure that the selected epitopes do not trigger allergic reactions in the host, enhancing the vaccine's safety profile [24]. Additionally, the allergenicity potential of viral proteins and epitopes was predicted by the AllerCatPro 2.0 web server [34], providing further validation to prevent allergenic responses. The toxicity of discovered peptides was predicted by ToxinPred [35], as identifying non-toxic peptides is vital to minimize adverse effects and ensure the overall safety of the vaccine candidates [24, 28].

Prediction of subcellular localisation and number of transmembrane helices

The subcellular localization of the viral proteins in infected host cells was predicted using Virus-mPLoc [36]. This prediction is important as it provides insights into where the viral proteins are likely to interact within the host cells, aiding in the identification of potential targets for vaccine development [17, 20, 21, 24]. The number of transmembrane helices was predicted using the TMHMM Server v.2.0 [37], as the presence of transmembrane regions can influence the protein's accessibility and immunogenicity, which are critical factors in the selection of suitable vaccine candidates [20, 24, 28].

Prediction of signal peptides in selected proteins

All selected monkeypox viral proteins were analyzed by the Signal-BLAST bioinformatic tool to predict signal peptides in the selected protein sequences [38]. Predicting

signal peptides is essential because they play a crucial role in directing proteins to their proper cellular locations, influencing their interaction with the host's immune system, and potentially enhancing the immunogenicity of the vaccine candidates [22].

Prediction of similarity with host proteome and prediction of BetaWrap motifs

The signal peptide for all proteins was predicted using BlastP (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) based on their similarity to the host proteome, with Homo sapiens selected as the host organism. Virus proteins that showed positive results in this analysis were excluded from further analysis.The prediction of BetaWrap motifs for the viral proteins was performed by BetaWrap online server [39]. BetaWrap motifs provide structural integrity and enhance the proper folding and presentation of epitopes to the immune system, making them important for effective antigen recognition. Proteins without these motifs may exhibit reduced stability, which could compromise their efficacy in eliciting a robust immune response [16, 40]. Therefore, only proteins containing BetaWrap motifs were selected for further analysis.

Prediction of B cell epitopes

Linear B cell epitopes for the viral proteins were predicted using Bcepred [41] and Bepipred Linear Epitope Prediction 2.0 online servers running under IEDB (the Immune Epitope Database) [42]. These tools were employed because identifying linear B cell epitopes is crucial for understanding which protein regions are likely to be recognized by antibodies. Selecting epitopes with scores greater than the 0.5 threshold helps ensure that the identified epitopes have a high potential to stimulate an immune response, which is vital for designing effective vaccine candidates [17, 20, 21, 24, 28].

Prediction of MHC-I and MHC-II epitopes

To predict MHC-I epitopes, we selected twelve commonly used MHC-I alleles: HLA-A-01.01, HLA-A-02.01, HLA-A-03.01, HLA-A-24.02, HLA-A-26.01, HLA-B-07.02, HLA-B-08.01, HLA-B-27.05, HLA-B-39.01, HLA-B-40.01, HLA-B-58.01, and HLA-B-15.01. For MHC-II epitope prediction, we selected seven different MHC-II alleles: DRB 1* 03.01, DRB 1* 07.01, DRB 1* 15.01, DRB 3* 01.01, DRB 3* 02.02, DRB 4* 01.01, and DRB 5* 01.01, which were obtained from https://www.ebi.ac.uk/ipd/mhc/ and tools.iedb.org/population/ We used the IEDB online server (https://www.iedb.org/) for predicting the alleles [42]. Additionally, the selection of epitopes was based on their antigenicity and IC50 values. The antigenicity of each epitope was assessed using a defined threshold, where scores above 0.5 were considered indicative of probable antigens. Furthermore, IC50 values were utilized as a critical parameter to evaluate the binding affinity of each epitope, with values below 50 nM indicating strong binding. These criteria were employed to ensure that the selected epitopes possess a high potential to elicit an effective immune response, thus facilitating the development of a robust vaccine candidate [20, 21]. In selecting the 12 MHC-I and 7 MHC-II alleles, we considered both their prevalence and immunological relevance across diverse populations, ensuring broad population coverage [42]. Additionally, we prioritized alleles that are most effective in generating strong immune responses, making them ideal candidates for vaccine design [42–44]. This approach allowed us to focus on epitopes that are not only widely recognized but also capable of eliciting potent immune reactions, optimizing the immunogenic potential of the predicted epitopes [42–44]. Expanding the number of alleles beyond this selection would have increased computational demands without significantly enhancing the coverage or effectiveness of the immune response [42–45]. Thus, our selection balances population diversity with immunological efficiency, ensuring optimal peptide selection for the immune system.

Docking analysis with MHC-I and II alleles

The HDOCK server (http://hdock.phys.hust.edu.cn/) was used to dock each modeled epitope ligand to its specific MHC-I allele receptor, and the results were visualized using the UCSF Chimera 1.15 tool [46]. The MHC-I and MHC-II receptor alleles specific to each epitope were retrieved from the Protein Data Bank for the docking analyses. The alleles and their respective PDB IDs are as follows: HLA-A*0201 (PDB: 3UTQ), HLA-A*2402 (PDB: 3NFN), and HLA-B*4001 (PDB: 6IEX). These specific alleles were used to ensure accurate docking simulations, providing a detailed understanding of the interaction between the predicted epitopes and MHC molecules. Epitopes with high antigenicity value, low IC50 value, non-toxicity, non-allergenicity, and high solubility were selected for docking. Similarly, each modeled epitope ligand was docked to its specific MHC-II allele receptor using the HDOCK server and visualized on UCSF Chimera 1.15 tool [46]. Additionally, obtaining and imaging three-dimensional structures of peptides was also done with UCSF Chimera 1.15 tool [46].

Molecular dynamics simulation analysis of predicted epitope with MHC-I and II alleles

In this study, molecular dynamics simulations were conducted utilizing GROMACS software (http://www.mdtutorials.com/gmx/complex/index.html), with the AMBER

ff14SB force field applied for precise energy calculations and system interactions [47, 48]. The protein-peptide complexes were solvated in an SPC (Simple Point Charge) water model within a triclinic box, ensuring ample space for molecular movement and minimizing boundary effects. The system was neutralized using NaCl, resulting in a physiological salt concentration of 0.15 Molar to explore its impact on system stability [47, 48]. Energy minimization was performed using the Steepest Descent method over 5000 steps, ensuring that the system reached a local energy minimum and was adequately prepared for the subsequent simulation stages. Equilibration of the system was achieved through both NVT (constant Number of particles, Volume, and Temperature) and NPT (constant Number of particles, Pressure, and Temperature) ensembles, conducted at 300 K and 1.0 bar for 300 ps, to ensure a stable and physiologically relevant simulation environment [49, 50]. The molecular dynamics simulations were run for a total of 100 ns, employing the Leap frog integrator for numerical integration of the equations of motion, and a total of 5000 frames were saved throughout the simulation to provide a detailed and comprehensive representation of the system's conformational space and dynamics [47–50].

In silico analysis of immune response effects of selected peptides following docking and molecular dynamics simulations

Selected peptides after both docking and MDs, were subjected to in silico analyses to evaluate their immunogenic potential. The IFNepitope server was used to predict the ability of the peptides to induce interferon-gamma (IFN-γ) production, utilizing a cutoff score of 0.51 to classify peptides as positive or negative IFN-γ inducers [26]. To assess interleukin-4 (IL-4) induction potential, the IL4Pred tool was employed, using a support vector machine (SVM) algorithm, with results presented as SVM scores [27]. Finally, the cytotoxic T lymphocyte (CTL) epitope prediction was conducted using the CTLpred server, which uses an artificial neural network (ANN) model. Peptides with scores above the 0.51 cutoff were classified as CTL epitopes [26, 27]. These analyses aimed to identify peptides with strong immunostimulatory properties for potential use in vaccine development [26–28].

Results

Predicted physico-chemical parameters

The selected proteins had varying numbers of amino acids, ranging from 53 to 509, with A26 being the largest at ~59,179.08 Dalton molecular weight and A15.5 being the smallest at 6190.54 Dalton molecular weight (Table 1 and Supplementary 1). The physico chemical parameters of vaccine candidates derived from these proteins are summarized in the tables and the other peptides are summarized in the Supplementary materials (1–6). The full sequence of all selected proteins were analyzed, and their theoretical pI values were between 4.10 and 9.76. The estimated half-life for these proteins was more than 30 h in mammalian reticulocytes (in vitro), more than 20 h in yeast (in vivo), and more than 10 h in E. coli (in vivo) by using Expasy protparam. Out of the 39 selected proteins, according to the predicted instability index, the stable proteins were F13L, E8, I2L, I5L, G3, L1, J1, H2, D13, A9, A14, A15, A15.5, A16 (377 aa), A21, A27, A28, A30, A34, A38, A43, and A56 (Table 1 and Supplementary 1). The aliphatic index varied significantly among all proteins, ranging from 69.28 to 148.87. The grand average of hydropathicity (GRAVY) values was generally negative, with 17 of the 39 selected proteins having positive GRAVY values and the remaining 22 having negative values. A15.5 protein had the highest positive value of 1.496.

Predicted secondary structures

According to the results obtained from the selected proteins, the alpha helix content was found to be between ~0 and 52.73%, the extended strand content was between ~13.16 and 49.37%, and the random coil content was between ~35.44 and 65.57%. A43 and A15.5 were predicted to have the lowest alpha helix content (score: 0) (Table 2 and Supplementary 2).

Solubility, transmembrane helices, localization and antigenicity

Based on the solubility estimation of the 39 analyzed proteins, F5L, E10R, L5, J1, H3, A9, A15.5, A16 (133 aa), A16 (340 aa), A21, A27, A28, and A30 were determined to be soluble, while the remaining proteins were found to be insoluble (Table 3 and Supplementary 3).

The number of transmembrane helices varied from 0 to 5 among the selected proteins, with the highest number observed in the A38 protein. Upon examination of the subcellular localization estimates, it was predicted that the selected proteins were located in the host endoplasmic reticulum, host cell membrane, and host cytoplasm, with the majority being located in the host cytoplasm (Table 3 and Supplementary 3).

Antigenicity was assessed for the selected 39 proteins, and their values were predicted to be between 0.4044 and 0.9010. A30 had the highest antigenicity value with a score of 0.9010, while A36 (228 aa) had the lowest score of 0.4044 (Table 3). Non-antigenic proteins are listed in Table 3 along with their predicted scores (Table 3 and Supplementary 3).

Table 1 Physico-chemical parameter results predicted by ExPASyProtParam

Protein name	Number of amino acids	Molecular weight	Theoretical PI	Total number of negatively charged residues (Asp + Glu)	Total number of positively charged residues (Arg + Lys)	The estimated half-life (hour)			The instability index	Aliphatic index	GRAVY
						Mammalian reticulocytes (in-vitro)	Yeast (in vivo)	E.coli (in vivo)			
K2L	375	42,984.30	6.97	39	39	30	> 20	> 10	40.72 unstable	81.65	−0.199
F5L	318	36,180.22	6.06	32	28	30	> 20	> 10	41.01 unstable	96.16	−0.087
A9	100	11,289.97	8.88	7	10	30	> 20	> 10	39.13 stable	89.70	−0.020
A43	196	23,004.88	8.63	14	18	30	> 20	> 10	31.15 stable	104.39	0.211
A56	313	34,451.95	4.10	49	19	30	> 20	> 10	31.55 stable	80.93	−0.302
B5	317	35,145.87	4.67	39	23	30	> 20	> 10	41.79 unstable	77.98	−0.182

Table 2 Secondary structures results predicted by GOR IV

Protein name	Alpha helix (%)	Extended strand (%)	Random coil (%)
K2L	26.93	21.07	52.00
F5L	16.35	31.13	52.52
A9	33.00	20.00	47.00
A43	0.00	38.27	61.73
A56	6.39	35.46	58.15
B5	10.09	32.49	57.41

Signal peptide

K2L, F5L, A9, A43, A56, and B5 proteins were all predicted to have four different parameters: Sensitivity, Specificity, Balanced prediction, and Cleavage Site. G3, J1, L5, H2, H3, D13, A13, A14, A15.5, A21, A26, A27, A28, A30, A34, and A38 proteins were predicted to have cleavage sites, while no results were obtained for the other proteins (Table 4 and Supplementary 4).

Allergenicity, BetaWrap motifs and host proteome similarity

Among the analyzed proteins, A56 was predicted to exhibit allergenic properties for the IgE epitopes, but all other proteins were predicted to be non-allergenic for the Meme/Mast Motif (Table 5 and Supplementary 5). Only K2L, F5L, F13L, D13, A43, and B5 proteins were predicted to have BetaWrap Motifs (P-Value); other proteins did not have this motif. Similarities were predicted between the analyzed F13L, A38, A56, A34, and D8 viral proteins and host proteins (Table 5 and Supplementary 5).

Out of the analyzed proteins, K2L, F5L, A9, A43, A56, and B5 were predicted to have four different parameters of Sensitivity, Specificity, Balanced prediction, and Cleavage Site (Table 4). However, only two of them (K2L and B5) have positive BlastP results, and A56 viral protein has both positive allergenicity and BlastP results (Table 5). Based on these results, we have decided that the viral proteins F5L, A9, and A43 are appropriate for use in our next steps.

Predicted B cell epitopes

Linear B-cell epitopes for the F5L, A9, and A43 viral proteins were predicted using Bcepred and IEDB. The antigenicity value of A9 (0.4001) was lower than the antigenicity value of its epitopes. For both F5L and A43, the prediction results showed that almost all epitopes had higher antigenicity values than their respective proteins. Although both A43 and A9 had only one predicted B-cell epitope, F5L viral protein was predicted to have four predicted B-cell epitopes.

Table 3 Solubility, transmembrane helices, localization and antigenicity results predicted by SolPro, TMHMM, Virus-mPLoc and Vaxijen, respectively

Proteins	SolPro	TMHMM	Virus-mPLoc	Vaxijen v2.0 value
K2L	0.621132 InSoluble	0	Host Cytoplasm	0.5839/Probable Antigen
F5L	0.697036 Soluble	1	Host cell membrane. Host cytoplasm	0.5857/Probable Antigen
A9	0.582528 Soluble	2	Host cell membrane. Host cytoplasm	0.40451/Probable Antigen
A43	0.567605 InSoluble	3	Host cytoplasm	0.5384/Probable Antigen
A56	0.732497 InSoluble	1	Host endoplasmic reticulum. Host cytoplasm	0.4726/Probable Antigen
B5	0.553990 InSoluble	1	Host cell membrane, Secreted	0.5786/Probable Antigen

Table 4 The signal peptide results predicted by Signal-BLAST

Protein name	Sensitivity	Specificity	Balanced Prediction	Cleavage Site
K2L	+	+	+	+
F5L	+	+	+	+
A9	+	+	+	+
A43	+	+	+	+
A56	+	+	+	+
B5	+	+	+	+

Table 5 Allergenicity, BetaWrap Motifs, and host proteome similarity results predicted by AlgPred, BetaWrap, and BlastP, respectively

Protein name	AlgPred		BetaWrap Motifs (P Value)	BlastP
	IgE	Meme/Mast Motif		
K2L	–	Non-allergen	0.041	**Yes**
F5L	–	Non-allergen	0.10	No
A9	–	Non-allergen	No	No
A43	–	Non-allergen	0.0094	No
A56	+	Non-allergen	No	**Yes**
B5	–	Non-allergen	0.086	**Yes**

"–" represent no allergen while "+" represent allergen

"Yes" indicates positive similarity within the host proteome, suggesting potential immunogenic effects, while "No" signifies that the protein is a non-allergen

The highest predicted epitope derivated from F5L score was 1.3549 (LEISSTF), while the least predicted epitope (derivated from A9 protein) score was 0.5335 (LYISEQDDKKNTNNDSNSNNDKRNVSSINSNSSH). All predicted B-cell epitopes are shown in Table 6.

Predicted epitopes for MHC-I and MHC-II

Several MHC-I epitopes were predicted as probable antigens. The antigenicity values of the obtained epitopes for MHC-I were generally higher than those of their own proteins. Among the F5L, A9, and A43 viral proteins, an epitope (KRVNISLTCL) with the highest antigenicity value (2.4684) was predicted in the F5L protein. A9 viral protein had the least number of epitopes belonging to HLA-A*02:01, while other F5L and A43 viral proteins had more than two predicted epitopes for MHC-I (Table 7). The IC50 results were predicted to be lower than 50, with scores ranging from 10.62 nano molar (nM) to 48.72 nM, while the percentile rank ranged from 0.02 to 0.26. All predicted results for MHC-I are shown in Table 7.

Similar to MHC-I epitopes, numerous MHC-II epitopes were predicted as probable antigens. Antigenicity values of the obtained MHC-II epitopes were generally higher than those of their own proteins. Among the F5L, A9, and A43 viral proteins, an epitope (GRFGYVPYVGYKCI) of A9 protein was predicted with the highest antigenicity value for MHC-II (1.754). The predicted scores were taken between

Table 6 B cell epitopes predicted by both Bcepred and IEDB and antigenicity value predicted by Vaxijen v2.0

Protein name	Antigenicity value	B cell epitopes	Antigenicity value for epitopes
F5L	0.5857 ANTIGEN	ALATEDKT	1.0716 (Probable antigen)
		LEISSTF	1.3549 (Probable antigen)
		LNNNDNA	0.6546 (Probable antigen)
		NDLSSMTSQLQNDD	0.4380 (Probable antigen)
A43	0.5384 ANTIGEN	TILTSKH	0.8396 (Probable antigen)
A9	0.4001 ANTIGEN	LYISEQDDKKNTNNDSNSNNDKRNVSSINSNSSH	0.5335 (Probable antigen)

Table 7 Epitopes specific to selected MHC-I alleles

Protein name	Allele	Start	End	Peptide	IC50	Percentile Rank	Antigenicity
F5L	HLA-A*03:01	89	98	GTYKGVSHLK	10.62	0.02	0.7496 (Probable antigen)
	HLA-A*24:02	213	222	CYLFSQNYSF	26.98	0.05	1.0568 (Probable antigen)
	HLA-B*15:01	188	197	SLALEISSTF	29.09	0.14	1.0412 (Probable antigen)
	HLA-B*58:01	36	45	VMSHINYTSW	32.26	0.19	1.2436 (Probable antigen)
	HLA-A*01:01	54	63	LATEDKTSGY	34.59	0.08	1.0586 (Probable antigen)
	HLA-B*27:05	69	78	KRVNISLTCL	39.83	0.11	2.4684 (Probable antigen)
	HLA-B*15:01	132	141	VKIRCEITSF	48.72	0.26	1.4941 (Probable antigen)
A9	HLA-A*02:01	38	47	KLRPNSFWFV	10.74	0.09	1.1107 (Probable antigen)
	HLA-A*02:01	54	63	SMIMYLVLGI	25.26	0.22	0.9124 (Probable antigen)
A43	HLA-B*15:01	11	20	TMSIMPVLTY	12.29	0.06	0.7838 (Probable antigen)
	HLA-A*01:01	27	36	FHSEDIELCY	18.67	0.04	1.2985 (Probable antigen)
	HLA-B*27:05	60	69	YRYNFINRTF	19.71	0.04	0.7327 (Probable antigen)
	HLA-A*01:01	160	169	YINDRYNDIY	23.44	0.05	1.0274 (Probable antigen)
	HLA-A*02:01	8	17	SILTMSIMPV	26.88	0.24	0.9341 (Probable antigen)
	HLA-B*15:01	143	152	RSMCIAIIGY	30.94	0.16	1.2370 (Probable antigen)

> 0.5, while the percentile ranks were between 0.11 and 1.30. All predicted results for MHC-II are shown in Table 8.

Toxicity, allergenicity, solubility of predicted epitopes for MHC-I and MHC-II Alleles.

The antigenicity values of the epitopes obtained for MHC-I and MHC-II were estimated and the results of the predicted epitopes with the best scores were shown in Tables 9 and 10. Although no toxicity and hydrophobicity were predicted for all peptides, only RSMCIAIIGY (belonging to A43 viral protein) was predicted to have toxicity. Results for toxicity, allergenicity, and resolution of epitopes with higher scores were shown in Tables 9 and 10.

Docking results

The epitopes with the best scores for predicted F5L, A9, and A43 viral proteins (shown in Tables 9 and 10) were used for docking analyses with MHC-I and MHC-II alleles (Table 11, Supplementary 6). Since RSMCIIIIGY had toxicity and hydrophobicity, it was not considered for docking analysis. The predicted docking score for each predicted peptide with MHC-I alleles was given in Table 11 and Supplementary 6. Based on the docking results using HDOCK, the best interactions between MHC-I and predicted peptides were − 221.27 kcal/mol KRVNISLTCL-6IEX.PDB) and − 211.73 kcal/mol (KRVNISLTCL-3UTQ.PDB) (Table 11). On the other hand, the best interactions between MHC-II and predicted epitopes were − 234.88 kcal/mol (GRFGYVPYVGYKCI-3UTQ.PDB) and − 223.27 kcal/mol (GRFGYVPYVGYKCI-6IEX.PDB), where the peptide of A9 (GRFGYVPYVGYKCI) had the best scores with MHC-II alleles. The docking score between MHC-II and predicted epitopes was higher than the score between MHC-I and predicted epitopes (Table 11).

Images of the two highest-scoring peptide-protein docking results are also given in Fig. 1. The three-dimensional structures of the epitopes in both A(KRVNISLTCL-6IEX.PDB) and B(GRFGYVPYVGYKCI-3UTQ.PDB) are given in orange.

In light of these findings(Fig. 1, Tables 7, 8, 9, 10 and 11), the final selected epitopes demonstrated superior characteristics compared to others in terms of non-toxicity, non-allergenicity, solubility, strong antigenicity, and favorable docking interactions. Therefore, these epitopes were chosen for further analysis. Following their selection, the epitopes were subjected to molecular dynamics simulations and immunogenicity analysis to ensure their potential efficacy in vaccine development.

Energy minimization and molecular dynamics simulation of selected epitopes with MHC-I/II.

According to the docking results, the interactions with the best exothermic energy were selected for MD simulation. Those with the best scores for MHC-I were KRVNISLTCL-6IEX.PDB (− 212.03), while those with the best scores for MHC-II were GRFGYVPYVGYKCI-3UTQ.PDB (− 234.88 kcal/mol) (Table 11 and Fig. 1). Two of these complexes were taken for MD smulations. After all necessary adjustments such as energy minimization, solvation, etc., for the epitopes were made based on the GROMACS protein–ligand complex tutorial, we obtained all MDs results (Figs. 2 and 3). Based on our results, the The Root Mean Square Deviation (RMSD) for our complexes were changed

Table 8 Epitopes specific to selected MHC-II alleles

Protein Name	Allele	Start	End	Peptide	Score	Percentile Rank	Antigenicity
F5L	HLA-DRB3*01:01	43	56	TSWYYNDKVIALAT	0.8674	0.11	0.886 (Probable antigen)
	HLA-DRB3*01:01	42	55	YTSWYYNDKVIALA	0.8222	0.16	0.9841 (Probable antigen)
	HLA-DRB3*01:01	44	57	SWYYNDKVIALATE	0.7925	0.20	0.9128 (Probable antigen)
	HLA-DRB1*07:01	217	230	SQNYSFHKTLNVRN	0.7743	0.46	0.6687 (Probable antigen)
	HLA-DRB5*01:01	298	311	IVVIAAIAIYKRSK	0.7165	0.52	0.9797 (Probable antigen)
	HLA-DRB1*07:01	216	229	FSQNYSFHKTLNVR	0.6966	0.63	0.4065 (Probable antigen)
	HLA-DRB1*07:01	218	231	QNYSFHKTLNVRNI	0.6713	0.71	0.7362 (Probable antigen)
	HLA-DRB5*01:01	111	124	VKANIIDLTGRVRY	0.6694	0.73	1.251 (Probable antigen)
	HLA-DRB5*01:01	112	125	KANIIDLTGRVRYL	0.6499	0.84	1.2026 (Probable antigen)
	HLA-DRB5*01:01	48	61	NDKVIALATEDKTS	0.6025	1.10	0.9438 (Probable antigen)
A9	HLA-DRB1*15:01	377	390	KRTVFAHISHTINI	0.6038	1.10	0.8359 (Probable antigen)
	HLA-DRB1*07:01	260	273	KPYYGNTDNKFISY	0.6021	1.10	0.5294 (Probable antigen)
	HLA-DRB1*07:01	259	272	TKPYYGNTDNKFIS	0.5994	1.10	0.4525 (Probable antigen)
	HLA-DRB3*02:02	475	488	RTLTFNFTPKIFFR	0.5955	0.37	1.1946 (Probable antigen)
	HLA-DRB1*07:01	203	216	NKEFVYVPELSFIG	0.5864	1.20	0.8888 (Probable antigen)
	HLA-DRB3*02:02	259	272	TKPYYGNTDNKFIS	0.5756	0.42	0.4525 (Probable antigen)
	HLA-DRB3*02:02	473	486	GTRTLTFNFTPKIF	0.5734	0.44	1.119 (Probable antigen)
	HLA-DRB3*01:01	273	286	YPGYSQDEKDYIDA	0.5626	0.68	0.4185 (Probable antigen)
	HLA-DRB1*15:01	376	389	TKRTVFAHISHTIN	0.5623	1.20	0.7527 (Probable antigen)
	HLA-DRB1*15:01	82	95	GRFGYVPYVGYKCI	0.5528	1.30	1.754 (Probable antigen)
A43	HLA-DRB3*02:02	57	70	HIPYRYNFINRTFS	0.8042	0.12	0.9249 (Probable antigen)
	HLA-DRB5*01:01	96	109	PSLIVSLSGNLKYN	0.8198	0.21	0.7191 (Probable antigen)
	HLA-DRB3*02:02	56	69	EHIPYRYNFINRTF	0.7046	0.23	0.8068 (Probable antigen)
	HLA-DRB1*15:01	21	34	SSSIFRFHSEDIEL	0.8682	0.27	0.4834 (Probable antigen)
	HLA-DRB5*01:01	50	63	NIKYIPEHIPYRYN	0.7750	0.29	0.8938 (Probable antigen)
	HLA-DRB3*02:02	58	71	IPYRYNFINRTFSV	0.6747	0.30	1.0497 (Probable antigen)
	HLA-DRB5*01:01	97	110	SLIVSLSGNLKYND	0.7439	0.40	0.9004 (Probable antigen)
	HLA-DRB5*01:01	95	108	NPSLIVSLSGNLKY	0.7353	0.42	0.7765 (Probable antigen)
	HLA-DRB5*01:01	49	62	VNIKYIPEHIPYRY	0.6674	0.74	1.0399 (Probable antigen)
	HLA-DRB3*02:02	57	70	HIPYRYNFINRTFS	0.8042	0.12	0.9249 (Probable antigen)

Table 9 Toxicity, allergenicity, solubility and antigenicity for predicted epitopes with the best scores for the MHC-I alleles

Protein	Peptide	Toxicity	Hydrophobicity	Allergen	Solubility	Antigen
F5L	KRVNISLTCL	Non-toxin	−0.16	–	0.829714 Soluble	2.4684 (Probable antigen)
	VKIRCEITSF	Non-toxin	−0.13	–	0.902184 Soluble	1.4941 (Probable antigen)
A43	FHSEDIELCY	Non-toxin	−0.07	–	0.819464 Soluble	1.2985 (Probable antigen)
	YINDRYNDIY	Non-toxin	−0.30	–	0.994430 Soluble	1.0274 (Probable antigen)
	RSMCIAIIGY	**Toxin**	**+0.09**	–	0.986503 Soluble	1.2370 (Probable antigen)
A9	KLRPNSFWFV	Non-toxin	−0.12	–	0.844255 Soluble	1.1107 (Probable antigen)
	SMIMYLVLGI	Non-toxin	−0.35	–	0.645661 Soluble	0.9124 (Probable antigen)

The toxin one was highlighted in bold

Table 10 Toxicity, allergenicity, solubility and antigenicity for predicted epitopes with the best scores for the MHC-II alleles

Protein	Peptide	Toxicity	Hydrophobicity	Allergen	Solubility	Antigen
A43	IPYRYNFINRTFSV	Non-toxin	−0.15	–	0.864714 Soluble	1.0497 (Probable antigen)
	VNIKYIPEHIPYRY	Non-toxin	−0.13	–	0.992610 Soluble	1.0399 (Probable antigen)
A9	RTLTFNFTPKIFFR	Non-toxin	−0.15	–	0.991933 Soluble	1.1946 (Probable antigen)
	GTRTLTFNFTPKIF	Non-toxin	−0.07	–	0.990994 Soluble	1.1190 (Probable antigen)
	GRFGYVPYVGYKCI	Non-toxin	−0.09	–	0.998358 Soluble	1.754 (Probable antigen)
	NKEFVYVPELSFIG	Non-toxin	**+0.03**	–	0.533742 Soluble	0.8888 (Probable antigen)
F5L	VKANIIDLTGRVRY	Non-toxin	−0.19	–	1.000000 Soluble	1.2510 (Probable antigen)
	KANIIDLTGRVRYL	Non-toxin	−0.19	–	1.000000 Soluble	1.2026 (Probable antigen)
	YTSWYYNDKVIALA	Non-toxin	−0.01	–	0.741413 Soluble	0.9841 (Probable antigen)
	IVVIAAIAIYKRSK	Non-toxin	**+0.04**	–	0.999945 Soluble	0.9797 (Probable antigen)
	NDKVIALATEDKTS	Non-toxin	−0.23	–	0.826801 Soluble	0.9438 (Probable antigen)
	SWYYNDKVIALATE	Non-toxin	−0.06	–	0.716450 Soluble	0.9128 (Probable antigen)
	TSWYYNDKVIALAT	Non-toxin	−0.03	–	0.713492 Soluble	0.8860 (Probable antigen)
	QNYSFHKTLNVRNI	Non-toxin	−0.28	–	0.963317 Soluble	0.7362 (Probable antigen)

"–" represents an absence of both allergenic potential and hydrophobicity, whereas the symbol "+" denotes hydrophobicity

from 0.1 to 0.6 (Fig. 2). It was determined that the peptide in KRVNISLTCL-3NFN.PDB complex preserved its structural integrity better than other complexes, without any significant conformational changes based on RMSD results while the RMSD results in the GRFGYVPYVGYKCI-6IEX.PDB complexes were determined to be the most fluctuating compared to the other results (Fig. 2).

Table 11 The docking results (kcal/mol) for predicted F5L, A9, and A43 viral proteins's epitopes with the MHC-I/II alleles

Protein	Predicted Epitope	MHC Alleles	HDOCK
F5L	KRVNISLTCL	3UTQ.PDB	−211.73
F5L	KRVNISLTCL	3NFN.PDB	−205.03
F5L	**KRVNISLTCL**	**6IEX.PDB**	**−221.27**
A9	**GRFGYVPYVGYKCI**	**3UTQ.PDB**	**−234.88**
A9	GRFGYVPYVGYKCI	3NFN.PDB	−209.60
A9	GRFGYVPYVGYKCI	6IEX.PDB	−223.27

Bold values indicate the most favorable binding affinities, where lower kcal/mol values represent stronger predicted binding interactions with MHC alleles

In the RMSF analysis of the given molecular dynamics simulations, the GRFGYVPYVGYKCI-6IEX.PDB complex, it was determined that the atoms in this complex were more mobile than the other complexes (Fig. 2). However, it was determined that this GRFGYVPYVGYKCI-6IEX.PDB complex had the best exothermic score (−700,000) in the energy graph (Fig. 3). On the other hand, it was determined that the KRVNISLTCL-3UTG.PDB complex had a worse exothermic score than the others in the energy graph (Fig. 3). Both these tables enable a comparative analysis of protein dynamics, with the RMSD and RMSF metrics providing key insights into the stability and flexibility of protein structures during simulations. The visual and statistical information together offer a comprehensive overview that can support conclusions about the dynamic behavior of the proteins under study.

Results of in silico immunogenicity analysis of selected peptides

The in silico immunostimulatory potential of the peptides GRFGYVPVYGYKCI and KRVNISLTCL was evaluated

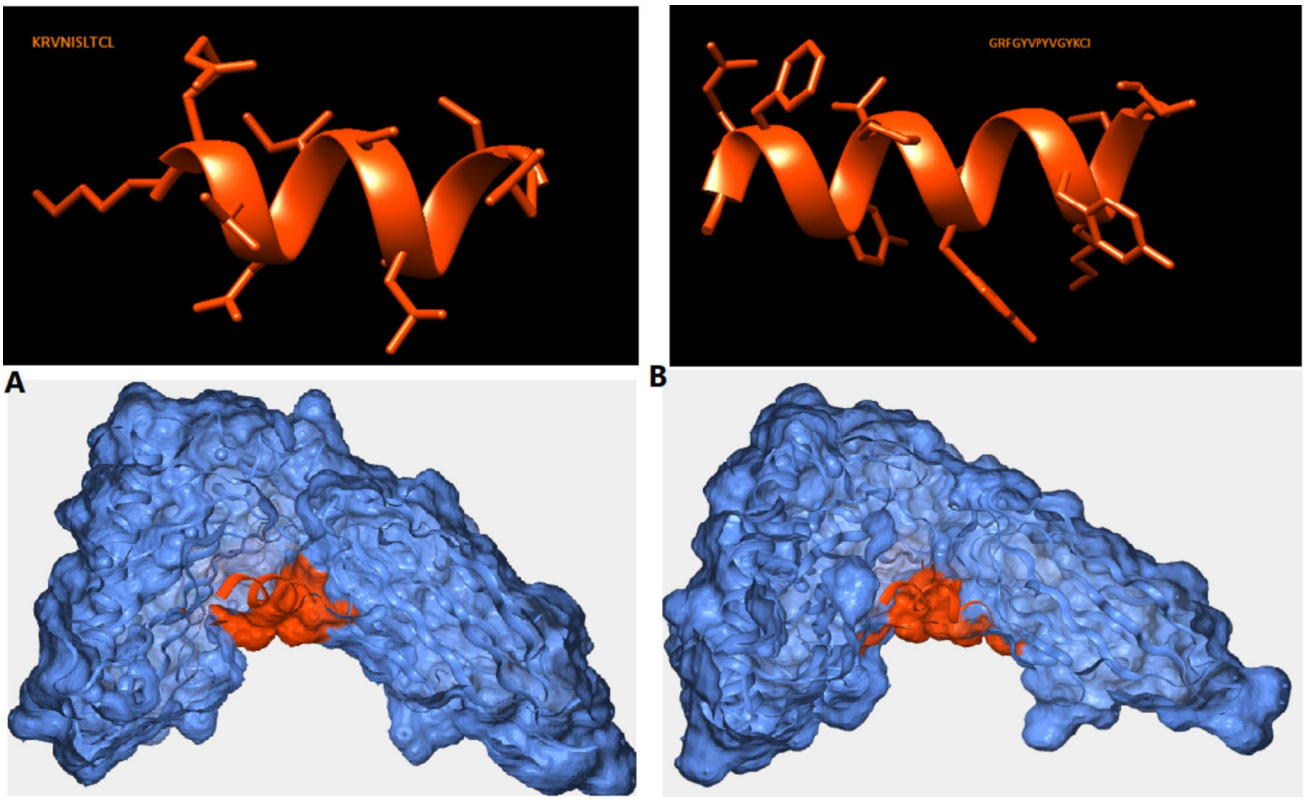

Fig. 1 represents the visual outcomes of the two peptide-protein placements with the highest docking scores. The images depict the configurations of KRVNISLTCL-6IEX.PDB (**A**) and GRFGYVPYVGYKCI-3UTQ.PDB (**B**), which achieved the highest scores in the docking analysis

using IFNepitope, IL4Pred, and CTLpred tools. The peptide GRFGYVPVYGYKCI was predicted to be a positive inducer of IFN-γ by IFNepitope, with a strong SVM score of 1.34 in IL4Pred, suggesting its potential to induce IL-4 cytokine responses. Moreover, CTLpred predicted GRFGYVPVYGYKCI as a CTL epitope with a high score of 0.970, further supporting its potential as a robust immunogenic peptide. In contrast, the peptide KRVNISLTCL was predicted to be a negative IFN-γ inducer by IFNepitope, and although IL4Pred suggested it as an IL-4 inducer with a lower SVM score of 0.36, CTLpred classified it as a non-epitope with a very low score of 0.080. These results suggest that GRFGYVPVYGYKCI is more likely to elicit a strong immune response compared to KRVNISLTCL.

Discussion

The basic idea behind all types of vaccination is that a vaccine should produce an effective immune response faster than the infecting virus itself. However, conventional vaccines that are developed based on a variety of biochemical assays may elicit antibodies in vaccinated individuals, but they require expensive, time-consuming, and potentially allergenic in vitro cultures of harmful microbes, which involve serious safety concerns when applying [15–17, 19, 27]. On the other hand, reverse vaccinology (RV) plays an important role in the development of vaccines by allowing in silico analysis of the genome of pathogens, which can significantly reduce the time and cost required to identify candidate peptides [17, 20, 51, 52]. RV strategy conducted with the help of in silico analysis is quite important because it enables the determination of the antigenic features of the pathogen and the identification of the sequences with the highest antigenicity, while also allowing for the assessment of whether these sequences are allergens, possible viral proteins that will activate the T and B epitopes, toxic, soluble, and safe before starting experiments in the laboratory environment [2, 22–27]. These studies show the necessity, reliability, and importance of bioinformatics preliminary analysis in target setting. Therefore, the basic step of our current study is to conduct bioinformatic analysis, and in vitro studies may be carried out in line with the data obtained from these analyses.

The 39 proteins of MPXV were not selected randomly. The reason being that all selected proteins are either assembly, budding, or surface proteins of MPXV [53–55]. Therefore, the first step of our study was to determine whether

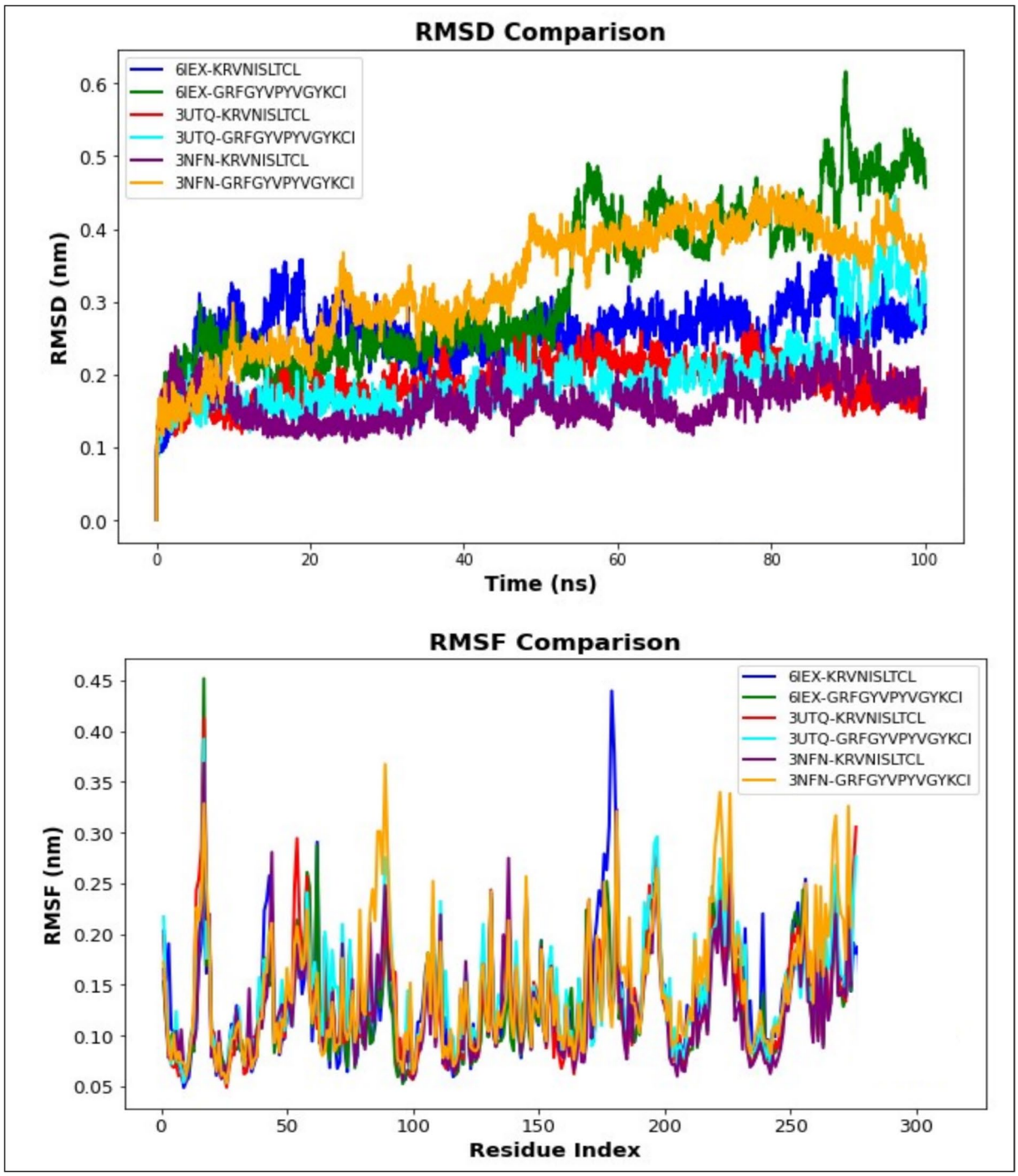

Fig. 2 provides a comparison of RMSD and RMSF values across six different molecular dynamics simulations. RMSD (top plot) shows the average deviation of protein backbone atoms from a reference structure over time, where lower values, such as those seen in KRVNISLTCL-3NFN.PDB, indicate greater structural stability. RMSF (bottom plot) highlights the flexibility of specific residues within the protein structures, with higher RMSF values, such as those observed in GRFGYVPYVGYKCI-6IEX.PDB, indicating greater mobility. Each simulation is distinguished by a unique color for clarity in comparison

these 39 proteins could be vaccine candidates. To develop vaccines that are safe and effective, it is crucial to optimize their sensitivity, specificity, balanced prediction, and cleavage site analysis [56]. By doing so, vaccines can be designed to provide robust protection against targeted pathogens while minimizing the risk of unwanted side effects or adverse reactions [56–58]. When predicting epitope candidates, sensitivity refers to the ability of the prediction method to accurately

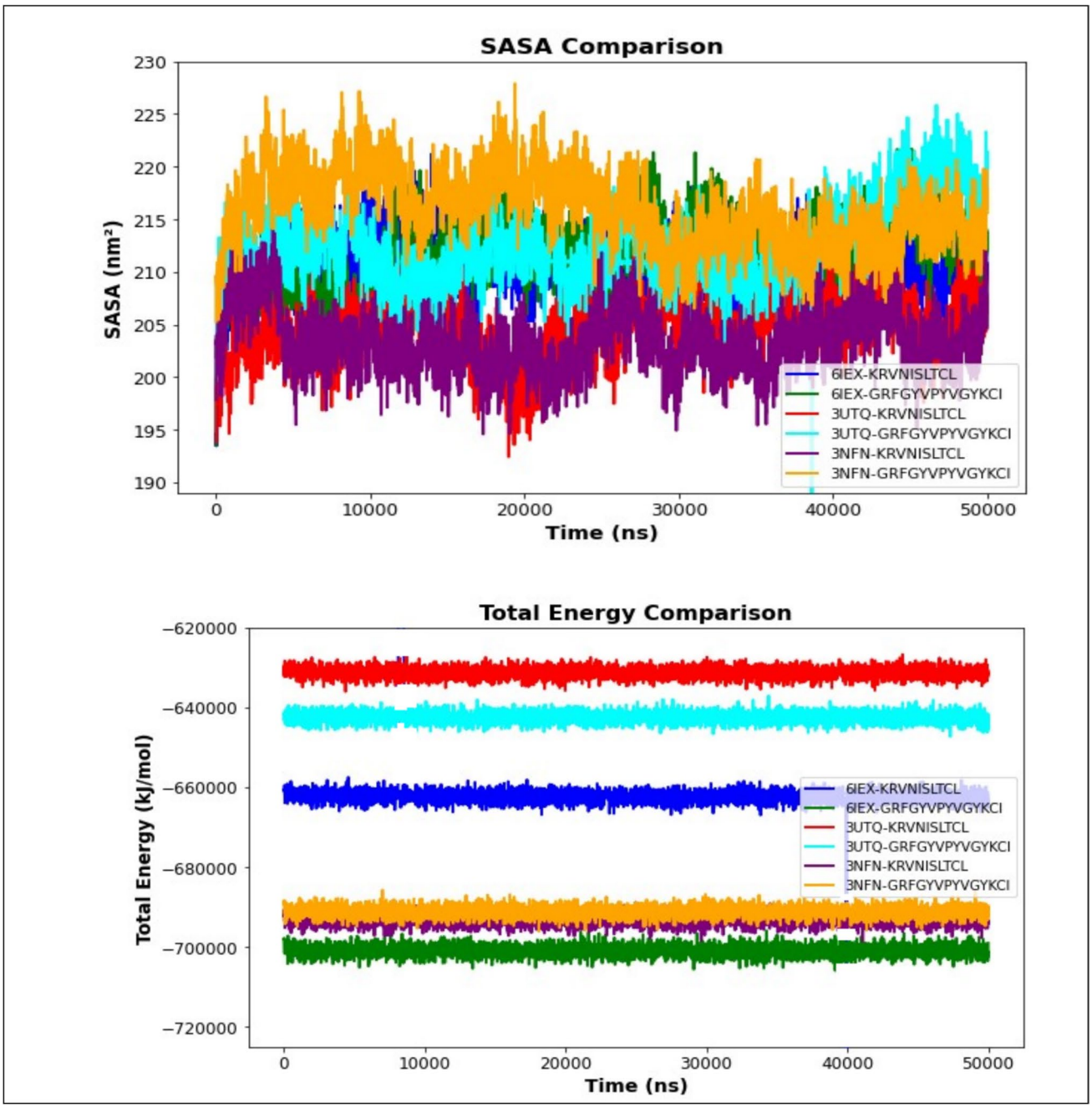

Fig. 3 provides an analysis of SASA values obtained from various molecular dynamics simulations and presents the energy profiles of protein structures as determined by molecular dynamics simulations. Changes in SASA can reflect conformational changes in the protein structure, including folding, unfolding, and interactions with other molecules. The total energy of the system is a critical factor in assessing the stability and viability of molecular conformations

identify true positive epitopes, while specificity refers to the ability to correctly exclude non-epitope regions as false positives. Achieving a balanced prediction means finding a compromise between these two aspects [59, 60]. And, only K2L, F5L, A9, A43, A56, and B5 proteins were predicted to have four different parameters of sensitivity, specificity, balanced prediction, and cleavage site (Table 4 and Supplementary 4). However, K2L and B5 have BlastP results, and A56 viral protein has allergenicity result (Table 5). Therefore, F5L, A9, and A43 viral proteins were deemed appropriate to be used in our present study. Physico-chemical characteristics including molecular weight, isoelectric point, and hydrophobicity are critical determinants that govern the efficacy of a protein vaccine [59]. And all predicted results are in both

Table 1 and Supplementary 1. The physicochemical analysis showed that F5L and A9 proteins had a negative GRAVY value, indicating that they were hydrophilic and had a better interaction with surrounding water molecules [60]. On the other hand, A43 protein had a positive GRAVY value, so it might not be good for interaction with surrounding water molecules [60] (Table 1). Besides, stable and soluble characteristics features are important parameters for biophysical studies on epitope-based vaccine design [21, 61]. F5L and A9 proteins have stable and soluble characteristics (Table 1).

As stated in the literature, natively unfolded protein regions and alpha-helical coiled coils peptides have been reported to be important forms of "structural antigens". Both of these structural forms could fold into their native structure and thus be recognized by antibodies naturally induced in response to infection [62]. In addition to having alpha-helical secondary structures features, F5L and A9 proteins showed a big molecular weight, and fewer than two transmembrane helices, facilitating cloning, expression, and purification [63] (Tables 2 and Table 3). Understanding the estimated half-life of peptide candidates or proteins is crucial for optimizing their pharmacokinetic properties, guiding formulation strategies, determining dosing regimens, and assessing their therapeutic efficacy and safety. All of our predicted samples have long estimated half-life (more than 10 h) (Table 1). Moreover, F5L and A9 had sensitivity, specificity, balanced prediction, and cleavage site (Table 4). F5L and A9 proteins with these prominent physico-chemical properties can be used as vaccine candidate antigens.

In addition to the physico-chemical properties, other predicted parameters such as the presence of a signal that creates a response in the immune system, non-allergenicity of the related proteins, sub-cellular location, having BetaWrap motifs, and their sensitivity and specificity are very important for vaccine candidates [16, 21, 51, 64, 65]. Having signal peptides is so important for proteins because they affect immune responses and possess high epitope densities [38]. Moreover, Dangi and colleagues reported that most known vaccine candidates also possess signal peptides [16]. Hence, it is worthwhile to predict signal peptides in proteins prior to epitope predictions. According to our results, K2L, F5L, A9, A43, A56, and B5 proteins were all predicted to have four different parameters of sensitivity, specificity, balanced prediction, and cleavage site (Table 4). However, two of them (K2L and B5) have BlastP results. It is clear that there should not be similarity with host proteins because this situation may affect the host's autoimmunity in a bad way [16]. Therefore, proteins that have similarities with host proteins were not considered in the present study. Additionally, A56 viral protein has an allergenicity result (Table 5). The candidate proteins should not have allergenicity because this feature affects the reactions (type-1 hypersensitivity) for the host [66, 67]. Therefore, the viral proteins F5L, A9, and A43 were deemed appropriate for use in our next steps. However, the A43 protein has a positive GRAVY value, so it might not interact well with surrounding water molecules [60] (Table 1). In addition, candidates with antigenicity can trigger the immune system and both F5L and A9 were predicted to be antigenic (Table 3) [68]. The prediction of sub-cellular location is important because it aids in the identification of potential vaccine candidates [51, 64, 65], and this prediction may help in the formulation of better therapeutic options against the virus [16]. F5L and A9 protein's sub-cellular location within the host were predicted to be in the host cell membrane and host cytoplasm, while A43 was predicted to be in the host cytoplasm (Table 3). Based on the literature, BetaWrap motifs are dominant in virulence factors of pathogens, and if proteins are predicted to possess such motifs, then they are appropriate to be taken under reverse vaccinology studies [16]. Therefore, our candidate proteins should have this feature. Even though A43 and F5L have BetaWrap motifs, A9 has no BetaWrap motifs (also K2L, F13L, D13, A43, and B5 proteins have BetaWrap motifs, Table 5). As a result, F5L is the best protein to use for RV due to having all these features.

The next step of our study was to predict epitope regions specific to B and T cells in the F5L, A9, and A43 viral proteins. Although the GRAVY value of the A49 protein was predicted to be positive (Table 1), candidate epitopes from this protein may not be hydrophobic [69]. Therefore, we tried to find epitope regions specific to B and T cells in the F5L, A9, and A43 viral proteins. While A43 and A9 have only one predicted B cell epitope, the F5L viral protein has four predicted B cell epitopes. The highest predicted epitope in F5L has a score of 1.3549 (LEISSTF), and even the least score in A9 is 0.5335 (LYISEQDDKKNTNNDSNSNNDKRNVSSINSNSSH) (Table 6). Like B cell epitopes, epitopes that were predicted with a high antigenicity value for MHC-I and MHC-II were found (Table 7). MHC-II (major histocompatibility complex class II) molecules activate and propagate CD4+T cells, which can activate and guide other immune cells such as B cells and CD8+T cells. On the other hand, MHC-I (major histocompatibility complex class I) molecules present peptides from intracellular proteins to CD8+T cells, inducing CD8+T cell responses against viruses. Since MHC-I/II molecules have these important roles [70, 71], it was extremely important to obtain epitopes that may trigger these molecules [71]. The length of epitopes for MHC-I is important because MHC-I molecules have a narrow peptide-binding groove that can accommodate peptides of a specific length [72–75]. Therefore, 8–10 amino acids in length epitopes should be preferred for MHC-I molecules [72, 74, 76]. So, we took ten length for finding our epitopes. Among the F5L, A9, and A43 viral proteins, an epitope (KRVNISLTCL) that had the highest antigenicity value (2.4684) was predicted in

the F5L protein. A9 viral protein had the least number of epitopes that belong to HLA-A*02:01, while the other F5L and A43 viral proteins had more than two predicted epitopes for MHC-I (Table 7). Among the F5L, A9, and A43 viral proteins, an epitope (GRFGYVPYVGYKCI) that had the highest antigenicity value (1.754) was predicted in the A9 protein for MHC-II (Table 8). Candidate epitopes with the highest antigenicity in MHC-I and MHC-II were shown in Table 10. Toxicity, allergenicity, and solubility properties of these epitopes with high antigenicity were also determined. Next, the predicted epitopes were used to dock with the most common alleles PDB: 3UTQ, PDB: 3NFN, and PDB: 6IEX, whose 3D structure is known (Table 11). Epitope toxicity is a major concern when designing vaccines, as toxicity may cause adverse effects on the body either directly or indirectly [77]. Additionally, the hydrophobicity of the epitope can affect its ability to interact with other vaccine components and its stability in aqueous solutions, while solubility may affect its formulation [77, 78]. Moreover, epitopes should not be allergenic, as they can harm the host [79, 80]. Finally, the presence of a cleavage site in an epitope can enhance the processing, immunogenicity, and presentation of the epitope, making it an important feature to consider when designing epitopes for vaccines [81, 82]. Our epitope candidates have been carefully selected, taking into account important features such as solubility, non-allergenicity, and remarkable antigenicity. All of the above [77–82] information were predicted in Tables 9 and 10. Therefore, the epitopes KRVNISLTCL and GRFGYVPYVGYKCI, which possess all of these features, could be promising vaccine candidates [77–82]. We recommend that researchers aiming to develop vaccines in this area consider utilizing these two epitopes.

The docking scores obtained from the HDOCK server serve as preliminary quantitative estimates of the binding affinities between predicted epitopes and MHC alleles, as presented in Table 11. These scores, which inversely correlate with interaction strength—with more negative values indicating higher affinity—provide crucial insights into the molecular interactions at play [83]. To further investigate these interactions, MD simulations were conducted, offering detailed insights into the stability and dynamics of the peptide-MHC complexes by analyzing parameters such as RMSD, RMSF, SASA, and the total energy of the system(Figs. 2 and 3) [38, 84, 85]. RMSD values ranged from 0.1 to 0.6, reflecting varying levels of structural integrity. The KRVNISLTCL-3NFN.PDB complex showed minimal conformational changes and lower RMSD values, indicating stable interactions between the peptide and the MHC molecule, which aligns with the findings of Maruyama et al. and other studies where lower RMSD values are associated with better structural preservation [48, 85, 86]. In contrast, the GRFGYVPYVGYKCI-6IEX.PDB complex exhibited more significant RMSD fluctuations, which may indicate more dynamic and less stable interactions (Fig. 2).However, as seen in other studies, higher RMSD values do not always signify unfavorable interactions, as they can also indicate a flexible binding interface that may still be energetically favorable [87, 88]. RMSF analysis further supports this observation (Fig. 2). The atoms in GRFGYVPYVGYKCI-6IEX.PDB were more mobile, reflecting increased flexibility. This increased mobility may correlate with reduced binding affinity, aligning with observations made by Poudyal et al. on the flexibility of peptide-MHC interactions [84, 88]. However, the complex still exhibited a favorable exothermic energy score, suggesting that despite its structural fluctuations, the interaction remained energetically stable—a phenomenon often seen in peptide-MHC interactions [88, 89]. The SASA values provided additional insight into the degree of peptide exposure to the solvent, with lower SASA values in KRVNISLTCL-3NFN.PDB suggests a more compact structure. This correlation between RMSD stability and SASA compactness aligns with previous literature, which highlights that stable peptide-MHC complexes tend to have reduced solvent-exposed surface areas, potentially enhancing immune recognition [89]. Despite the GRFGYVPYVGYKCI complex showing higher RMSD and RMSF values, its superior exothermic energy score points to the potential of a dynamic equilibrium where the complex, while flexible, remains energetically favorable [86, 90]. Conversely, the KRVNISLTCL-3UTG.PDB complex, despite its lower RMSD fluctuations, showed a less favorable energy score, indicating that although structurally stable, it may not be energetically optimized, which could affect its biological efficacy [86]. These insights into the RMSD, RMSF, SASA, and energy profiles of the peptide-MHC complexes provide a deeper understanding of the biophysical underpinnings of antigen presentation, crucial for predicting the potential efficacy of peptide-based therapeutics [85–90].

The peptide GRFGYVPYVGYKCI showed strong potential to induce immune responses, particularly through the IFN-γ pathway, with a high SVM score of 1.34 in IL4Pred and a CTLpred score of 0.970, indicating its robust immunogenic potential [26]. Studies have demonstrated that IFN-γ plays a critical role in activating both innate and adaptive immune responses, enhancing antigen presentation by MHC molecules, and promoting the differentiation of Th1 cells, which are crucial for a cytotoxic T-cell response [86, 91]. Peptides capable of inducing IFN-γ production are considered ideal candidates for vaccines, as they can elicit a broad immune response [91]. The high CTLpred score further supports this peptide's ability to stimulate CTL activity, which is essential for clearing intracellular pathogens like viruses [91, 92]. Moreover, IL-4, which is associated with Th2-type immune responses, aids in humoral immunity by promoting antibody production. While IFN-γ is traditionally linked to cell-mediated immunity, the ability

of GRFGYVPYVGYKCI to also potentially induce IL-4 responses (based on the IL4Pred score) suggests it may provoke a balanced immune response, engaging both Th1 and Th2 pathways [91–93]. This could make the peptide a more versatile vaccine candidate, as vaccines that stimulate both humoral and cellular immunity are often more effective in the long-term protectionity of KRVNISLTCL: In contrast, the peptide KRVNISLTCL was predicted to have limited immunostimulatory potential. It was classified as a negative IFN-γ inducer by IFNepitope and showed a significantly lower SVM score of 0.36 in IL4Pred, indicating its limited capacity to induce IL-4 cytokine responses. This suggests that KRVNISLTCL may not effectively engage the immune system through the Th1 or Th2 pathways [93]. Additionally, its very low CTLpred score of 0.080 indicates that it is unlikely to function as a strong CTL epitope, reducing its potential efficacy in vaccine development [94]. In conclusion, the MD simulation and docking analyses, combined with in silico immunogenicity assessments, underscore the potential of GRFGYVPYVGYKCI as a promising candidate for vaccine development. Despite the dynamic fluctuations observed in its structure, its overall energetic stability and strong immunostimulatory potential highlight its promise as a robust immunogenic peptide. On the other hand, the relatively lower immunogenic potential of KRVNISLTCL suggests that while it may form stable structural interactions with MHC molecules, it may not be as effective in eliciting a strong immune response. These insights are instrumental in understanding the biophysical and immunogenic properties of peptide-MHC interactions, which is crucial for advancing peptide-based therapeutics in vaccine development.

Conclusion

In this study, we employed a reverse vaccinology (RV) approach to identify potential vaccine candidates against the monkeypox virus (MPXV). RV enables in silico analysis of viral proteins to predict key properties like antigenicity, allergenicity, toxicity, and immunogenic potential, streamlining the process before in vitro testing. We focused on 39 MPXV proteins involved in viral assembly, budding, and surface interactions, as these proteins play crucial roles in the virus's attachment to the host cell, fusion with the cell membrane, and entry into the cell, making them strategic targets for vaccine development. Our analysis identified the F5L protein as the most promising candidate due to its hydrophilic nature (negative GRAVY value), stability, solubility, and alpha-helical structure, which allows for effective antibody recognition. Additionally, F5L has a high molecular weight, fewer than two transmembrane helices, and a long half-life (over 10 h), optimizing it for cloning, expression, and sustained therapeutic effects. It also features multiple B-cell epitopes and epitopes for both MHC-I and MHC-II, supporting a balanced immune response. Its non-allergenic nature, similarity to host proteins, and presence of BetaWrap motifs further solidify its potential for vaccine development. The A9 and A43 proteins were also found to be promising due to their signal peptides and favorable antigenicity, despite some limitations (e.g., A43's positive GRAVY value and A9's lack of BetaWrap motifs). Molecular docking and dynamics simulations affirmed the potential of the epitopes KRVNISLTCL and GRFGYVPYVGYKCI; KRVNISLTCL showed remarkable structural stability, while GRFGYVPYVGYKCI exhibited dynamic yet energetically favorable interactions with MHC molecules. In silico immunogenicity analysis further supported GRFGYVPYVGYKCI's ability to induce a strong immune response, particularly through IFN-γ and CTL activity. In summary, our RV-based analysis highlights the F5L protein and epitopes KRVNISLTCL and GRFGYVPYVGYKCI as strong vaccine candidates. These findings provide a foundation for future experimental validation and underscore the importance of bioinformatics in advancing vaccine development. Further, wet lab studies are recommended to explore their immunogenic potential for an effective MPXV vaccine.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11030-024-11033-1.

Acknowledgements The authors would like to acknowledge that this paper is submitted in partial fulfillment of the requirements for PhD degree at Yildiz Technical University

Author contribution E.A. (Emre Aktaş) wrote the paper, designed the research, conducted Molecular Dynamics Simulations, and performed other analyses; O.U.S. (Osman Uğur Sezerman) designed the research, wrote the paper, supervised the analyses, and served as the supervisor for the study; M.Ö. (Murat Özer), K.K. (Kevser Kübra Kırboğa), A.E.K. (Ahmet Efe Köseoğlu), and N.Ö.Ö (Nehir Özdemir Özgentürk) contributed to the writing of the manuscript. All authors contributed to the revisions of the manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Alakunle E, Kolawole D, Diaz-Canova D, Alele F, Adegboye O, Moens U, Okeke MI (2024) A comprehensive review of monkeypox virus and mpox characteristics. Front Cell Infect Microbiol 14:1360586. https://doi.org/10.3389/fcimb.2024.1360586
2. Hu B, Guo H, Zhou P, Shi ZL (2021) Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol 19(3):141–154
3. Maldonado MS, Lucchetti AJ, Pacheco RAP, Cevallos LCM, Saavedra EUZ, Zapata LRP et al (2023) Epidemiologic characteristics and clinical features of patients with monkeypox virus

infection from a hospital in Peru between July and September 2022. Int J Infect Dis 129:175–180. https://doi.org/10.1016/j.ijid.2023.01.045
4. Scarpa F, Azzena I, Ciccozzi A, Branda F, Locci C, Perra M, Pascale N, Romano C, Ceccarelli G, Terrazzano G et al (2024) Update of the genetic variability of monkeypox virus Clade IIb Lineage B.1. Microorganisms 12(9):1874. https://doi.org/10.3390/microorganisms12091874
5. Hossain FMA, Bappy MNI, Robin TB, Ahmad I, Patel H, Jahan N et al (2024) A review on computational studies and bioinformatics analysis of potential drugs against monkeypox virus. J Biomol Struct Dyn 42(12):6091–6107
6. Agarwal P, Shukla N, Bhatia A, Mahfooz S, Narayan J (2024) Comparative genome analysis reveals driving forces behind monkeypox virus evolution and sheds light on the role of ATC trinucleotide motif. Virus Evol 10(1):veae043. https://doi.org/10.1093/ve/veae043
7. Saxena SK, Ansari S, Maurya VK, Kumar V, Islam S, Srivastava N (2022) Re-emerging human monkeypox: a major public-health debacle. J Med Virol 95(1):e27902
8. Beer EM, Rao VB (2019) A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. PLoS Negl Trop Dis 13(10):e0007791. https://doi.org/10.1371/journal.pntd.0007791
9. Dou YM, Yuan H, Tian HW (2022) Monkeypox virus: past and present. World J Pediatr 19(2):224–230. https://doi.org/10.1007/s12519-022-00618-1
10. Farahat RA, AlAseri Z, Memish ZA, Saeedi M, Alzahrani F, Zowawi HM (2022) Human monkeypox disease (MPX). Infezioni in Medicina 30(3):372–391 (**PMID: 36348660**)
11. Ulloque-Badaracco JR, Alarcón-Braga EA, Hernandez-Bustamante EA, Al-kassab-Córdova A, Benites-Zapata VA, Bonilla-Aldana DK, Rodriguez-Morales AJ (2022) Acceptance towards monkeypox vaccination: a systematic review and meta-analysis. Pathogens 11(11):1248. https://doi.org/10.3390/pathogens11111248
12. Ortiz-Martínez Y, Sarmiento L, Bonilla-Aldana DK, Rodríguez-Morales AJ, Henao-Martínez AF (2022) Monkeypox—a description of the clinical progression of skin lesions: a case report from Colorado, USA. Ther Adv Infect Dis. https://doi.org/10.1177/20499361221117
13. Aldhaeefi M, AlHogbani T, AlBalwi M, AlBadr A, AlHarbi H (2022) The 2022 human monkeypox outbreak: Clinical review and management guidance. Am J Health Syst Pharm 80(1):44–52
14. Kugelman JR, Johnston SC, Mulembakani PM, Kisalu NK, Lee MS, Koroleva G et al (2014) Genomic variability of monkeypox virus among humans, Democratic Republic of the Congo. Emerg Infect Dis 20(2):232–239. https://doi.org/10.3201/eid2002.130118
15. Goodswen SJ, Kennedy PJ, Ellis JT (2013) A guide to in silico vaccine discovery for eukaryotic pathogens. Brief Bioinform 14(6):753–774
16. Dangi M, Kumari R, Singh B, Chhillar AK (2018) Advanced in silico tools for designing of antigenic epitope as potential vaccine candidates against Coronavirus. Bioinform Seq Struct Phylogeny 15:329–357. https://doi.org/10.1007/978-981-13-1562-6_15
17. Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omidi Y (2020) Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. Drug Discov Today 25(4):1034–1042. https://doi.org/10.1016/j.drudis.2020.03.006
18. Bhattacharya M, Sharma AR, Patra P, Ghosh P, Sharma G, Patra BC et al (2020) Immunoinformatics approach to understand molecular interaction between multi-epitopic regions of SARS-CoV-2 spike-protein with TLR4/MD-2 complex. Infect Genet Evol 85:104587. https://doi.org/10.1016/j.meegid.2020.104587
19. Ghosh P, Chakraborty P, Bhattacharya M, Sharma AR, Sharma G, Patra BC, Lee SS (2021) A novel multi-epitopic peptide vaccine candidate against *Helicobacter pylori*: In-silico identification, design, cloning and validation through molecular dynamics. Int J Pept Res Ther 27(2):1149–1166. https://doi.org/10.1007/s10989-020-10157-w
20. Miah MM, Tabassum N, Zinnia MA, Islam ABMMd (2022) Drug and anti-viral peptide design to inhibit the monkeypox virus by restricting A36R protein. Bioinform Biol Insights. https://doi.org/10.1177/11779322221141164
21. Can H, Güner A, Gülmez A, Kalayci AS, Gündüz A (2020) In silico discovery of antigenic proteins and epitopes of SARS-CoV-2 for the development of a vaccine or a diagnostic approach for COVID-19. Sci Rep 10:22387. https://doi.org/10.1038/s41598-020-79645-9
22. de Araújo LP, de Melo Santos NC, Corsetti PP, de Almeida LA (2024) Immunoinformatic approach for rational identification of immunogenic peptides against host entry and/or exit Mpox proteins and potential multiepitope vaccine construction. J Infect Dis 229(Supplement_2):S285–S292. https://doi.org/10.1093/infdis/jiad443
23. Dallavilla T, Rowland J, Avogaro A, Spera P (2020) Bioinformatic analysis indicates that SARS-CoV-2 is unrelated to known artificial coronaviruses. Eur Rev Med Pharmacol Sci 24(9):4558–4564
24. Abdi SAH, Islam ABMMd, Ashraf S, Dangi M, Chhillar AK, Khanday FA (2022) Multi-epitope-based vaccine candidate for monkeypox: an in silico approach. Vaccines 10(10):1564
25. Romero-Lopez JP, Carnalla-Cortez E, Pacheco-Olvera LD, Cuevas-Fernandez CA, Canizalez-Roman A, Andrade-Villanueva JF (2021) A bioinformatic prediction of antigen presentation from SARS-CoV-2 spike protein revealed a theoretical correlation of HLA-DRB1*01 with COVID-19 fatality in Mexican population: an ecological approach. J Med Virol 93(4):2029–2038
26. Gharbavi M, Danafar H, Amani J, Sharafi A (2021) Immunoinformatics analysis and expression of a novel multi-domain antigen as a vaccine candidate against glioblastoma. Int Immunopharmacol 91:107265. https://doi.org/10.1016/j.intimp.2020.107265
27. Bai Y, Zhou M, Wang N, Yang Y, Wang D (2024) Designing a candidate multi-epitope vaccine against transmissible gastroenteritis virus based on immunoinformatic and molecular dynamics. Int J Mol Sci 25(16):8828. https://doi.org/10.3390/ijms25168828
28. Bhattacharya M, Chatterjee S, Nag S, Dhama K, Chakraborty C (2022) Designing, characterization, and immune stimulation of a novel multi-epitopic peptide-based potential vaccine candidate against monkeypox virus through screening its whole genome encoded proteins: an immunoinformatics approach. Travel Med Infect Dis 50:102481
29. Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38(7):3022–3027. https://doi.org/10.1093/molbev/msab120
30. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005. Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, pp 571–607. https://doi.org/10.1385/1-59259-890-0:571
31. Cheng J, Randall AZ, Sweredoski MJ, Baldi P (2005) SCRATCH: A protein structure and structural feature prediction server. Nucleic Acids Res 33(Web Server issue):W72–W76
32. Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 266:540–553. https://doi.org/10.1016/S0076-6879(96)66034-0
33. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8(1):4
34. Saha S, Raghava GPS (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34(Web Server issue):W202–W209

35. Nguyen MN, Rai DK, Siddharth P, Gidudu P, Nagata AL, Wickner PG et al (2022) AllerCatPro 2.0: a web server for predicting protein allergenicity potential. Nucleic Acids Res 50(W1):W36–W43. https://doi.org/10.1093/nar/gkac446
36. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery Consortium, Raghava GPS (2013) In silico approach for predicting toxicity of peptides and proteins. PLoS ONE 8(9):e73957
37. Shen HB, Chou KC (2010) Virus-mPLoc: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites. J Biomol Struct Dyn 28(2):175–186. https://doi.org/10.1080/07391102.2010.10507351
38. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305(3):567–580
39. Bradley P, Cowen L, Menke M, King J, Berger B (2001) BETAWRAP: successful prediction of parallel β-helices from primary sequence reveals an association with many microbial pathogens. Proc Natl Acad Sci USA 98(26):14819–14824. https://doi.org/10.1073/pnas.251267298
40. Ribas-Aparicio RM, Castelán-Vega JA, Jiménez-Alberto A, Monterrubio-López GP, Aparicio-Ozores G (2017) The impact of bioinformatics on vaccine design and development. IntechOpen. https://doi.org/10.5772/intechopen.69273
41. Chen Z, Zhu Y, Sha T et al (2021) Design of a new multi-epitope vaccine against Brucella based on T and B cell epitopes using bioinformatics methods. Epidemiol Infect 149:e136. https://doi.org/10.1017/S0950268821001229
42. Mukherjee S, Tworowski D, Detroja R, Mukherjee SB, Frenkel-Morgenstern M (2020) Immunoinformatics and structural analysis for identification of immunodominant epitopes in SARS-CoV-2 as potential vaccine targets. Vaccines 8(2):290. https://doi.org/10.3390/vaccines8020290
43. Venkatesh G, Grover A, Srinivasaraghavan G, Rao S (2020). MHCAttnNet: predicting MHC-peptide bindings for MHC alleles classes I and II using an attention-based deep neural model. Bioinformatics 36(Supplement_1):i399–i406
44. Schijns V, Majhen D, van der Ley P, Thakur A, Summerfield A, Berisio R, Nativi C, Fernández-Tejada A, Alvarez-Dominguez C, Gizurarson S et al (2021) Rational vaccine design in times of emerging diseases: the critical choices of immunological correlates of protection, vaccine antigen and immunomodulation. Pharmaceutics 13(4):501. https://doi.org/10.3390/pharmaceutics13040501
45. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
46. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H (2022) GROMACS in the cloud: a global supercomputer to speed up alchemical drug design. J Chem Inf Model 62(7):1691–1711
47. Maruyama Y, Igarashi R, Ushiku Y, Mitsutake A (2023) Analysis of protein folding simulation with moving root mean square deviation. J Chem Inf Model 63(5):1529–1541
48. Lemkul JA (2018) From proteins to perturbed Hamiltonians: a suite of tutorials for the GROMACS-2018 molecular simulation package, v1.0. Living J Comput Mol Sci 1(1):5068
49. Biswas S, Mahmud S, Mita MA, Afrose S, Hasan MR, Sultana Shimu M et al (2022) Molecular docking and dynamics studies to explore effective inhibitory peptides against the spike receptor binding domain of SARS-CoV-2. Front Mol Biosci 8:791642
50. Chaudhuri R, Kulshreshtha D, Raghunandanan MV, Ramachandran S (2014) Integrative immunoinformatics for mycobacterial diseases in R platform. Syst Synth Biol 8(1):27–39
51. Piri-Gharaghie T, Doosti A, Mirzaei SA (2022) Identification of antigenic properties of *Acinetobacter baumannii* proteins as novel putative vaccine candidates using reverse vaccinology approach. Appl Biochem Biotechnol 194(11):4892–4914
52. Sbrana E, Xiao SY, Newman PC, Brown ST, Shope RE (2007) Pathology of West African monkeypox. J Comp Pathol 136(1):28–35
53. Dabrowski PW, Radonić A, Kurth A, Nitsche A (2011) Genome-wide comparison of cowpox viruses reveals a new clade related to Variola virus. PLoS ONE 6(5):e79953
54. Osadebe LU, Ogbuniwe IJ, Moa B, Doshi RH, Carlson AL, Dang VC et al (2021) Monkeypox virus transmission by aerosol and immune response of nonhuman primates. Emerg Infect Dis 27(4):916–919
55. Krammer F (2020) SARS-CoV-2 vaccines in development. Nature 586(7830):516–527
56. Khatoon N, Pandey RK, Prajapati VK, Kumar M (2021) Designing effective vaccines: current perspectives and challenges. Eur J Pharm Sci 157:105623
57. Wang Z, Lorenzi JCC, Muecksch F, Finkin S, Viant C, Gaebler C et al (2021) Enhanced SARS-CoV-2 neutralization by dimeric IgA. Sci Transl Med 13(577):eabf1555
58. Singh H, Raghava GPS (2019) Designing of peptide vaccine against SARS-CoV-2. Front Immunol 11:1–14
59. Droppa-Almeida D, Franceschi E, Padilha FF (2018) Immune-informatic analysis and design of peptide vaccine from multi-epitopes against *Corynebacterium pseudotuberculosis*. Bioinform Biol Insights 12:1177932218755337
60. Shey RA, Ghogomu SM, Esoh KK, Nebangwa DN, Shintouo CM, Nongley NF et al (2019) In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Sci Rep 9:4409
61. Corradin G, Villard V, Kajava AV (2007) Protein structure based strategies for antigen discovery and vaccine development against malaria and other pathogens. Endocr Metab Immune Disord Drug Targets 7(4):259–265
62. Meunier M, Guyard-Nicodème M, Hirchaud E, Parra A, Chemaly M, Dory D et al (2016) Identification of novel vaccine candidates against campylobacter through reverse vaccinology. J Immunol Res 2016:5715790
63. Pizza M, Scarlato V, Masignani V, Giuliani MM, Aricò B, Comanducci M et al (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287(5459):1816–1820
64. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R et al (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26(13):1608–1615
65. Dimitrov I, Bangov I, Flower DR, Doytchinova I (2014) AllerTOP vol 2—a server for in silico prediction of allergens. J Mol Model 20:227
66. Costa J, Mafra I, Carrapatoso I, Oliveira MBPP (2022) Are physicochemical properties shaping the allergenic potency of plant allergens? Clin Rev Allergy Immunol 62(1):37–63
67. Mittal A, Khattri A, Verma V (2022) Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. PLoS Pathog 18(4):e1010260
68. Moreira RS, Filho VB, Calomeno NA, Wagner G, Miletti LC (2022) EpiBuilder: a tool for assembling, searching, and classifying B-cell epitopes. Bioinform Biol Insights 16:11779322221095220
69. Pamer EG (2018) Immune responses to *Listeria monocytogenes*. Nat Rev Immunol 18(6):517–528. https://doi.org/10.1038/nri1461
70. Khorraminejad-Shirazi M, Ghasemi M, Abdolalizadeh J, Razavi SM (2018) MHC Class I and Class II molecules: structure,

function, and roles in immune response. Rev Med Microbiol 29(2):67–72
71. Sarkar S, Malmberg C, Ul Islam S, Alizadeh M, Wahlgren M, Sandalova T (2021) Understanding the relationship between epitope length and binding affinity for MHC Class I: a combined experimental and computational approach. Front Immunol 12:717131
72. Gonzalez-Paz L, Ding Y, Nowotny C, Zhang D, Geller R, Shapiro BA et al (2021) Structural deformability induced in proteins of potential interest associated with COVID-19 by binding of homologues present in ivermectin: Comparative study based in elastic networks models. J Mol Liq 340:117284
73. Zaib S, Irfan M, Ashraf N, Abro A, Hassan FU, Jamal M (2023) Designing multi-epitope monkeypox virus-specific vaccine using immunoinformatics approach. J Infect Public Health 16(1):107–116
74. Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC et al (2021) The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. Nat Microbiol 6(7):899–909
75. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ et al (2020) Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med 12:68
76. Marthandan N, Kanakasabai S, Pavithra R, Mohan V (2021) Epitope-based vaccine designing approach against viral infections: An overview of recent advancements. VirusDisease 32(4):375–386
77. Nayak AK, Chakraborty A, Shukla S, Kumar N, Samanta S (2024) An immunoinformatic approach for developing a multi-epitope subunit vaccine against monkeypox virus. In Silico Pharmacol 12(1):42
78. Kaur G, Dufour JM (2012) Cell culture media formulation and optimization: challenges and strategies for successful mammalian cell culture. Biotechnol Adv 30(6):1448–1460
79. Li D, Li Y, Zhang W, Yang H, Zhao X, Zhang H (2019) Designing an epitope-based peptide vaccine against chikungunya virus. Biomed Res Int 2019:8306518
80. Ternette N, Busch R, Rammensee HG (2015) Cleavage prediction improves the immunogenicity of HLA class I epitopes. Mol Immunol 68(2):514–519
81. Rathore AS, Kapil P, Sharma RK (2019) Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. Drug Discovery Today 24(3):634–649. https://doi.org/10.1016/j.drudis.2020.03.006
82. Berenger F, Kumar A, Zhang KYJ, Yamanishi Y (2021) Lean-docking: exploiting ligands' predicted docking scores to accelerate molecular docking. J Chem Inf Model 61(5):2341–2352. https://doi.org/10.1021/acs.jcim.0c01452
83. Aktaş E, Saygılı İ, Kahveci E, Tekbıyık Z, Özgentürk NÖ (2023) Bioinformatic investigation of Nipah virus surface protein mutations: molecular docking with Ephrin B2 receptor, molecular dynamics simulation, and structural impact analysis. Microbiol Immunol 67(12):501–513. https://doi.org/10.1111/1348-0421.13098
84. Poudyal M, Patel K, Gadhe L, Sawner AS, Kadu P, Datta D et al (2023) Intermolecular interactions underlie protein/peptide phase separation irrespective of sequence and structure at crowded milieu. Nat Commun 14(1):6199. https://doi.org/10.1038/s41467-023-41864-9
85. Ghobadi Z, Mahnam K, Shakhsi-Niaei M (2022) In-silico design of peptides for inhibition of HLA-A*03-KLIETYFSK complex as a new drug design for treatment of multiple sclerosis disease. J Mol Graph Model 111:108079. https://doi.org/10.1016/j.jmgm.2021.108079
86. Wang CY, Peng WJ, Kuo BS, Ho YH, Wang MS, Yang YT et al (2023) Toward a pan-SARS-CoV-2 vaccine targeting conserved epitopes on spike and non-spike proteins for potent, broad and durable immune responses. PLoS Pathog 19(4):e1010870. https://doi.org/10.1371/journal.ppat.1010870
87. Zhao Y, Zeng C, Massiah MA (2015) Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. PLoS ONE 10(4):e0124377
88. Liu W, Liu R, Qin Q et al (2024) Interaction mechanisms of ACE inhibitory peptides: molecular docking and molecular dynamics simulation studies on five wheat gluten derived peptides. Eur Food Res Technol 250:2133–2146. https://doi.org/10.1007/s00217-024-04526-8
89. Yadav PK, Antonyraj CB, Ahamed SIB, Srinivas S (2017) MD simulation analysis in terms of H-Bond, SASA, RMSD, and Rg for thrombin. PLOS ONE. Figure. https://doi.org/10.1371/journal.pone.0181216.g008
90. Weber JK, Morrone JA, Kang S, Zhang L, Lang L, Chowell D, Krishna C, Huynh T, Parthasarathy P, Luan B, Alban TJ, Cornell WD, Chan TA (2024) Unsupervised and supervised AI on molecular dynamics simulations reveals complex characteristics of HLA-A2-peptide immunogenicity. Brief Bioinform 25(1):bbad504. https://doi.org/10.1093/bib/bbad504
91. Pandey A, Madan R, Singh S (2022) Immunology to immunotherapeutics of SARS-CoV-2: identification of immunogenic epitopes for vaccine development. Curr Microbiol 79:306. https://doi.org/10.1007/s00284-022-03003-3
92. Song J, Wang M, Zhou L et al (2023) A candidate nanoparticle vaccine comprised of multiple epitopes of the African swine fever virus elicits a robust immune response. J Nanobiotechnol 21:424. https://doi.org/10.1186/s12951-023-02210-9
93. Xu Y, Zhu F, Zhou Z et al (2024) A novel mRNA multi-epitope vaccine of Acinetobacter baumannii based on multi-target protein design in immunoinformatic approach. BMC Genom 25:791. https://doi.org/10.1186/s12864-024-10691-7
94. Adam KM (2021) Immunoinformatics approach for multi-epitope vaccine design against structural proteins and ORF1a polyprotein of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). Trop Dis Travel Med Vaccines 7:22. https://doi.org/10.1186/s40794-021-00147-1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.