International Journal of Peptide Research and Therapeutics (2023) 29:60
https://doi.org/10.1007/s10989-023-10535-0

REVIEW PAPER

Advances in Computational and Bioinformatics Tools and Databases for Designing and Developing a Multi-Epitope-Based Peptide Vaccine

Mohammad Mahfuz Ali Khan Shawan[1] · **Ashish Ranjan Sharma**[2] · **Sajal Kumar Halder**[1] · **Tawsif Al Arian**[3] · **Md. Nazmussakib Shuvo**[4] · **Satya Ranjan Sarker**[5] · **Md. Ashraful Hasan**[1]

Accepted: 11 May 2023 / Published online: 23 May 2023

Abstract

A vaccine is defined as a biologic preparation that trains the immune system, boosts immunity, and protects against a deadly microbial infection. They have been used for centuries to combat a variety of contagious illnesses by means of subsiding the disease burden as well as eradicating the disease. Since infectious disease pandemics are a recurring global threat, vaccination has emerged as one of the most promising tools to save millions of lives and reduce infection rates. The World Health Organization reports that immunization protects three million individuals annually. Currently, multi-epitope-based peptide vaccines are a unique concept in vaccine formulation. Epitope-based peptide vaccines utilize small fragments of proteins or peptides (parts of the pathogen), called epitopes, that trigger an adequate immune response against a particular pathogen. However, conventional vaccine designing and development techniques are too cumbersome, expensive, and time-consuming. With the recent advancement in bioinformatics, immunoinformatics, and vaccinomics discipline, vaccine science has entered a new era accompanying a modern, impressive, and more realistic paradigm in designing and developing next-generation strong immunogens. *In silico* designing and developing a safe and novel vaccine construct involves knowledge of reverse vaccinology, various vaccine databases, and high throughput techniques. The computational tools and techniques directly associated with vaccine research are extremely effective, economical, precise, robust, and safe for human use. Many vaccine candidates have entered clinical trials instantly and are available prior to schedule. In light of this, the present article provides researchers with up-to-date information on various approaches, protocols, and databases regarding the computational designing and development of potent multi-epitope-based peptide vaccines that can assist researchers in tailoring vaccines more rapidly and cost-effectively.

Keywords Vaccine · Microbial infection · Epitope-based peptide vaccine · Immune response · Immunoinformatics · Vaccinomics · Computational tools · Databases

Mohammad Mahfuz Ali Khan Shawan and Ashish Ranjan Sharma have contributed equally to this work.

✉ Mohammad Mahfuz Ali Khan Shawan
mahfuz_026shawan@juniv.edu

✉ Ashish Ranjan Sharma
researchskeletal@gmail.com

✉ Md. Ashraful Hasan
ashrafulhasan@juniv.edu

[1] Department of Biochemistry and Molecular Biology, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

[2] Institute for Skeletal Aging & Orthopedic Surgery, Hallym University-Chuncheon Sacred Heart Hospital, Chuncheon-si 24252, Gangwon-do, Republic of Korea

[3] Department of Pharmacy, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

[4] Department of Botany, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

[5] Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh

Introduction

A vaccine is an immunobiological substance from a disease-causing pathogen that triggers the immune system to elicit an effective immune response against that specific pathogen (Khan et al. 2022a). They destroy the lethality of an infectious microorganism analogous to natural immunity (Dey et al. 2022a). Infectious diseases caused by microbial pathogens like viruses, bacteria, and fungi are globally responsible for increased morbidity and mortality (Mahapatra et al. 2022a). To date, over 6.8 million people have already died of COVID-19 (Coronavirus Disease 2019) pandemic caused by SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) (Sahoo et al. 2022), and the death toll is increasing day by day (Shawan et al. 2021a, b). Other viruses, such as HIV (Human Immunodeficiency Virus), Ebola, Zika, Dengue, etc., have a horrendous death rate and are walking on the same track (Xin et al. 2023). Besides viruses, deadly bacteria are also responsible for numerous infectious diseases (Dey et al. 2022a, b; Khan et al. 2022a, b). Upon getting the chance, commensal bacteria like *Staphylococcus aureus* may become slaughterous (Oh et al. 2016). In this context, vaccines are a blessing through medicine and act as a game changer by offering protection against various deadly infectious diseases, saving millions of lives. They have raised life expectancy in developed and underdeveloped countries (Xin et al. 2023).

At present, multi-epitope-based peptide vaccine design and development is an emerging area of research that focuses on using specific components of a pathogen, known as epitopes, to create a vaccine (Abass et al. 2022). Epitopes are short amino acid sequences recognized by the immune system and trigger an immune response. By using epitopes, vaccines can be designed to target specific parts of a pathogen, leading to a more targeted and effective immune response (Shawan et al. 2014). Epitope-based peptides are desired vaccine candidates due to their simpler production, non-infectious property, and chemical stability (Obaidullah et al. 2021). One of the main promises of epitope-based peptide vaccine design is its potential for relatively quick, cheap, and rapid development, as it only requires the production of a small number of antigenic peptides rather than the entire pathogen, making them ideal for use in response to emerging infectious diseases. Another advantage of this type of vaccine is its potential for improved safety (Dey et al. 2022a). Traditional vaccines use either inactivated or attenuated forms of the pathogen, which can cause adverse reactions and/or autoimmune responses in some individuals. On the other hand, epitope-based peptide vaccines are biologically harmless and highly effective at eliciting the desired immune response (Purcell et al. 2007; Kar et al. 2020; (Mahapatra et al. 2022a). The molecular mechanism of action of an epitope-based peptide vaccine is depicted in Fig. 1 (Kar et al. 2020).

The natural immune response can be triggered/evoked by entire or parts of microorganisms that may act as antigens, which can elicit a host's immune response and produce antibodies against those antigens. Antigenicity is the capacity of an antigen to react with a particular antibody and is linked to immunoreactivity and/or immunogenicity. Immunoreactivity and/or immunogenicity is a complex network of antigen-specific biological reactions mediated by the humoral immunity of the host's adaptive immune system (Shawan et al. 2014). During the exposure of an antigen to the immune system, B-cells are stimulated and differentiated into plasma cells with the aid of CD4+helper T-cells, producing antigen-specific antibodies (Nicholson et al. 2016). In addition, the immune system also relies on CD8+cytotoxic T-cells and IFN (Interferons, a group of cytokines) along with CD4+helper T-cells to neutralize the antigen. The T-cell-mediated immune response deeply relies on the MHC (Major Histocompatibility Complex) molecules and is analogous to the binding of an antigen with its specific antibody. The human leukocyte antigen (HLA) gene encodes MHC peptide molecules. Every HLA allele stands for a peptide set found on the infected cell surface and identified by the receptors on T-cells (TCRs). Thus, both T-cell and B-cell subsequently provide cellular and humoral immunity, which are critically needed to evoke an effective immune response (Rakib et al. 2020).

The conventional approach to designing and developing an efficient vaccine candidate requires identifying target antigens, conducting in-depth research, and establishing an immunological correlation with the vaccine construct (Rappuoli et al. 2019). Traditional/experimental approach toward vaccine development is time-consuming, expensive, fraught with challenges, and requires the cultivation of large amounts of the pathogen. The process typically takes significant time to construct a commercially viable vaccine and involves a high rate of failure. That is why researchers are extremely interested in designing and developing vaccines using computer-assisted tools and techniques (Obaidullah et al. 2021). Recent research has shown that *in silico* approaches toward vaccine design are much more effective than earlier methods (Pyasi et al. 2021). Using novel resources (computational tools, techniques, and databases) and similar bioinformatics strategies, this process successfully establishes potent vaccine candidates that can induce strong immune responses against different types of human infectious pathogens like viruses [i.e., SARS-CoV-2 (Srivastava et al. 2022), mammarenavirus (Khan et al. 2022b) etc.], bacteria [i.e., *Achromobacter xylosoxidans* (Khan et al. 2022a), *Enterococcus faecium* (Dey et al. 2022a), *Klebsiella pneumoniae* (Dey et al. 2022b), *Acinetobacter baumannii*

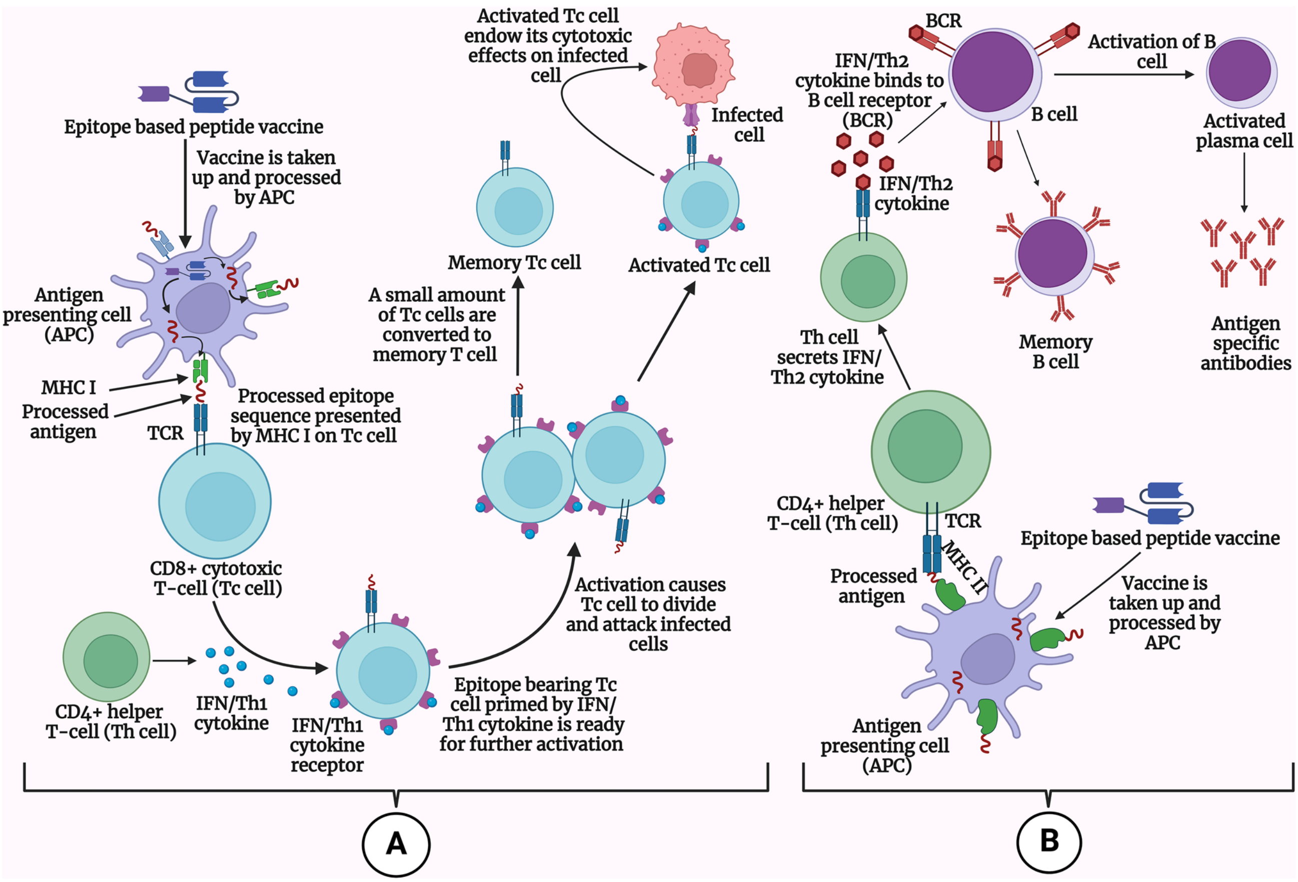


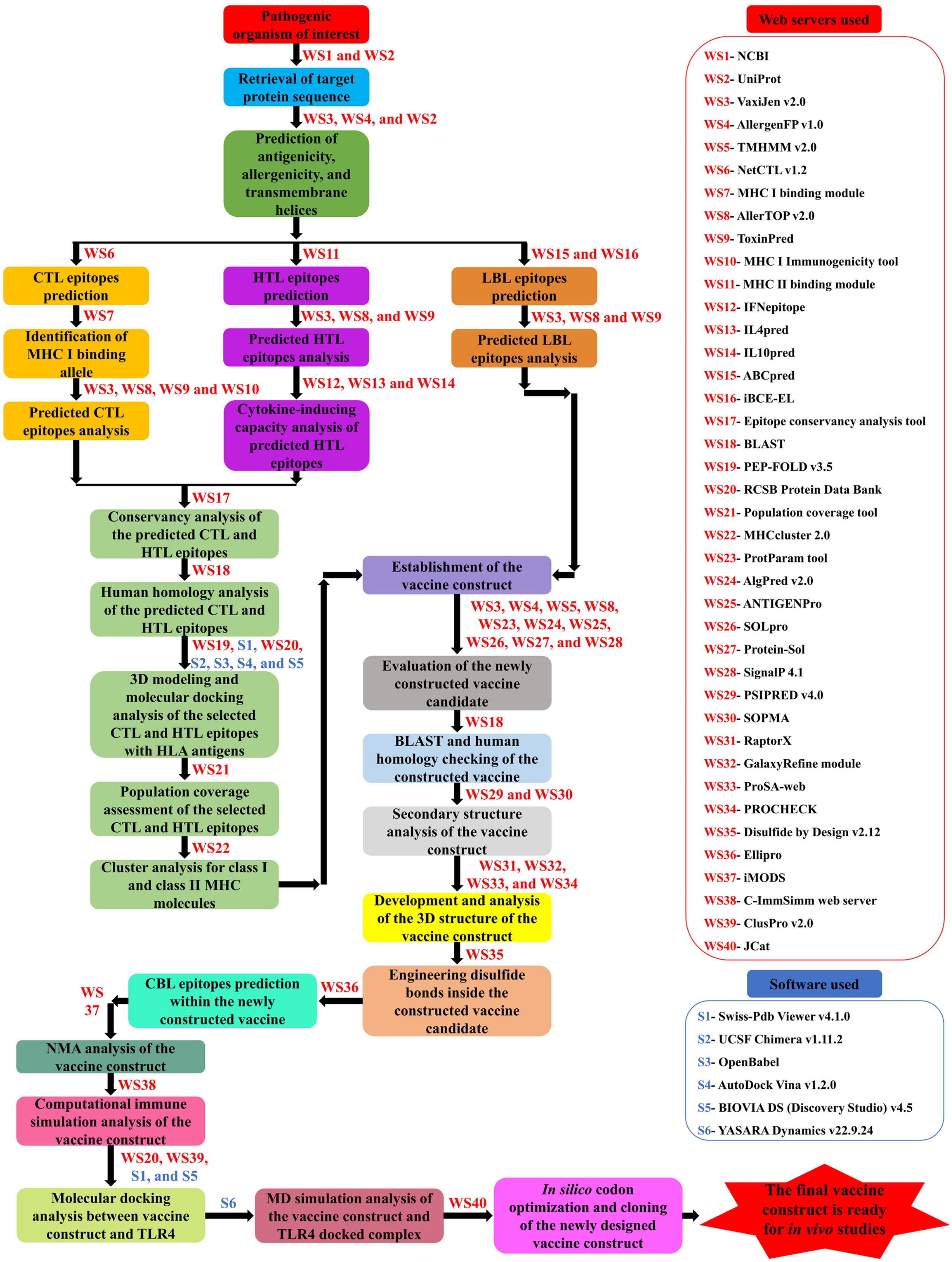
Fig. 1 Molecular mechanism of action of epitope-based peptide vaccine triggering cellular and humoral immunity. **(A)** The vaccine is taken up, processed, and presented by antigen-presenting cells (APC) with the help of the MHC I receptor to the T-cell receptor (TCR) of CD8^{+} cytotoxic T-cell (Tc-cell). This interaction activates the Tc-cell development and elicits the production of IFN/Th1 cytokines by CD4^{+} helper T-cell (Th-cell). IFN/Th1 cytokine results in the activation of Tc-cells to divide and attack the infected cell. The activated Tc-cells are also converted to memory Tc-cells. **(B)** Likewise, the antigenic vaccine is taken up, processed, and presented by MHC II of APC to TCR of Th-cell. This causes Th-cell activation, resulting in the secretion of IFN/Th2 cytokines. IFN/Th2 cytokine activates B-cells which differentiate into activated plasma cells and memory B-cells. Activated plasma cells and memory B-cells are capable of producing antigen-specific antibodies that can neutralize an infection. This figure was generated using BioRender.com

(Mahapatra et al. 2022a) etc.], as well as fungi [i.e., *Candida auris* (Khan et al. 2022c). Creating a safe and new vaccine using *in silico* design and development requires expertise in reverse vaccinology, multiple vaccine databases, and high-throughput methods. Databases such as Cytomegalovirus-db, Mammarenavirus-db, Hantavirus-db etc., are the repository of valuable information regarding experimentally validated vaccine components ((Khan et al. 2021a; (Khan et al. 2021a; (Khan et al. 2021a). In contrast, high-throughput methods are potent bioinformatics protocols to anticipate novel vaccine candidates (Srivastava et al. 2022). Furthermore, peptide candidates as potent epitope vaccines having improved expression patterns can be detected by *in silico* models that use various computational algorithms. These robust and more sophisticated algorithms are the hub for identifying immune epitopes against T and B cells. Various high-throughput screening approaches have already been developed to evaluate a vaccine construct's efficacy (Abass et al. 2022).

In this article, we provide an outline for designing and developing multi-epitope-based peptide vaccines with the aid of different bioinformatics/immunoinformatics tools, database repositories, and computational algorithms in a simple, basic, and straightforward fashion. We expect that developments in bioinformatics and computational technologies will make vaccinology protocols more effective and accessible for researchers, enhancing the future of immunology.

Materials and methods

The complete step-by-step methodology for the *in silico* designing and developing a multi-epitope-based peptide vaccine is visualized in a flow chart in Fig. 2. All the web


Pathogenic organism of interest
WS1 and WS2
Retrieval of target protein sequence
WS3, WS4, and WS2
Prediction of antigenicity, allergenicity, and transmembrane helices
WS6
WS11
WS15 and WS16
CTL epitopes prediction
HTL epitopes prediction
LBL epitopes prediction
WS7
WS3, WS8, and WS9
WS3, WS8 and WS9
Identification of MHC I binding allele
Predicted HTL epitopes analysis
Predicted LBL epitopes analysis
WS3, WS8, WS9 and WS10
WS12, WS13 and WS14
Predicted CTL epitopes analysis
Cytokine-inducing capacity analysis of predicted HTL epitopes
WS17
Conservancy analysis of the predicted CTL and HTL epitopes
WS18
Human homology analysis of the predicted CTL and HTL epitopes
WS19, S1, WS20, S2, S3, S4, and S5
3D modeling and molecular docking analysis of the selected CTL and HTL epitopes with HLA antigens
WS21
Population coverage assessment of the selected CTL and HTL epitopes
WS22
Cluster analysis for class I and class II MHC molecules
Establishment of the vaccine construct
WS3, WS4, WS5, WS8, WS23, WS24, WS25, WS26, WS27, and WS28
Evaluation of the newly constructed vaccine candidate
WS18
BLAST and human homology checking of the constructed vaccine
WS29 and WS30
Secondary structure analysis of the vaccine construct
WS31, WS32, WS33, and WS34
Development and analysis of the 3D structure of the vaccine construct
WS35
Engineering disulfide bonds inside the constructed vaccine candidate
WS36
CBL epitopes prediction within the newly constructed vaccine
WS 37
NMA analysis of the vaccine construct
WS38
Computational immune simulation analysis of the vaccine construct
WS20, WS39, S1, and S5
Molecular docking analysis between vaccine construct and TLR4
S6
MD simulation analysis of the vaccine construct and TLR4 docked complex
WS40
In silico codon optimization and cloning of the newly designed vaccine construct
The final vaccine construct is ready for in vivo studies
Web servers used
WS1- NCBI
WS2- UniProt
WS3- VaxiJen v2.0
WS4- AllergenFP v1.0
WS5- TMHMM v2.0
WS6- NetCTL v1.2
WS7- MHC I binding module
WS8- AllerTOP v2.0
WS9- ToxinPred
WS10- MHC I Immunogenicity tool
WS11- MHC II binding module
WS12- IFNepitope
WS13- IL4pred
WS14- IL10pred
WS15- ABCpred
WS16- iBCE-EL
WS17- Epitope conservancy analysis tool
WS18- BLAST
WS19- PEP-FOLD v3.5
WS20- RCSB Protein Data Bank
WS21- Population coverage tool
WS22- MHCcluster 2.0
WS23- ProtParam tool
WS24- AlgPred v2.0
WS25- ANTIGENPro
WS26- SOLpro
WS27- Protein-Sol
WS28- SignalP 4.1
WS29- PSIPRED v4.0
WS30- SOPMA
WS31- RaptorX
WS32- GalaxyRefine module
WS33- ProSA-web
WS34- PROCHECK
WS35- Disulfide by Design v2.12
WS36- Ellipro
WS37- iMODS
WS38- C-ImmSimm web server
WS39- ClusPro v2.0
WS40- JCat
Software used
S1- Swiss-Pdb Viewer v4.1.0
S2- UCSF Chimera v1.11.2
S3- OpenBabel
S4- AutoDock Vina v1.2.0
S5- BIOVIA DS (Discovery Studio) v4.5
S6- YASARA Dynamics v22.9.24

◂**Fig. 2** A schematic illustration exhibiting the overall systematic immunoinformatics strategy/approach adopted for the *in silico* epitope curation, designing, and development of a multi-epitope-based peptide vaccine. Initially, an antigenic target protein sequence from a desired microbe (virus, bacteria, fungi etc.) is extracted to select promiscuous T-cell (Tc and Th) and B-cell (LBL) epitopes. Appropriate linkers can then join these novel epitopes to construct a multi-epitope-based peptide vaccine candidate. After the evaluation (BLAST, disulfide engineering, CBL epitope prediction, NMA, and immune simulation) and structural assessment (2D and 3D), the newly formulated vaccine construct can be subjected to molecular docking analysis with TLR4 immune receptor. A molecular dynamics simulation is carried out to predict the stability of the docked complex. This flowchart is generated using Microsoft Office (PowerPoint) 2019

addresses with additional comments on different servers/databases and software that are used in the vaccinomics approach are listed in Tables 1 and 2.

Retrieval of Target Protein Sequence

The amino acid sequence of the target protein from desired pathogenic microbes can be acquired using different protein databases like National Center for Biotechnology Information (NCBI) (Database resources of the NCBI 2016) or UniProt (The UniProt Consortium 2021). This retrieved amino acid sequence is used to generate a novel vaccine construct. The NCBI and UniProt databases provide a huge amount of biological protein information (Narang et al. 2021; Panda et al. 2022). The amino acid sequence of the target protein can be extracted in FASTA format (Shawan et al. 2014, 2018).

Target Protein Sequence Analysis

Considering the default threshold value, the target protein's antigenicity can be determined using the VaxiJen v2.0 web server (Shawan et al. 2014). Afterward, allergenicity of the target protein can be detected using AllergenFP v1.0 server (Dimitrov et al. 2014b). Later, the TMHMM v2.0 server can be used to predict the target protein's transmembrane (TM) helices (Doytchinova and Flower 2007). Ultimately, non-allergic and highly antigenic amino acid sequences with less TM helicase are selected for further evaluation (Dey et al. 2022a).

Prediction and Analysis of CTL (Cytotoxic T Lymphocyte) Epitopes

CTL Epitopes Prediction

Within the immune system, CTLs interact and kill the infectious cell, thus playing a crucial role in the host's defense mechanism. To detect the CTL epitopes within a target protein, NetCTL v1.2 server can be used, which anticipates 9-mer epitopes against 12 HLA antigen allele class I supertypes (A1, A2, A3, A24, A26, B7, B8, B27, B39, B44, B58, and B62). Taking the default threshold values (C terminal cleavage- 0.15, epitope identification- 0.75, and antigen processing transport efficiency- 0.05) in consideration, this tool detects epitopes with great precision, and the CTL epitopes having the highest combined score are then selected for further analysis (Larsen et al. 2007).

Identification of MHC I Binding Allele

After the detection of CTL epitopes, the MHC I binding allele for each of the epitopes can be identified using MHC I binding module within IEDB (Immune Epitope Database) server. A consensus percentile rank score of less than or equal to 2.0 is usually considered to choose effective CTL epitopes, as a lower rank score represents higher affinity (Moutaftsi et al. 2006).

Predicted CTL Epitopes Analysis

Afterward, each of the refined CTL epitopes can be analyzed for antigenicity, allergenicity, toxicity, and immunogenicity through VaxiJen v2.0, AllerTOP v2.0, ToxinPred, and IEDB MHC I Immunogenicity tool of IEDB server respectively (Doytchinova and Flower 2007; Gupta et al. 2013; Calis et al. 2013; Dimitrov et al. 2014a). The CTL epitopes, which are highly antigenic, non-toxic, non-allergenic, and extremely immunogenic, are considered for vaccine preparation.

Prediction and Analysis of HTL (Helper T Lymphocyte) Epitopes

HTL Epitopes Prediction

HTLs are a crucial part of the adaptive immune system as they can identify foreign antigens and stimulate B-cell proliferation and CTLs to eliminate the infectious entity. HTL epitopes within a desired protein sequence can be forecasted through the MHC II binding tool from the IEDB server. This module detects 15-mer epitopes against HTLs, while a consensus percentile rank score equal to or less than 2.0 can be used as a threshold to anticipate efficient HTL epitopes. As for MHC I binding module, a lower percentile score suggests a higher binding affinity in this module (Wang et al. 2010).

Predicted HTL Epitopes Analysis

Each of the selected HTL epitopes can then be scrutinized for antigenicity, allergenicity, and toxicity using VaxiJen v2.0, AllerTOP v2.0, and ToxinPred server, respectively

Table 1 Web addresses with additional comments on different servers/databases that are implemented for *in silico* vaccine discovery process

Server/Database	Web address	Additional comment	Reference
National Center for Biotechnology Information (NCBI)	https://www.ncbi.nlm.nih.gov/	The NCBI is a wing of NIH (National Institute of Health) and is integrated with USNLM (the United States National Library of Medicine). This center is accredited and supported by the United States of America and houses a huge collection of significant resources related to biomedicine and biotechnology. Most important databases, bioinformatics tools, and computational services include GenBank, PubMed, BLAST, Gene, Genome, BioProject, Nucleotide, Protein, SNP etc., which are accessible online with Entrez search engine.	Database resources of the NCBI 2016
UniProt	https://www.uniprot.org/	UniProt database is a repository of high-quality protein sequences and functional information. Most of the information within this freely accessible database is mainly derived from various genome sequencing projects. This is maintained by the UniProt consortium.	The UniProt Consortium 2021
VaxiJen v2.0	http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html	VaxiJen v2.0 utilizes a new alignment-independent model to predict protective antigens considering ACC (Auto Cross Covariance) transformation of amino acid sequences and generating 70–89% accurate results.	Doytchinova and Flower 2007
AllergenFP v1.0	https://ddg-pharmfac.net/AllergenFP/index.html	With 88.9% accuracy, AllergenFP v1.0 server exploits a new alignment-free descriptor-based fingerprint approach to predict any amino acid sequence allergenicity.	Dimitrov et al. 2014b
TMHMM v2.0	https://services.healthtech.dtu.dk/service.php?TMHMM-2.0	TMHMM v2.0 predicts membrane protein topology based on HMM (Hidden Markov Model) and displays 97–98% accurate results. With a high degree of sensitivity and specificity (more than 99%), this server can differentiate between membrane and soluble proteins.	Krog et al. 2001
NetCTL v1.2	http://www.cbs.dtu.dk/services/NetCTL/	NetCTL v1.2 server has a high prediction capacity and maintains a threshold point of 0.90 with a sensitivity and specificity of 0.74 and 0.98, respectively. This server precisely detects the epitopes on the basis of two important neural networks: ANN (Artificial Neural Network) and WMNN (Weight Matrix Neural Network).	Larsen et al. 2007

Table 1 (continued)

Server/Database	Web address	Additional comment	Reference
MHC I binding module within the IEDB server	http://tools.iedb.org/mhci/	The IEDB server is a freely accessible resource with research data on B-cell, T-cell, and antibodies from humans and other species in the context of allergy, infectious disease, and autoimmunity. This database depends on the CONSENSUS algorithm to anticipate T and B cell epitopes.	Moutaftsi et al. 2006
AllerTOP v2.0	https://www.ddg-pharmfac.net/AllerTOP/	AllerTOP v2.0 server isolate allergens from non-allergens with a detection accuracy of 85.3% at 5-fold cross-validation. This database applies amino acid descriptors, ACC transformation method, and kNN (k-Nearest Neighbour) algorithm to detect allergens.	Dimitrov et al. 2014a
ToxinPred	http://crdd.osdd.net/raghava/toxinpred/	To predict the toxicity of different peptides, the ToxinPred server utilizes SVM (Support Vector Machine) algorithm along with a quantitative matrix model.	Gupta et al. 2013
MHC I Immunogenicity tool of IEDB server	http://tools.iedb.org/immunogenicity/	MHC I Immunogenicity tool of the IEDB server can only detect the immunogenicity of 9-mer peptide epitopes.	Calis et al. 2013
MHC II binding module within IEDB server	http://tools.iedb.org/mhcii/	To predict HTL epitopes efficiently, this tool depends on the widely accepted CONSENSUS method.	Wang et al. 2010
IFNepitope	http://crdd.osdd.net/raghava/ifnepitope/	IFNepitope server is used to detect IFN γ inducing regions within peptide molecules considering three classical models, namely motif-based, SVM-based, and hybrid approach-based algorithms. This tool searches for an IFN γ inducing epitope from an in-house dataset of 10,433 already established IFN γ inducing peptides.	Wang et al. 2008
IL4pred	https://webs.iiitd.edu.in/raghava/il4pred/	The operation of IL4pred server is based on an SVM algorithm with a prediction accuracy of 75.76%.	Dhanda et al. 2013
IL10pred	https://webs.iiitd.edu.in/raghava/il10pred/	The operation of the IL10pred server is based on an SVM algorithm with a prediction accuracy of 81.24%.	Nagpal et al. 2017
ABCpred	https://webs.iiitd.edu.in/raghava/abcpred/	ABCpred server identifies linear B-cell epitopes in a protein sequence considering PRNN (Partial Recurrent Neural Network) algorithm along with a single hidden layer. This database compares the query protein sequence with 2479 established continuous B-cell epitopes in the B-cell epitope database.	Saha and Raghava 2007

Table 1 (continued)

Server/Database	Web address	Additional comment	Reference
iBCE-EL	http://www.thegleelab.org/iBCE-EL/	iBCE-EL server consists of 9925 epitopes (4440 BCEs and 5485 non-BCEs) and predicts potential BCE considering six different ML algorithms (RF, ERT, AB, GB, SVM, and kNN).	Manavalan et al. 2018
Epitope conservancy analysis tool	http://tools.iedb.org/conservancy/	The epitope conservancy analysis tool within the IEDB resource calculates the extent of conservancy for a single or batch of epitopes or protein sequences at a given identity level. Currently, two modes of conservancy analysis are available, including linear and discontinuous epitope analysis.	Bui et al. 2007
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi	blastp module basically searches conserved domain databases based on PSSM (Position Specific Scoring Matrix).	Altschul et al. 1990
PEP-FOLD v3.5	https://bioserv.rpbs.univ-paris-diderot.fr/services/PEPFOLD3/	By applying the Taboo/Backtract sampling algorithm, PEP-FOLD v3.5 server predicts the conformation of small peptides with 5–50 amino acid residues in length.	Lamiable et al. 2016
RCSB Protein Data Bank	https://www.rcsb.org/	RCSB (Research Collaboratory for Structural Bioinformatics) PDB is a freely accessible database that provides knowledge about annotated 3D structures of biological macromolecules i.e. nucleic acids, proteins etc. This knowledge is required to study different biological macromolecules' shapes and molecular interactions.	Berman et al. 2002
Population coverage tool in IEDB server	http://tools.iedb.org/population/	The population coverage tool in the IEDB server utilizes HLA genotyping frequencies from the allele frequency database to assess the distribution with the expression pattern of HLA alleles. Presently, the allele frequency database has data on 115 countries and 21 different ethnic groups.	Bui et al. 2006
MHCcluster 2.0	https://services.healthtech.dtu.dk/service.php?MHCcluster-2.0	For MHC cluster assessment through MHCcluster 2.0, the NetMHCpan-2.8 prediction method can be utilized, where HLA prevalent and characterized module and HLA-DRB representative module can be selected for HLA class I and class II cluster analysis, respectively.	Thomsen et al. 2013
ProtParam tool within ExPASy server	https://web.expasy.org/protparam/	ProtParam program analyzes different protein parameters for a given protein and computes various physical and chemical properties by comparing proteomics datasets deposited in Swiss-Port or TrEMBL.	Gasteiger et al. 2005

Table 1 (continued)

Server/Database	Web address	Additional comment	Reference
AlgPred v2.0	https://webs.iiitd.edu.in/raghava/algpred2/	To determine the allergenicity of a given peptide sequence, AlgPred v2.0 utilizes SVM machine learning technology along with BLAST, motifs mapping, MEME/MAST, and IgE epitope mapping. For this task, a large dataset of 10,075 allergens and non-allergens are considered.	Saha and Raghava 2006
ANTIGENPro	https://scratch.proteomics.ics.uci.edu/	For antigenicity prediction, the ANTIGENPro tool in the SCRATCH protein prediction server utilizes cross-validation experiments within the combined dataset and displays 78% accuracy.	Magnan et al. 2010
SOLpro	http://scratch.proteomics.ics.uci.edu/	SOLpro within the SCRATCH protein prediction server predicts the solubility of a peptide using a two-stage SVM algorithm.	Magnan et al. 2009
Protein-Sol	https://protein-sol.manchester.ac.uk/	Protein-Sol calculates the solubility of a protein by comparing its solubility with *E. coli* proteins' average solubility from the PopAvrSol dataset.	Hebditch et al. 2017
SignalP 4.1	http://www.cbs.dtu.dk/services/SignalP-4.1/	SignalP 4.1 apply a deep neural network-based approach to detect signal peptides.	Nielsen 2017b
PSIPRED v4.0	http://bioinf.cs.ucl.ac.uk/psipred/	PSIPRED v4.0 predicts a protein secondary structure from PSI-BLAST by adopting two feed-forward neural network algorithms.	Buchan et al. 2013
SOPMA	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html	SOPMA is widely used for predicting the secondary structure of proteins. It applies a neural network algorithm and utilizes a statistical analysis of protein structures of the protein data bank.	Geourjon and Deléage 1995
RaptorX	http://raptorx.uchicago.edu/	RaptorX forecasts protein structure and function using the deep learning method. It generates a 3D structure against a query protein based on numerous template threading techniques.	Wang et al. 2016
GalaxyRefine module within the GalaxyWEB server	https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE	To refine a 3D model, the GalaxyRefine module into the GalaxyWEB server depends on the CASP9 method along with HHsearch and PROMALS3D program at Seok-server.	Ko et al. 2012
ProSA-web	https://prosa.services.came.sbg.ac.at/prosa.php	ProSA-web is a protein structure-analyzing tool used for validating protein models.	Wiederstein and Sippl 2007

Table 1 (continued)

Server/Database	Web address	Additional comment	Reference
Verify3D	https://www.doe-mbi.ucla.edu/verify3d/	Verify3D application assesses a protein model's compatibility by generating an atomic (3D) profile calculated from correct protein structures with its own amino acid sequences (1D). The analysis is carried out by accrediting a structural class based on its location and environment (loop, polar, alpha, beta, nonpolar etc.) and comparing the findings with a good structure. A query model with a high Verify3D score is considered well-refined.	Eisenberg et al. 1997
ERRAT	https://www.doe-mbi.ucla.edu/errat/	ERRAT appraises the overall quality of a pre-refined protein model (3D structure) by considering the pattern of different non-bonded atomic interactions. This program is too sensitive to small-scale errors and results in notable improvement in the final model than the initial one. A model of interest with a higher ERRAT score (> 50) is accepted as a high-quality model.	Colovos and Yeates 1993
PROCHECK	https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/	PROCHECK is a protein structure validating application used for Ramachandran plot determination and visualization.	Morris et al. 1992
Disulfide by Design v2.12	http://cptweb.cpt.wayne.edu/DbD2/	Disulfide by Design v2.12 uses a modified DbD algorithm to engineer disulfide linkages by predicting amino acid pairs that more likely form a disulfide bond if mutated to Cys (cysteines).	Craig and Dombkowski 2013
ElliPro: Antibody Epitope Prediction tool within the IEDB analysis resource	http://tools.iedb.org/ellipro/	Using residue clustering, the ElliPro tool predicts the discontinuous B-cell epitopes inside a 3D protein model and modifies Thornton's algorithm.	Ponomarenko and Bourne 2007
iMODS	https://imods.iqfr.csic.es/	For the NMA of a biological macromolecule, iMODS uses an improved affine model-based domain dynamics algorithm.	López-Blanco et al. 2014
C-ImmSim web server	http://150.146.2.1/CIMMSIM/index.php	The C-ImmSim application hired a machine learning approach based on PSSM (Position Specific Protein Matrix).	Dellagostin et al. 2017
ClusPro v2.0	https://cluspro.bu.edu/login.php	The ClusPro v2.0 server utilizes a web-based automated program with three complex steps to simulate protein-protein/protein-peptide molecular docking. Those steps are structural rearrangement by energy minimization, clustering of lowest energy structure, and rigid body docking.	Land and Humble 2018

Table 1 (continued)

Server/Database	Web address	Additional comment	Reference
JCat	http://www.jcat.de/	The freely available JCat database is integrated with the PRODORIC server.	Alam et al. 2019

(Doytchinova and Flower 2007; Gupta et al. 2013; Dimitrov et al. 2014a). Later on, extremely antigenic, non-allergic, and non-toxic epitopes against HTLs can further be considered to check their cytokine-inducing capacity.

Cytokine-inducing Capacity Analysis of Predicted HTL Epitopes

In microbial infection, interferon-gamma (IFN γ) plays a pivotal role in specific and innate immune responses with the activation of macrophages and natural killer cells. IFNepitope server can be applied to predict and design potent IFN γ inducing MHC II binding HTL epitopes with an accuracy of 81.39% (Wang et al. 2008; Ashrafi et al. 2019). The interleukin-4 (IL-4) and interleukin-10 (IL-10) inducing ability of the selected HTL epitopes can be evaluated by IL4pred and IL10pred servers, respectively, with a threshold value of 0.2 and − 0.3 (Dhanda et al. 2013; Nagpal et al. 2017). After the analysis, HTL epitopes having all three cytokine-inducing capacities are chosen to construct the final vaccine candidate.

Prediction and Analysis of LBL (Linear B Lymphocyte) Epitopes

LBL Epitopes Prediction

Antigens having epitopes capable of eliciting B-cell response are critical mediators for antibody-associated humoral immunity. ABCpred server is the most popular one to identify LBL epitopes within a given set of protein sequences with a threshold of 100 for sensitivity, specificity, and accuracy (Saha and Raghava 2007). Subsequently, the probability score of each of the LBL epitopes can be predicted using iBCE-EL server considering default parameters (Manavalan et al. 2018).

Predicted LBL Epitopes Analysis

The predicted LBL epitopes' antigenicity, allergenicity, and toxicity can be assessed through VaxiJen v2.0, AllerTOP v2.0, and ToxinPred server, respectively, accepting default parameters (Doytchinova and Flower 2007; Gupta et al. 2013; Dimitrov et al. 2014a). LBL epitopes having good scores are then chosen for vaccine construction.

Conservancy Analysis of the Predicted CTL and HTL Epitopes

The conservancy (conservation across antigens) of the previously selected MHC I and MHC II epitopes can be analyzed with the help of the epitope conservancy analysis tool under the hood of the epitope analysis tool in the IEDB server. For

Table 2 Web addresses with additional comments on different software that are implemented in *in silico* vaccine discovery process

Software	Web address	Additional comment	Reference
Swiss-Pdb Viewer v4.1.0	https://swiss-pdb-viewer.software.informer.com/4.1/	Swiss-Pdb Viewer calculates and minimizes a protein molecule's energy in addition to restoring distorted geometries utilizing GROMOS 43B1 force field.	Guex and Peitsch 1997
UCSF Chimera v1.11.2	https://www.cgl.ucsf.edu/chimera/download.html	The program UCSF Chimera is developed to study supramolecular assemblies, density maps, docking scores, sequence alignments, and MD (Molecular Dynamics) trajectories of biological macromolecules.	Pettersen et al. 2004
OpenBabel	https://sourceforge.net/projects/openbabel/	OpenBabel is an open-access software that helps search, analyze, convert, and store various chemical and biochemical data using the SMARTS method.	O'Boyle et al. 2011
AutoDock Vina v1.2.0	https://vina.scripps.edu/	AutoDock Vina v1.2.0 is a docking engine within the AutoDock suite program suitable for efficient protein-ligand interaction analysis. This tool has another program called AutoGrid which is valuable for calculating grids within a docked structure.	Trott and Olson 2009
BIOVIA DS (Discovery Studio) v4.5	https://discover.3ds.com/discovery-studio-visualizer-download	Discovery Studio v4.5 is a program that assists in simulating small and macromolecular systems considering CHARM, MODELLER, DELPHI, and DMol3 algorithms.	Accelrys Software Inc: San Diego 2012
YASARA Dynamics (v22.9.24)	http://www.yasara.org/downloads.htm	YASARA Dynamics (v22.9.24) is authorized by PVL (Portable Vector Language) and uses NOVA, YAMBER, and AMBER force fields to conduct an MD simulation.	Land and Humble 2018
SnapGene v6.1	https://www.snapgene.com/	SnapGene v6.1 is a popular tool used for DNA letter evaluation and manipulation.	Solanki and Tiwari 2018

sequence identity, this tool helps recognize the opening of a single epitope in a range of strains with a threshold value greater than or equal to 100 (Bui et al. 2007). MHC epitopes with 100% maximum identity can be selected to construct a vaccine candidate.

Human Homology Analysis of the Predicted CTL and HTL Epitopes

Identifying homologous epitopes within human proteome is vital to design a potent vaccine, as similar epitopes with humans may hamper eliciting an adequate immune response. The epitope homology to the human proteome can be determined by BLAST (Basic Local Alignment Search Tool) module, mainly blastp (protein BLAST) within the NCBI database. In this analysis, a search for homologous sequences can be done using default parameters by selecting *Homo sapiens* (taxid: 9606) at a threshold e-value of 0.05 (Altschul et al. 1990; Mehla and Ramana 2016). Non-homologous epitopes of humans with an e-value below 0.05 can be selected for vaccine construction (Mehla and Ramana 2016).

3D Modeling and Molecular Docking Analysis of the Selected CTL and HTL Epitopes with HLA Antigens

CTL and HTL Epitopes Modeling

To design a reliable vaccine, evaluating the binding affinity of HLA alleles with CTL and HTL epitopes is crucial and can be done by exploiting molecular docking studies. For that, the epitopes (CTL and HTL) must first be modeled with the sOPEP scheme of the PEP-FOLD v3.5 server employing 200 simulations (Lamiable et al. 2016).

Molecular Docking Between CTL and HTL Epitopes with HLA Alleles

Before molecular docking simulation, the energy of each modeled epitope can be computed and minimized with Swiss-Pdb Viewer v4.1.0 software. The 3D structures with the lowest energy are then considered (Guex and Peitsch 1997). Two widely distributed alleles, namely HLA-A*01:01 and HLA-DRB1*01:01, can be selected to represent MHC I and MHC II alleles to examine the binding affinity with CTL and HTL epitopes. To check the molecular interaction, the 3D X-ray crystallographic structure of HLA-A*01:01 and HLA-DRB1*01:01 can be downloaded in pdb format from the RCSB protein data bank bearing PDB ID of 6AT9 and 1QEW, respectively. To validate the docking simulation, co-crystalized ligands within the PDB structures can be considered the positive control (Berman et al. 2002). The UCSF Chimera v1.11.2 is a freely available software for preparing large protein molecules. The preparation can be done by eliminating attached ligands from the co-crystalized structure and adding hydrogens GM (Gasteiger-Marsili) charges (Pettersen et al. 2004). Afterward, OpenBabel can be used to minimize ligand energy and save both the structure (protein and ligand) files into pdbqt format (O'Boyle et al. 2011). AutoDock Vina v1.2.0 is a widely used, more reliable, and cited software utilized for molecular docking simulation (Rahman et al. 2016). Throughout the molecular interaction analysis, all the parameters can be kept at default, and the grid box for HLA-A*01:01 and HLA-DRB1*01:01 can be set at (X)60.64 × (Y)73.76 × (Z)45.49 Å and (X)61.25 × (Y)48.69 × (Z)72.95 Å respectively. The results of docking studies are denoted as negative values (kcal mol^{-1}), and a lower score indicates strong binding affinity (Trott and Olson 2009). BIOVIA DS (Discovery Studio) v4.5 can be utilized to visualize the molecular docking simulation results, and the figure can be generated using UCSF Chimera (Accelrys Software Inc: San Diego 2012).

Population Coverage Assessment of Selected CTL and HTL Epitopes

The expression and the distribution pattern of HLA alleles (class I and II) differ by ethnic groups and regions around the globe. Population coverage analysis is pivotal for developing an effective epitope-based peptide vaccine. The population coverage of the selected CTL and HTL epitopes can be assessed by the population coverage tool in the IEDB server. After the calculation, predicted CTL and HTL epitopes and their corresponding HLA binding alleles (MHC I, MHC II, and combined) can be analyzed (Bui et al. 2006).

Cluster Analysis for Class I and Class II MHC Molecules

In humans, the genes for both classes of MHC molecules are highly polymorphic, and this extreme polymorphism in HLA antigens encompasses hundreds of thousands of alleles. MHC I and II molecules with similar binding affinity can be recognized by MHC clustering analysis with the help of the MHCcluster 2.0 server. Considering the default parameters, this tool generates phylogenetic trees and excessively intuitive heat-maps of the effective cluster between MHC class I and II molecules (Thomsen et al. 2013).

Establishment of the Vaccine Construct

The effective vaccine construct can be formulated by combining previously selected CTL, HTL, and LBL epitopes that have outperformed others based on different selection criteria with each other. For this addition, CTL, HTL, and LBL epitopes can be linked with AAY (Ala-Ala-Tyr), GPGPG (Gly-Pro-Gly-Pro-Gly), and KK (Lys-Lys) linkers, respectively (Dorosti et al. 2019). The AAY linker improves the immunogenicity of a vaccine candidate by influencing protein stability and epitope presentation capacity. The glycine-proline (GPGPG) and bi-lysine (KK) linker facilitate immune processing and immunogenic activity of the newly constructed vaccine, respectively (Nain et al. 2020). To achieve a stronger immune response, an adjuvant like the 50 S ribosomal protein subunit L7/L12 (TLR4 agonist) can be linked at the starting end of the construct with a bifunctional EAAAK linker (Glu-Ala-Ala-Ala-Lys) (Olejnik et al. 2018).

Evaluation of the Newly Constructed Vaccine Candidate

Physicochemical Property Analysis of the Vaccine Construct

The physicochemical properties, i.e., the number of amino acids, molecular weight (MW), theoretical pH (pI), amino acids composition, the total number of negatively charged residues, the total number of positively charged residues, atomic composition, formula, extinction coefficient, estimated half-life, instability index (II), aliphatic index (AI), and grand average of hydropathicity (GRAVY) of the formulated vaccine can be assessed using ProtParam tool within ExPASy proteomic server (Gasteiger 2003; Narang et al. 2021; Panda et al. 2022).

Allergenicity, Antigenicity, and Solubility Profile Analysis of the Vaccine Construct

A newly designed vaccine construct must exhibit non-allergenicity, extreme antigenicity, and high solubility to elicit a strong immune response. The allergenicity profiling can be determined by AllerTop v2.0, AllergenFP v1.0, and AlgPred v2.0 server (Saha and Raghava 2006; Dimitrov et al. 2014b). The antigenicity of the construct can be assessed with VaxiJen v2.0 and ANTIGENPro server (Doytchinova and Flower 2007; Magnan et al. 2010). The solubility of a vaccine can be analyzed through the SOLpro tool, and a given peptide is expected to be soluble if the calculated score is greater than or equal to 0.5 (Magnan et al. 2009). For a better understanding, another solubility prediction server, namely Protein-Sol, can be utilized, and a protein with a solubility score greater than 0.45 is considered highly soluble (Hebditch et al. 2017). Next, the transmembrane helices and potential signal peptides within the vaccine construct can be determined using TMHMM v2.0 and SignalP 4.1 server (Krogh et al. 2001; Nielsen 2017a; Panda et al. 2022).

BLAST and Human Homology Checking of the Constructed Vaccine

To minimize an autoimmune response, relative homology analysis between the final vaccine candidate and human proteome can be done with the BLASTp module of the PSIBLAST algorithm within the NCBI database (Altschul et al. 1990; Altschul et al. 1997; Narang et al. 2022). In this step, a search must be restricted to *H. sapiens* (taxid:9606), and the query sequence must exhibit less than 40% human homology.

Secondary Structure Analysis of the Vaccine Construct

The secondary structure, as well as the peptide configuration of the final vaccine, can be examined through PSIPRED v4.0 and SOPMA applications (Geourjon and Deléage 1995; Buchan et al. 2013). Considering default parameters, the two servers calculate the percentage of 2D configurations such as alpha helix, random coil, and beta-turn. The PSIPRED v4.0 and SOPMA servers generate the secondary structure of a query protein sequence with a result accuracy of 78.1% and 80%, respectively (Montgomerie et al. 2006).

Development and Analysis of the Tertiary (3D) Structure of the Vaccine Construct

Homology Modeling to Create the 3D Model of the Constructed Vaccine

The RaptorX web server can be employed to build a 3D model of the vaccine candidate. To predict the tertiary structure, this server applies a homology modeling technique, and a 3D model having the lowest p-value is admitted as the finest model (Wang et al. 2016).

3D Model Refinement and Validation

A vaccine model's tertiary (3D) structure can be refined using the GalaxyRefine module on the GalaxyWEB server, which generates five refined models as output. These refined models are ranked according to the score of different parameters, including GDT-HA, RMSD, MolProbity, Clash score, Poor rotamers, and Rama favored (Ko et al. 2012). Afterward, the refined model can be validated with a ProSA-web server that calculates the Z-score of that particular model. This server can be used to analyze the stereochemical quality of a protein model by evaluating the geometry of both individual residues and the overall structure (Wiederstein and Sippl 2007). Then the validated model can be further assessed using Verify3D and ERRAT web servers. Verify3D algorithm assesses a query protein model with its three-dimensional profile obtained from X-ray crystallographic, NMR spectroscopic, and/or computational methods (Eisenberg et al. 1997). In contrast, the ERRAT program assesses a 3D model by identifying imprecise regions within a protein structure based on the errors resulting from the random distribution of atoms (Colovos and Yeates 1993). The PROCHEK application can be used to assess the Ramachandran plot, providing valuable information about the overall quality of the refined vaccine model. Based on dihedral angles [psi (ψ) and phi (ϕ)], the Ramachandran plot visualizes the percentage of amino acid residues within the most favored, generously allowed, additionally allowed, and disallowed regions. A good quality model should have over 90% of amino acid residues in its most favored region (Morris et al. 1992).

Engineering Disulfide Bonds Inside the Constructed Vaccine Candidate

Disulfide bonds within a protein molecule are critical to stabilizing the tertiary/quaternary structure, interactions, and dynamics. Next to the refinement, the vaccine construct can be submitted to Disulfide by Design v2.12 server for disulfide engineering. For disulfide bridging, default values (in°) can be kept for $\chi3$ and Cα-Cβ-Sγ angles. The angle of $\chi3$ ranging between -87 to $+97°$ and the energy score of less than 2.2 kcal/mol suggests an effective disulfide bridging (Craig and Dombkowski 2013).

Scanning for CBL (Conformational B Lymphocyte) Epitopes Within the Newly Formulated Vaccine

The CBL epitopes within the formulated vaccine construct can be predicted with the help of the ElliPro: Antibody Epitope Prediction tool within the IEDB analysis resource. The discontinuous B-cell epitopes can be detected by allowing a minimum protein index (PI) score of 0.5 and a maximum distance between the residue's center of mass (R) 6 Å as the default value. A larger value for R and PI indicates a larger conformational B-cell epitope and greater solvent accessibility, respectively (Ponomarenko and Bourne 2007).

Normal Mode Analysis (NMA) of the Vaccine Construct

NMA is highly required to understand the spontaneous functional motion of a protein complex in its internal (dihedral) coordinates. The iMODS server can be used to analyze the normal mode of the designed vaccine candidate. This quicker and cost-effective MD (Molecular Dynamic) simulation analysis technique facilitates the prediction of the eigenvalues, deformability, B-factors, and covariance (López-Blanco et al. 2014) .

Computational Immune Simulation Analysis of the Constructed Vaccine

A vaccine candidate's immunogenicity and immune response can be understood by exploiting the C-ImmSim web server. This server applies an immune simulation technique, setting the parameters as defaults (Dellagostin et al. 2017).

Molecular Docking Simulation Study Between Vaccine Construct and TLR4 (Toll-Like Receptor) Complexes

Computer-assisted molecular docking assessment can predict the molecular interaction and binding affinity of TLR and vaccines. TLRs are extremely associated with strong immunity (Rafi et al. 2022).

TLR Preparation

For docking analysis, the X-Ray crystallographic structure of the human TLR4 complex with MD-2 and LPS (PDB ID 4G8A) can be downloaded from the RCSB protein data bank bearing a resolution of 2.4 Å. The ligands, along with B, C, and D chains, can be removed by BIOVIA DS (Discovery Studio) v4.5. Later on, the energy of the protein structure can be minimized with Swiss-Pdb Viewer v4.1.0 applying GROMOS 43B1 force field (Guex and Peitsch 1997; Berman et al. 2002; Accelrys Software Inc: San Diego 2012).

Docking Simulation Analysis

Next, the vaccine candidate and the prepared TLR4 can be docked by a protein-protein docking server, i.e., ClusPro v2.0 (Land and Humble 2018). The TLR4-vaccine docked complex with the lowest docking score can be considered to have high-affinity binding, and the molecular interaction can be observed using BIOVIA DS (Discovery Studio) v4.5 (Mahapatra et al. 2022b).

MD (Molecular Dynamics) Simulation Study of the Vaccine Construct and TLR4 Docked Complex

Molecular dynamics simulation allows researchers to examine the potential vaccine's molecular and atomic motions. The molecular dynamics simulation is employed to analyze the association between the receptor proteins (TLRs) and the vaccine candidate (multi-epitope-based subunit vaccine) (Kozakov et al. 2017). The molecular docking technique initially determines the stability between the vaccine-receptor complex, which is further supported and verified by molecular dynamics simulation (Mahapatra et al. 2022b). The process generally suggests whether the developed vaccine would trigger TLR stimulation which could support higher immune reactions inside the human body (Kozakov et al. 2017). The YASARA (Yet Another Scientific Artificial Reality Application) Dynamics (v22.9.24) software package may be adopted to analyze the MD simulation of the vaccine-TLR4 complex. During the simulation, AMBER14 forcefield can be employed (Chatterjee et al. 2018). Before the MD simulation, the complex is cleaned by deleting unknown ligands, water molecules, and metal ions. Similarly, H-bonded networks are optimized to reorder hydrogen bonds and add the missing ones (Pyasi et al. 2021). A simulation cell can solvate the protein complex using the TIP3P solvation model, where the solvent density value may be maintained at 0.997gL-1 (Harrach and Drossel 2014). The AMBER force fields are generally integrated with the most regularly utilized TIP3P solvent model. While the TIP3P framework has no impact on the thermodynamic characteristics of the solutes, it dramatically lowers the distances among these stages, speeding up the dynamics and thereby improving testing in the computations (Krieger et al. 2012). The protonation arrangement of proteins is critical for their structural rigidity. Before initiating a traditional MD simulation, the protonation stages should be established and assigned (Florová et al. 2010). The SCWRL algorithm manages the protonation state of every amino acid within a protein molecule which helps calculate each amino acid's pKa (acid dissociation constant) value. Furthermore, Na^+ and Cl^- can be added to preserve the physiological environment at pH 7.4 and 298 K temperatures (Krieger et al. 2012; Pyasi et al. 2021).The Particle Mesh Ewald (PME) approach can be used to calculate the long-range interactions, short-range Coulomb, and vdW contacts (Varma et al. 2006). When utilizing PME to handle electrostatic interactions, molecular dynamics simulations of protein in specified water are significantly impacted by adding $Cl^-$and Na^+ particles (Alam et al. 2019). When the ionic solution equilibrates, the protein's flexible regions' overall architecture and movements are influenced by the presence of salt ions and charge-stabilizing opposite ions (Alam et al. 2019). The steepest descent is preferable for reducing the high-energy characteristics of the starting configuration (Hsieh et al. 2009). Using the simulated annealing methods, the energy of the TLR4-vaccine docked complex can be minimized with the steepest gradient approaches. For the simulation process, the time step can be set as 2.0 fs, where long-range electrostatic interactions can be calculated with a cut of radius 8 Å (Grote et al. 2005). The simulation may be conducted for 100 ns and the trajectories can be stored following 100 fs intervals. The data within trajectory files can be used to analyze RMSD (Root Mean Square Deviation), RMSF (Root Mean Square Fluctuation), Rg (Radius of Gyration), SASA (Solvent Accessible Surface Area), and H-bonds (Solanki and Tiwari 2018). Despite several successes, MD simulation incorporates challenges like a lack of more refined force fields or superior computational power demanding more than a microsecond simulation time (Durrant and McCammon 2011).

Insilico Codon Optimization and Molecular Cloning of the Constructed Vaccine

Highly efficient cloning and expression properties of a multi-epitope-based peptide vaccine construct are needed to develop an effective vaccine. Therefore, effectual codon adaptation, optimization, and vaccine cloning can be carried out in *E. coli* K12 (Solanki and Tiwari 2018). Since human codons differ from *E. coli*, JCat (JAVA Codon Adaptation Tool), an online application, can be employed to reverse translate and optimize the final vaccine construct. This step increases the expression of the final vaccine construct into the *E. coli* host. JCat output for the adapted and optimized construct exhibits the nucleotide sequence, CAI (Codon Adaptation Index), and % of GC content, which are essential for proper expression in a particular host (Grote et al. 2005). For the effective expression of a vaccine construct, the CAI value must range from 0 to 1, while % of GC content must be within 30–70%. Finally, BglII and ApaI restriction sites can be added at the newly formulated vaccine's N and C terminal end. The freshly prepared vaccine codon sequence can be cloned into the pET-28a (+) vector using SnapGene v6.1 software (Solanki and Tiwari 2018).

Conclusion and Future Scope

Developing a swift and highly effective vaccinology technique is critical for responding to unexpected health catastrophes and lowering infection-related death rates. Vaccination via sparking the immune response offers protection against infectious diseases, reducing morbidity and mortality. Vaccine development must be efficient and prompt to tackle emergent health crises. However, conventional vaccine design and development procedures are time-consuming and expensive. On the contrary, computational vaccinology supported by vaccinomics and immunoinformatics strategies from that perspective has placed the world in an advantageous stage to screen and detect antigens of interest in an economically friendly and time-saving manner and develop vaccine candidates to combat the emergent pathogenic invasion. The wealth of genomics and proteomics data allows informatics to effectively expand its contribution to medical innovation, especially in vaccine science. In the post-genomic age, the construction of multi-epitope-based peptide vaccines has emerged as a unique concept. The availability of the entire microbial genome and proteome sequences and the applicability of bioinformatic tools/techniques for analyzing these sequences can be used to design multi-epitope-based peptide vaccines, which unleash the detection of top immunogenic protein candidates for vaccine development. Thus, designing a multi-epitope-based peptide vaccine offers a promising avenue for efficient and cost-effective therapy and generating a robust immune response against infectious disease.

This review delivers a modest, elementary, and typical procedure/protocol for designing and developing multi-epitope-based peptide vaccines with the aid of different databases, computational tools, and algorithms. Interested researchers/immunologists might utilize the information in this article to guide designing multi-epitope-based peptide vaccine candidates for subsequent pre-clinical and clinical studies. This concise and comprehensive review encompasses a range of essential resources and databases needed to identify the most potent as well as novel antigenic protein sequences (CTL, HTL, and LBL epitopes), assess MHC (both class I and II) binding, create a putative vaccine construct through homology modeling, analyze the interaction between the constructed vaccine and immune receptors (TLR4) using molecular docking and dynamics simulation, compute normal mode and immune simulation analysis of the vaccine candidate and finally molecular cloning of the newly constructed vaccine (Fig. 3). We hope that this summarized review may offer a more effective and accessible vaccinology protocol for future researchers allowing them to design vaccines according to the pathogen of interest computationally.

In the near future, multi-epitope-based peptide vaccine design and development will likely become the fastest-growing field of biological science, particularly in response to combatting infectious diseases. With bioinformatics and computational modeling advancements, researchers can predict epitopes more easily and

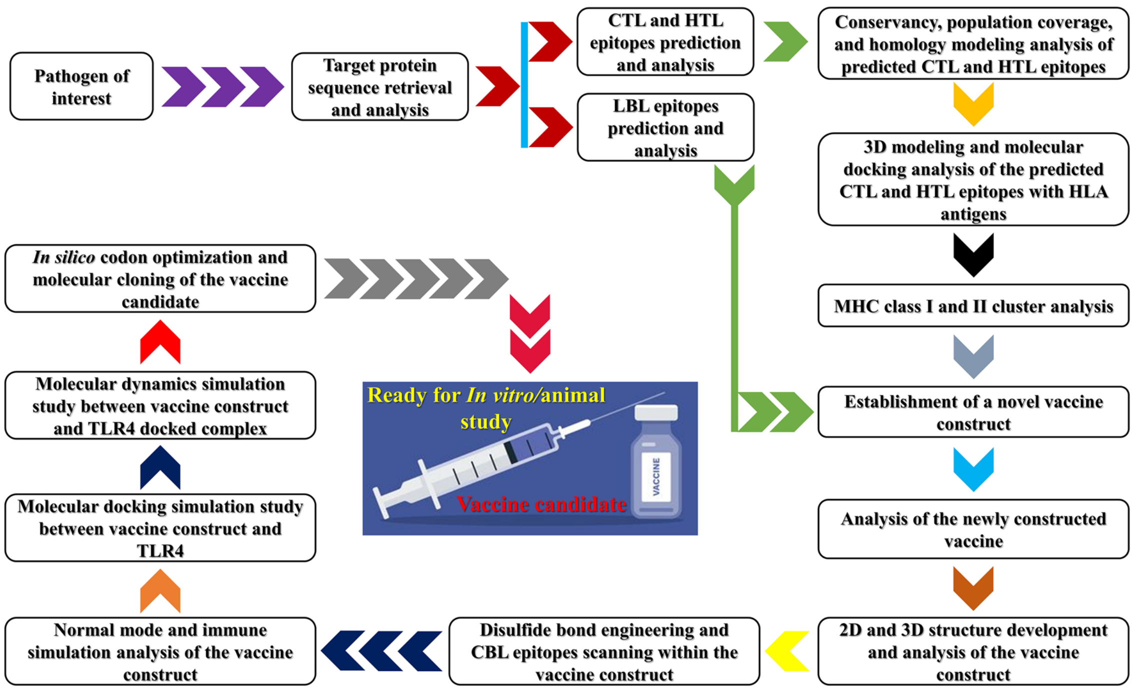


Fig. 3 Concise and comprehensive representation showing different applications employed for the vaccinomics/immunoinformatics governed multi-epitope-based peptide vaccine developmental process. This illustration is generated using Microsoft Office (PowerPoint) 2019

accurately, which are most likely to elicit a potent and effective immune response, making the development of new vaccines much more economically, rapid and efficient. Using multi-epitope-based peptide vaccines may help reduce the global burden of infectious diseases by providing a safe and effective means of preventing and treating those illnesses. Additionally, as our understanding of the immune system and the mechanisms of antigen recognition and presentation continues to grow, new strategies for enhancing the immunogenicity of epitopes and improving the efficacy and durability of epitope-based peptide vaccines are likely to emerge. Overall, multi-epitope-based peptide vaccine designing and development holds great promise for preventing and controlling infectious diseases in the years to come.

Acknowledgements This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1C1C1008694). The authors would like to thank the faculties of the Department of Biochemistry and Molecular Biology, Jahangirnagar University, for their tremendous support.

Author Contributions MMAKS, ARS, and MAH conceptualized and designed the study. MMAKS, MAH, SKH, TAA, and MNS did the investigation. NS, TAA, and MMAKS analyzed the data. MMAKS, MAH, and SKH did the methodology. SKH, MMAKS, ARS, and MAH wrote the original manuscript draft. MMAKS, ARS, SRS, and MAH reviewed and edited the manuscript. MMAKS, ARS, and MAH administered the project.

Declarations

Competing Interests The authors declare no competing interests.

References

Abass OA, Timofeev VI, Sarkar B (2022) Immunoinformatics analysis to design novel epitope based vaccine candidate targeting the glycoprotein and nucleoprotein of Lassa mammarenavirus (LASMV) using strains from Nigeria. J Biomol Struct Dynamics 40:7283–7302. https://doi.org/10.1080/07391102.2021.1896387

Accelrys Software Inc (2012) San Diego 2012. Accelrys Discovery Studio. Discovery Studio modeling Environment, Release 3.5. Accelrys Discovery Studio:Accelrys Software Inc, San Diego

Alam S, Hasan MDK, Manjur OHB (2019) Predicting and designing epitope ensemble vaccines against HTLV-1. J Integr Bioinform 16:20180051. https://doi.org/10.1515/jib-2018-0051

Altschul SF (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389

Altschul SF, Gish W, Miller W (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ashrafi H, Siraji MI, Showva NN, Hossain MM, Hossan T, Hasan MA, Shohael AM, Shawan MMAK (2019) Structure to function analysis with antigenic characterization of a hypothetical protein, HPAG1_0576 from *Helicobacter pylori* HPAG1. Bioinformation 15:456–466. https://doi.org/10.6026/97320630015456

Berman HM, Battistuz T, Bhat TN (2002) The protein data bank. Acta Crystallogr D Biol Crystallogr 58:899–907. https://doi.org/10.1107/S0907444902003451

Buchan DWA, Minneci F, Nugent TCO (2013) Scalable web services for the PSIPRED protein analysis Workbench. Nucleic Acids Res 41:W349–W357. https://doi.org/10.1093/nar/gkt381

Bui H-H, Sidney J, Dinh K (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics 7:153. https://doi.org/10.1186/1471-2105-7-153

Bui H-H, Sidney J, Li W (2007) Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC Bioinformatics 8:361. https://doi.org/10.1186/1471-2105-8-361

Calis JJA, Maybeno M, Greenbaum JA (2013) Properties of MHC class I presented peptides that enhance immunogenicity. PLoS Comput Biol 9:e1003266. https://doi.org/10.1371/journal.pcbi.1003266

Chatterjee N, Ojha R, Khatoon N, Prajapati VK (2018) Scrutinizing *Mycobacterium tuberculosis* membrane and secretory proteins to formulate multiepitope subunit vaccine against pulmonary tuberculosis by utilizing immunoinformatic approaches. Int J Biol Macromol 118:180–188. https://doi.org/10.1016/j.ijbiomac.2018.06.080

Colovos C, Yeates TO (1993) Verification of protein structures: patterns of non-bonded atomic interactions. Protein Sci 2:1511–1519. https://doi.org/10.1002/pro.5560020916

Craig DB, Dombkowski AA (2013) Disulfide by design 2.0: a web-based tool for disulfide engineering in proteins. BMC Bioinform 14:346. https://doi.org/10.1186/1471-2105-14-346

Database resources of the National Center for Biotechnology Information (2016) Nucleic Acids Res 44:D7–D19. https://doi.org/10.1093/nar/gkv1290

Dellagostin O, Grassmann A, Rizzi C (2017) Reverse vaccinology: an approach for identifying Leptospiral vaccine candidates. Int J Mol Sci 18:158. https://doi.org/10.3390/ijms18010158

Dey J, Mahapatra SR, Raj TK, Kaur T, Jain P, Tiwari A, Patro S, Misra N, Suar M (2022a) Designing a novel multi-epitope vaccine to evoke a robust immune response against pathogenic multidrug-resistant *Enterococcus faecium* bacterium. Gut Pathog 14:21. https://doi.org/10.1186/s13099-022-00495-z

Dey J, Mahapatra SR, Lata S, Patro S, Misra N, Suar M (2022b) Exploring *Klebsiella pneumoniae* capsule polysaccharide proteins to design multiepitope subunit vaccine to fight against pneumonia. Expert Rev Vaccines 21:569–587. https://doi.org/10.1080/14760584.2022.2021882

Dhanda SK, Gupta S, Vir P, Raghava GP (2013) Prediction of IL4 inducing peptides. Clin Dev Immunol 2013:263952. https://doi.org/10.1155/2013/263952

Dimitrov I, Bangov I, Flower DR, Doytchinova I (2014) AllerTOP v.2—A server for in silico prediction of allergens. J Mol Model 20:2278. https://doi.org/10.1007/s00894-014-2278-5

Dimitrov I, Naneva L, Doytchinova I, Bangov I (2014b) AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics 30:846–851. https://doi.org/10.1093/bioinformatics/btt619

Dorosti H, Eslami M, Negahdaripour M (2019) Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. J Biomol Struct Dyn 37:3524–3535. https://doi.org/10.1080/07391102.2018.1519460

Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform 8:4. https://doi.org/10.1186/1471-2105-8-4

Durrant JD, McCammon JA (2011) Molecular dynamics simulations and drug discovery. BMC Biol 9:71. https://doi.org/10.1186/1741-7007-9-71

Eisenberg D, Luthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404. https://doi.org/10.1016/S0076-6879(97)77022-8

Florová P, Sklenovský P, Banáš P, Otyepka M (2010) Explicit Water Models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact. J Chem Theory Comput 6:3569–3579. https://doi.org/10.1021/ct1003687

Gasteiger E (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788. https://doi.org/10.1093/nar/gkg563

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, Totowa, pp 571–607

Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics 11:681–684. https://doi.org/10.1093/bioinformatics/11.6.681

Grote A, Hiller K, Scheer M (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 33:W526–W531. https://doi.org/10.1093/nar/gki376

Guex N, Peitsch MC (1997) SWISS-MODEL and the swiss-pdb viewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723. https://doi.org/10.1002/elps.1150181505

Gupta S, Kapoor P, Chaudhary K (2013) In Silico approach for predicting toxicity of peptides and proteins. PLoS ONE 8:e73957. https://doi.org/10.1371/journal.pone.0073957

Harrach MF, Drossel B (2014) Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. J Chem Phys 140:174501. https://doi.org/10.1063/1.4872239

Hebditch M, Carballo-Amador MA, Charonis S (2017) Protein–Sol: a web tool for predicting protein solubility from sequence. Bioinformatics 33:3098–3100. https://doi.org/10.1093/bioinformatics/btx345

Hsieh W-T, Kuo M-K, Yau HF, Chang C-C (2009) A simple arbitrary phase-step digital holographic reconstruction approach without blurring using two holograms. OPT REV 16:466–471. https://doi.org/10.1007/s10043-009-0090-8

Kar T, Narsaria U, Basak S (2020) A candidate multi-epitope vaccine against SARS-CoV-2. Sci Rep 10:10895. https://doi.org/10.1038/s41598-020-67749-1

Khan T, Khan A, Nasir SN, Ahmad S, Ali SS, Wei D-Q (2021a) CytomegaloVirusDb: multi-omics knowledge database for cytomegaloviruses. Comput Biol Med 135:104563. https://doi.org/10.1016/j.compbiomed.2021.104563

Khan T, Khan A, Wei D-Q (2021b) MMV-db: vaccinomics and RNA-based therapeutics database for infectious hemorrhagic fever-causing mammarenaviruses. Database 2021:baab063. https://doi.org/10.1093/database/baab063

Khan A, Khan S, Ahmad S, Anwar Z, Hussain Z, Safdar M, Rizwan M, Waseem M, Hussain A, Akhlaq M, Khan T, Ali SS, Wei D-Q (2021c) HantavirusesDB: Vaccinomics and RNA-based therapeutics database for the potentially emerging human respiratory pandemic agents. Microb Pathog 160:105161. https://doi.org/10.1016/j.micpath.2021c.105161

Khan T, Abdullah M, Toor TF, Almajhdi FN, Suleman M, Iqbal A, Ali L, Khan A, Waheed Y, Wei D-Q (2022) Evaluation of the whole proteome of *Achromobacter xylosoxidans* to identify vaccine targets for mRNA and peptides-based vaccine designing against the emerging respiratory and lung cancer-causing bacteria. Front Med 8:825876. https://doi.org/10.3389/fmed.2021.825876

Khan T, Muzaffar A, Shoaib RM, Khan A, Waheed Y, Wei D-Q (2022b) Towards specie-specific ensemble vaccine candidates against mammarenaviruses using optimized structural vaccinology pipeline and molecular modelling approaches. Microb Pathog 172:105793. https://doi.org/10.1016/j.micpath.2022.105793

Khan T, Suleman M, Ali SS, Sarwar MF, Ali I, Ali L, Khan A, Rokhan B, Wang Y, Zhao R, Wei D-Q (2022c) Subtractive proteomics assisted therapeutic targets mining and designing ensemble

vaccine against *Candida auris* for immune response induction. Comput Biol Med 145:105462. https://doi.org/10.1016/j.compbiomed.2022.105462
Ko J, Park H, Heo L, Seok C (2012) GalaxyWEB server for protein structure prediction and refinement. Nucleic Acids Res 40:W294–W297. https://doi.org/10.1093/nar/gks493
Kozakov D, Hall DR, Xia B (2017) The ClusPro web server for protein–protein docking. Nat Protoc 12:255–278. https://doi.org/10.1038/nprot.2016.169
Krieger E, Dunbrack RL, Hooft RWW, Krieger B (2012) Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. In: Baron R (ed) Computational drug Discovery and Design. Springer New York, New York, NY, pp 405–421
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 305:567–580. https://doi.org/10.1006/jmbi.2000.4315
Lamiable A, Thévenet P, Rey J (2016) PEP-FOLD3: faster *de novo* structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 44:W449–W454. https://doi.org/10.1093/nar/gkw329
Land H, Humble MS (2018) YASARA: A Tool to Obtain Structural Guidance in Biocatalytic Investigations. In: Bornscheuer UT, Höhne M (eds) Protein Engineering. Springer New York, New York, NY, pp 43–67
Larsen MV, Lundegaard C, Lamberth K (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinform 8:424. https://doi.org/10.1186/1471-2105-8-424
López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P (2014) iMODS: internal coordinates normal mode analysis server. Nucleic Acids Res 42:W271–W276. https://doi.org/10.1093/nar/gku339
Magnan CN, Randall A, Baldi P (2009) SOLpro: accurate sequence-based prediction of protein solubility. Bioinformatics 25:2200–2207. https://doi.org/10.1093/bioinformatics/btp386
Magnan CN, Zeller M, Kayala MA (2010) High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics 26:2936–2943. https://doi.org/10.1093/bioinformatics/btq551
Mahapatra SR, Dey J, Jaiswal A, Roy R, Misra N, Suar M (2022a) Immunoinformatics-guided designing of epitope-based subunit vaccine from Pilus assembly protein of *Acinetobacter baumannii* bacteria. J Immunol Methods 508:113325. https://doi.org/10.1016/j.jim.2022.113325
Mahapatra SR, Dey J, Raj TK, Kumar V, Ghosh M, Verma KK, Kaur T, Kesawat MS, Misra N, Suar M (2022b) The potential of plant-derived secondary metabolites as novel drug candidates against *Klebsiella pneumoniae*: molecular docking and simulation investigation. South Afr J Bot 149:789–797. https://doi.org/10.1016/j.sajb.2022.04.043
Manavalan B, Govindaraj RG, Shin TH (2018) iBCE-EL: a New Ensemble Learning Framework for Improved Linear B-cell epitope prediction. Front Immunol 9:1695. https://doi.org/10.3389/fimmu.2018.01695
Mehla K, Ramana J (2016) Identification of epitope-based peptide vaccine candidates against enterotoxigenic *Escherichia coli*: a comparative genomics and immunoinformatics approach. Mol BioSyst 12:890–901. https://doi.org/10.1039/C5MB00745C
Montgomerie S, Sundararaj S, Gallin WJ, Wishart DS (2006) Improving the accuracy of protein secondary structure prediction using structural alignment. BMC Bioinformatics 7:301. https://doi.org/10.1186/1471-2105-7-301
Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12:345–364. https://doi.org/10.1002/prot.340120407
Moutaftsi M, Peters B, Pasquetto V (2006) A consensus epitope prediction approach identifies the breadth of murine TCD8+-cell responses to vaccinia virus. Nat Biotechnol 24:817–819. https://doi.org/10.1038/nbt1215
Nagpal G, Usmani SS, Dhanda SK (2017) Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. Sci Rep 7:42851. https://doi.org/10.1038/srep42851
Nain Z, Abdulla F, Rahman MM (2020) Proteome-wide screening for designing a multi-epitope vaccine against emerging pathogen *Elizabethkingia anophelis* using immunoinformatic approaches. J Biomol Struct Dyn 38:4850–4867. https://doi.org/10.1080/07391102.2019.1692072
Narang PK, Dey J, Mahapatra SR, Ghosh M, Misra N, Suar M, Kumar V, Raina V (2021) Functional annotation and sequence-structure characterization of a hypothetical protein putatively involved in carotenoid biosynthesis in microalgae. South Afr J Bot 141:219–226. https://doi.org/10.1016/j.sajb.2021.04.014
Narang PK, Dey J, Mahapatra SR, Roy R, Kushwaha GS, Misra N, Suar M, Raina V (2022) Genome-based identification and comparative analysis of enzymes for carotenoid biosynthesis in microalgae. World J Microbiol Biotechnol 38:8. https://doi.org/10.1007/s11274-021-03188-y
Nicholson LB (2016) The immune system. Essays Biochem 60:275–301. https://doi.org/10.1042/EBC20160017
Nielsen H (2017) Predicting Secretory Proteins with SignalP. In: Kihara D (ed) Protein function prediction. Springer, New York, NY, pp 59–73
Nielsen H (2017) Predicting Secretory proteins with SignalP. Methods Mol Biol 1611:59–73
O'Boyle NM, Banck M, James CA (2011) Open Babel: an open chemical toolbox. J Cheminform 3:33. https://doi.org/10.1186/1758-2946-3-33
Obaidullah AJ, Alanazi MM, Alsaif NA (2021) Immunoinformatics-guided design of a multi-epitope vaccine based on the structural proteins of severe acute respiratory syndrome coronavirus 2. RSC Adv 11:18103–18121. https://doi.org/10.1039/D1RA02885E
Oh J, Byrd AL, Park M (2016) Temporal stability of the human skin microbiome. Cell 165:854–866. https://doi.org/10.1016/j.cell.2016.04.008
Olejnik J, Hume AJ, Mühlberger E (2018) Toll-like receptor 4 in acute viral infection: too much of a good thing. PLoS Pathog 14:e1007390. https://doi.org/10.1371/journal.ppat.1007390
Panda SS, Dey J, Mahapatra SR, Kushwaha GS, Misra N, Suar N, Ghosh M (2022) Investigation on structural prediction of pectate lyase enzymes from different microbes and comparative docking studies with pectin: the economical waste from food industry. Geomicrobiol J 39:294–305. https://doi.org/10.1080/01490451.2021.1992042
Pettersen EF, Goddard TD, Huang CC (2004) UCSF Chimera–A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084
Ponomarenko JV, Bourne PE (2007) Antibody-protein interactions: benchmark datasets and prediction tools evaluation. BMC Struct Biol 7:64. https://doi.org/10.1186/1472-6807-7-64
Purcell AW, McCluskey J, Rossjohn J (2007) More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov 6:404–414. https://doi.org/10.1038/nrd2224
Pyasi S, Sharma V, Dipti K (2021) Immunoinformatics approach to design multi-epitope- subunit vaccine against bovine ephemeral fever disease. Vaccines 9:925. https://doi.org/10.3390/vaccines9080925
Rafi MdO, Al-Khafaji K, Sarker MdT (2022) Design of a multi-epitope vaccine against SARS-CoV-2: immunoinformatic and computational methods. RSC Adv 12:4288–4310. https://doi.org/10.1039/D1RA06532G

Rahman A, Ali MT, Shawan MMAK, Sarwar MG, Khan MA, Halim MA (2016) Halogen-directed drug design for Alzheimer's disease: a combined density functional and molecular docking study. SpringerPlus 5:1346. https://doi.org/10.1186/s40064-016-2996-5

Rakib A, Sami SA, Mimi NJ (2020) Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. Comput Biol Med 124:103967. https://doi.org/10.1016/j.compbiomed.2020.103967

Rappuoli R, Black S, Bloom DE (2019) Vaccines and global health: in search of a sustainable model for vaccine development and delivery. Sci Transl Med 11:eaaw2888. https://doi.org/10.1126/scitranslmed.aaw2888

Saha S, Raghava GPS (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res 34:W202–W209. https://doi.org/10.1093/nar/gkl343

Saha S, Raghava GPS (2007) Prediction methods for B-cell epitopes. Methods Mol Biol 409:387–394. https://doi.org/10.1007/978-1-60327-118-9_29

Sahoo P, Dey J, Mahapatra SR, Ghosh A, Jaiswal A, Padhi S, Prabhuswamimath SC, Misra N, Suar M (2022) Nanotechnology and COVID-19 convergence: toward new planetary health interventions against the pandemic. OMICS 26:473–488. https://doi.org/10.1089/omi.2022.0072

Shawan MMAK, Mahmud HA, Hasan MM, Parvin A, Rahman MN, Rahman SMB (2014) In silico modeling and immunoinformatics probing disclose the epitope based peptide vaccine against Zika virus envelope glycoprotein. Indian J Pharm Biol Res 2:44–57. https://doi.org/10.30750/ijpbr.2.4.10

Shawan MMAK, Hasan MA, Yesmin R, Hossan T, Hossain MM, Hasan MM, Parvin A, Morshed M, Salauddin NM, Sarker SR, Rahman MN, Rahman SMB (2018) tRNA diversification among uncultured archeon clones. Bioinformation 14:357–360. https://doi.org/10.6026/97320630014357

Shawan MMAK, Halder SK, Hasan MA (2021a) Luteolin and abyssinone II as potential inhibitors of SARS-CoV-2: an in silico molecular modeling approach in battling the COVID-19 outbreak. Bull Natl Res Cent 45:27. https://doi.org/10.1186/s42269-020-00479-6

Shawan MMAK, Sharma AR, Bhattacharya M, Mallik B, Akhter F, Shakil MS, Hossain MM, Banik S, Lee SS, Hasan MA, Chakraborty C (2021b) Designing an effective therapeutic siRNA to silence RdRp gene of SARS-CoV-2. Infect Genet evolution: J Mol Epidemiol evolutionary Genet Infect Dis 93:104951. https://doi.org/10.1016/j.meegid.2021.104951

Solanki V, Tiwari V (2018) Subtractive proteomics to identify novel drug targets and reverse vaccinology for the development of chimeric vaccine against *Acinetobacter baumannii*. Sci Rep 8:9044. https://doi.org/10.1038/s41598-018-26689-7

Srivastava S, Verma S, Kamthania M (2022) Computationally validated SARS-CoV-2 CTL and HTL multi-patch vaccines, designed by reverse epitomics approach, show potential to cover large ethnically distributed human population worldwide. J Biomol Struct Dyn 40:2369–2388. https://doi.org/10.1080/07391102.2020.1838329

The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489. https://doi.org/10.1093/nar/gkaa1100

Thomsen M, Lundegaard C, Buus S (2013) MHCcluster, a method for functional clustering of MHC molecules. Immunogenetics 65:655–665. https://doi.org/10.1007/s00251-013-0714-9

Trott O, Olson AJ (2009) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. https://doi.org/10.1002/jcc.21334

Varma S, Chiu S-W, Jakobsson E (2006) The influence of amino acid Protonation States on Molecular Dynamics Simulations of the bacterial porin OmpF. Biophys J 90:112–123. https://doi.org/10.1529/biophysj.105.059329

Wang P, Sidney J, Dow C (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4:e1000048. https://doi.org/10.1371/journal.pcbi.1000048

Wang P, Sidney J, Kim Y (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinform 11:568. https://doi.org/10.1186/1471-2105-11-568

Wang S, Li W, Liu S, Xu J (2016) RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Res 44:W430–W435. https://doi.org/10.1093/nar/gkw306

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410. https://doi.org/10.1093/nar/gkm290

Xin H, Wu P, Wong JY (2023) Hospitalizations and mortality during the first year of the COVID-19 pandemic in Hong Kong, China: an observational study. Lancet Reg Health - Western Pac 30:100645. https://doi.org/10.1016/j.lanwpc.2022.100645