ISSN: 3033-4896

Research Article

Med Biomedical Journal

In-Silico Design of Novel Multi-epitope Peptide Vaccine against NSCLC by Targeting GRB7/HER2/STAT3

Ali Gholamian Moghaddam[1], Debajyoti Majumdar[2], Parisa Yektay Sanati[3], Malihe Sarmadi[4], Prince AnimAddo[1], Ali Ebrahimi[5], Ali Sadeghi[6], Neda Asghari Kollahi[7], Sabina Dahal[1], Siva Murru[1]

1. School of Sciences, College of Arts, Education and Sciences, University of Louisiana at Monroe, LA 71209.
2. Department of Pathobiological Sciences, LSU School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA 71209.
3. Department of Biotechnology, Faculty of New Science and Technology, Semnan University, Semnan, Iran.
4. Clinical Research Development Unit, Imam Reza Hospital, Birjand University of Medical Sciences, Birjand, Iran.
5. Faculty of Nursing and Midwifery, Zabol University of Medical Sciences, Sistan and Baluchestan, Iran.
6. Birjand University of Medical Sciences, Birjand, Iran.
7. Shahid Akbarabadi Clinical Research Development Unit (ShACRDU), University of Medical Sciences, Tehran, Iran.

***CorrespondingAuthor**
Siva Murru (murru@ulm.edu)

Submitted: 24/06/2024 Accepted:06/07/2024 published:10/07/2024

Citation: Gholamian Moghaddam A., Majumdar D., Yektay Sanati P., Sarmadi M., AnimAddo P., Ebrahimi A., Sadeghi A., Asghari Kollahi N., Dahal S., Murru S., (2024) In-Silico Design of Novel Multi-epitope Peptide Vaccine against NSCLC by Targeting GRB7/HER2/STAT3. Med Biomedial J., 1(4), 11-27

Abstract

Lung cancer is the most common leading cause of cancer-related death in the world. The non-small cell lung cancer (NSCLC) involves approximately 80–85% of all lung cancer cases. At present, a novel approach in cancer treatment for NSCLC is vaccine therapy that seeks to induce adaptive anti-tumor immune responses by introducing tumor antigens into the body to prime the immune system. This stimulation aims to activate the host immune system, prompting the generation of tumor antigen-specific effector and memory T-cell responses while avoiding non-cancerous cells. The present study aimed to develop a multi-epitope peptide vaccine targeting key cancer antigens, including GRB7, HER2, and STAT3, known for their significant roles in tumor cell proliferation and invasion. By using the immunoinformatic analysis tools, the protein vaccine was designed, based on B cell epitopes and CD4 and CD8 induced epitopes, which are essential for enhancing immune responses against cancer cells. Physicochemical, antigenic, and allergenic properties were performed to ensure the safety and immunogenicity of the vaccine. Predictions of secondary and tertiary structures were conducted, along with refinement and validation processes to control any structural concerns. Finally, molecular docking was performed for the constructed vaccine. Further investigations are necessary to validate the safety and immunogenic potential of this novel vaccine for NSCLC before its clinical application in patients.

Keywords: NSCLC- Multiepitope Vaccine- Immunoinformatic- HER2- STAT3- GRB7.

Introduction

On a global scale, both cancer incidence and mortality rates are increasing. In particular, lung cancer makes up a significant portion (11%) of newly diagnosed cancers, but it represents a much smaller number (3%) of all people living with cancer. This is because lung cancer has a lower survival rate compared to other major cancers.[1] In the United States alone, it is estimated that there will be approximately 238,340 new cases of lung cancer in 2023, resulting in about 127,070 deaths[2]. It continues to maintain its position as the primary cause of cancer-related deaths worldwide, contributing to 18.4% of total fatalities and imposing a considerable societal and economic burden .[3,4] Surgical resection stands as the sole truly curative therapy; however, with relapse rates surpassing 40% following definitive resection, new-age adjuvant-based therapeutics represent a promising approach. [4,5] In advanced stages of the disease, systemic chemotherapy and/or localized irradiation can generate objective responses and alleviate symptoms. Nonetheless, these therapies provide only marginal enhancements in survival. The two-year survival rates for stages IIIB and IV non-small cell lung cancer (NSCLC) stand at 10.8% and 5.4% respectively, with corresponding five-year survival rates of 3.9% and 1.3% respectively.[6,7]

Throughout time, surgery, radiotherapy, and chemotherapy have stood as the primary triad in cancer treatment. But in contrast to conventional cancer treatments, immunotherapy harnesses the body's immune system to identify and combat cancer cells, providing a natural means to manage the disease's advancement. Cancer immunotherapy has come a long way from the early days of "Coley's toxins" to the identification of neoantigen targets for patient-tailored treatment. [8] It has rightfully earned its place as a fourth mode of treatment, next to the stalwarts like surgery, radiotherapy, and chemotherapy. The immune system has been recognized to play a dual role in a way that informs the phenotypic features of the cancer in the body. It has been said about cancers that they are 'wounds that do not heal',[9] implying the role of the immune system-mediated tissue repair machinery in the genesis of cancer in the body. Described in the literature was the ability of interferon-γ (IFN-γ) stimulation to induce rejection of tumor transplants [10] and likewise mice lacking IFN-γ responsiveness (IFN- γ receptor knockout etc.) or with an impaired adaptive immune system ($RAG2^{-/-}$) showed increased susceptibility to induced cancers.[11] On the other hand, it has also been observed that tumors derived in mice that lacked an intact immune system, were in general more immunogenic when transplanted in immunocompetent mice, whereas immunocompetent mice themselves select for a less immunogenic tumor phenotype growing in them.[12] Overall, immunotherapy functions by modulating the immune response, either by activating the body's immune defense system (e.g. CAR T-cells, therapeutic cancer vaccines etc.) or by employing biologically active compounds (e.g. checkpoint inhibitors etc.) to inhibit and deter tumor growth .[13]

Among immunotherapeutic agents, cancer vaccines, offer the unique advantage of generating a specific adaptive immune response to recognize and target the cancer cells. Target antigen selection is of paramount importance in the design of these vaccines. Targets are classified as either tumor-associated antigens (TAA) or tumor-specific antigens (TSA). TAAs are often self-antigens overexpressed or aberrantly expressed in the context of the cancer (e.g. cancer-testis antigens).[14] Since these are shared antigens, vaccine agents will need to contend with immune tolerance mechanisms in the body to be able to target the cancer effectively while minimizing off-target effects. On the other hand, TSAs are often neoantigens arising out of patient-specific mutations (single nucleotide variations, small insertions and deletions, etc.) or virally-transmitted mutations.[15] Being novel antigens, they are not protected by central and peripheral tolerance, and yet predicting the most strongly immunogenic epitopes out of these proteins is a challenge in itself. Whole exome sequencing combined with mass spectrometry was successfully used to identify tumor-specific mutations coupled with assessing the major histocompatibility complex (MHC)-binding ability of peptide fragments generated from these.[16] Summing up, a critically important advantage of vaccines against TAAs is the promise of an off-the-shelf solution which makes it both time and cost-effective.

The HER2 protein is part of the HER/ErbB family of receptor tyrosine kinases, comprising an extracellular region, a transmembrane domain, and a tyrosine kinase domain with a C-terminal regulatory region. HER2 lacks a known soluble ligand; instead, downstream signaling is initiated by forming dimers with other ligand bound HER family members. Moreover, HER2 exhibits reduced internalization and degradation, allowing it to remain activated for longer periods on the cell membrane.[17,18] The typical outcome of alterations in the HER2 gene/protein is the heightened activation of the receptor due to increased homo- or heterodimerization and autophosphorylation in the cytoplasmic domain. This activation can start many signaling pathways such as mitogen-activated protein kinase (MAPK), protein kinase C (PKC), and signal transducers and activators of transcription (STAT), which can induce rapid and abnormal cell proliferation.[17-19]

Growth factor receptor-bound protein-7 (Grb7) is a multitask adaptor protein that has an essential role in the activation of epidermal growth factor receptor (EGFR) which has been recognized for its role in various physiological and pathological pathways. The overexpression of Grb7 has significant cooperation with erythroblastic leukemia viral oncogene homolog (ERBB) in the progress of advanced human cancers. Grb7 has been involved in EGF-regulated STAT3 phosphorylation by making a complex with STAT3 which can increase the aggregation of STAT3 in the nucleus. Interestingly, when nuclear accumulation of STAT3 has been increased, it can lead to the transcription of genes that are involved in lung cancer expansion and migration.[20] Therefore, Grb7 has been considered as a potential therapeutic target in the context of NSCLC treatment.[20]

The Signal Transducer and Activator of Transcription 3 (STAT3) is a member of the STAT family of proteins, which function as both signal transducers and transcription factors. This family comprises seven identified members: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6.[21,22] A significant number of growth factors, cytokines, and hormones are involved in the activation of STAT3.[22,23] Receptor tyrosine kinases like EGFR and vascular endothelial growth factor receptor (VEGFR) can phosphorylate STAT3 upon ligand binding.[22,24] Above 50 % of NSCLC cases, STAT3 can remain constantly activated.[22,25] Overexpression of STAT3 is related to advanced clinical stages of cancer and metastasis of lymph nodes and drug resistance.[22,26] Moreover, mutations in receptor tyrosine kinases like EGFR are linked to over-activation of STAT3 signaling in NSCLC.[27] Furthermore, disturbances in STAT3 regulators such as PTP, PIAS, or SOCS proteins are also evident in NSCLC tumors, leading to heightened levels of phosphorylated STAT3.[28,29] Crucially, STAT3 is not essential for the growth and survival of normal cells, making it a valuable target specific to cancer. [30] So far, well-characterized tumor-promoting functions of STAT3 signaling in NSCLC include the promotion of angiogenesis, cell survival, cancer cell stemness, drug resistance, and evasion of anti-tumor immunity.

Cancer vaccines are beginning to emerge as a new avenue in adjuvant immunotherapy. So far, a total of 8 cancer vaccines have seen usage in the clinical setting worldwide.[31] Peptide vaccines in trials have demonstrated mixed results.[32-35] However, it is still among the most actively studied categories of cancer vaccines in the advanced stages of clinical trials. This work focuses on designing a multiepitope immunotherapeutic peptide vaccine

targeted against HER 2, STAT3, and Grb7. All three are TAAs with relevance in the context of lung cancer.

Materials and Methods:

Protein sequence retrieval

The amino acid sequence of GRB7 (Accession no. AAH06535.1), STAT3 (Accession no. KAI2583079.1), and ERBB2 (Accession no. NP_001369711.1) were Collected and saved in FASTA format from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih/Blast).

Identification and selection of T-cell and B-cell epitopes

T cell epitopes that are identified by T cell receptors can bind to MHC molecules and it can induce immune response. MHC-I molecules can introduce peptides to CD8+ T cells but on the other hand, the MHC-II molecules can present peptides to CD4+ T cells.[36] For this purpose, The IEDB database (http://tools.immuneepitope.org/MHC-I/), Net-MHC 4.0 (http://imed.med.ucm.es/Tools/rankpep.html) server, and RANKPEP server (http://www.cbs.dtu.dk/services/NetMHC/) were utilized to predict all potential MHC-I binding epitopes for peptide antigens derived from GRB7, STAT3, and HER2 proteins.

the Immune Epitope Database and Analysis Resource (IEDB) is a free database that has a variety of applications in developing new vaccines, diagnostics, and therapeutics. This tool can evaluate different epitopes, and it is a benchmarked method to predict MHC class I and class II binding. [37] The NetMHC 4.0 server predicts interactions between MHC-I molecules and peptides by analyzing ligand data and assessing the binding affinity of the MHC-I molecules[38]. Moreover, the RANKPEP that used for the prediction of peptide-MHC class I (MHC-I) binding[39].

Utilizing the Immune Epitope Database (http://tools.iedb.org/MHC-Ii/) and NetMHC-IIpan 4.0 server (http://www.cbs.dtu.dk/services/NetMHC-IIpan/), MHC-II binding epitopes for GRB7, STAT3, and HER2 proteins were forecasted. Finally, The NetCTL 1.2 server was employed to compute CTL epitopes for the specified proteins. This web server Combining the MHC class I affinity, TAP transport efficiency, and proteasomal cleavage for prediction of cytotoxic T lymphocytes (CTL) epitopes.[40]

B cell epitopes are the specific parts of an antigen that are identified by the B cell receptor.[41] To predict linear B-cell epitopes of the targeted proteins the Immune Epitope Database (http://tools.iedb.org/bcell/) was utilized. Subsequently, the selected peptide sequences underwent assessment for allergenicity and antigenicity using the AllergenFP v.1.0 server (http://ddg-pharmfac.net/AllergenFP/) and VaxiJen 2.0 server (http://www.ddg-pharmfac.net/vaxijen/), respectively.

Construction of the multi-epitope vaccine construct

In the formulation of the final multi-epitope vaccine, the most promising epitopes were chosen from GRB7, STAT3, and HER2 proteins, along with three adjuvants: Heparin-binding hemagglutinin: (HBHA), Mannose-binding protein (MBP), and Large ribosomal subunit protein bL12 (P9WHE3). Various linkers, including EAAAK, AYY, AK, KFER, and GPGPG, were utilized to bridge the selected epitopes and adjuvants in the vaccine construct.

Assessment of Physicochemical Properties, Allergenicity, solubility and Antigenicity

Following the development of the ultimate multiepitope vaccine, Expasy's ProtParam online server (http://web.expasy.org/protparam/) was employed to predict various physicochemical properties of the vaccine. These properties encompassed the total count of amino acids, atoms, and their composition, as well as the weight and formula of the molecule. Additionally, predictions were made for the total number of positive and negative residues, instability, aliphatic index, isoelectric point (pI), and extinction coefficients, along with the grand average of hydropathicity (GRAVY).

Antigenicity is the pattern of antibody responses that the host can show against a particular pathogens.[42] The allergenicity and antigenicity of the vaccine were evaluated by AllergenFP v.1.0 server (http://ddg-pharmfac.net/AllergenFP/). Moreover, The vaccine's antigenicity was also validated by ANTIGENpro (http://scratch.proteomics.ics.uci.edu) and VaxiJen 2.0 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) server. Finally, the solubility of the vaccine was also evaluated by the SOLpro (https://scratch.proteomics.ics.uci.edu/).

Prediction of the secondary structure

The secondary structure of the multiepitope vaccine was anticipated by using three servers: The GOR4 (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) and PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/). The GOR method is one of the most common for secondary structure prediction. The GOR method has benefits compared to neural network-based methods since it clearly specifies the factors considered for the prediction and the ones that are disregarded.[43] Self-optimized prediction method (SOPM) can predict 69.5% of amino acids in a comprehensive database of 126 non-homologous protein chains described by a three-state model of secondary structure, consisting of α-helix, β-sheet, and coil.[44] The PSIPRED server utilizes two feed-forward neural networks to analyze the output Protein Specific Iterated BLAST.

Tertiary structure modeling, refinement, and validation of the multi-epitope vaccine

Three-dimensional (3D) protein structures offer crucial insights into the molecular mechanisms of protein function, facilitating the design of experiments like site-directed mutagenesis, investigations into disease-related mutations, and the development of specific inhibitors based on structure. Although significant advancements in experimental structure determination techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), these processes remain time-

consuming and may not always yield success. SWISS-MODEL (http://swissmodel.expasy.org) is an automated server designed for the comparative modeling of three-dimensional (3D) protein structures.[45] The refinement of the 3D model of the vaccine was utilized by the Swiss-model server (https://swissmodel.expasy.org/). To validate the model, the Ramachandran plot was acquired from the SAVES v6.0 server (https://saves.mbi.ucla.edu/) to assess the stereochemical quality of the predicted vaccine model.

Identification of discontinuous B-cell epitopes

By using the 3D structure of multiepitope vaccine, all feasible discontinuous B-cell epitopes are predicted by Ellipro of the IEDB database (http://tools.iedb.org/ellipro/) This database facilitates in predicting discontinuous (conformational) B cell epitopes within the validated 3D model of the vaccine, employing three algorithms that leverage protrusion index (PI) values to characterize the ellipsoidal shape of the protein, measure the PI of residues, and analyze clusters of neighboring residues.

Molecular docking

Molecular docking investigations involving the multiepitope vaccine and the mentioned receptor proteins were conducted using CLUSPRO 2.0 (cluspro.bu.edu/login.php). The Cluspro 2.0 server predicted more than 30 models for the intended proteins each. Among them, the models with the best center energy score and lowest energy score were chosen for subsequent analysis.[46]

Results and Discussion:

This study employed a series of sequential steps, starting with the selection of target proteins. Subsequently, epitopes were predicted from the selected proteins, encompassing MHC-I, MHC-II, CTL, and B cell epitopes. Next, the vaccine was mapped, followed by an assessment of its physicochemical properties. Predictions of secondary and tertiary structures of the vaccine, and molecular docking analyses were then conducted. (**Figure 1**)

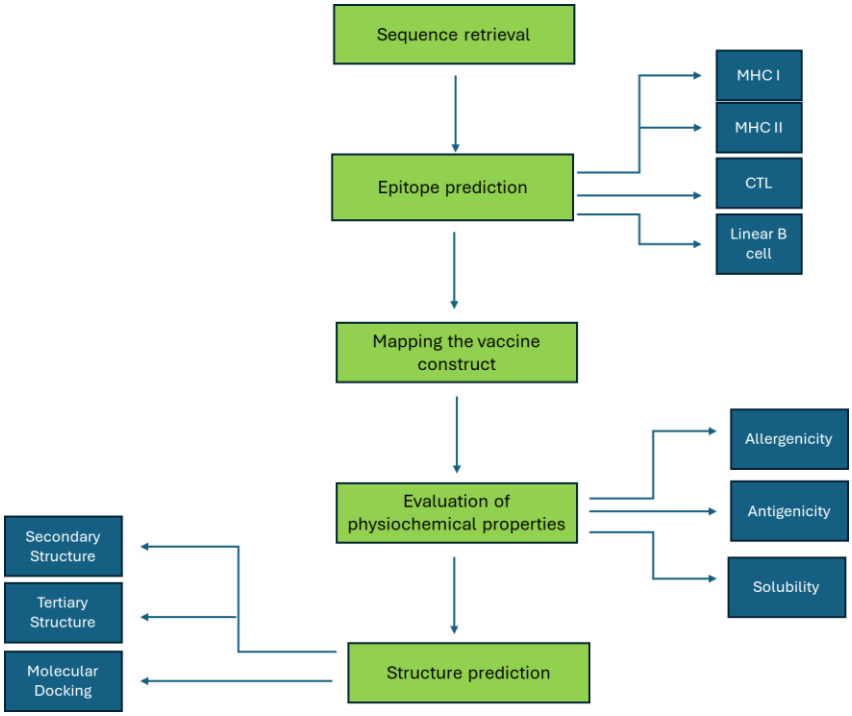


Figure 1. Outline of the study

1. Collection of data and protein sequences

The amino acid sequence of GrB7 (Accession no. AAH06535.1), STAT3 (Accession no. KAI2583079.1), and ERBB2 (Accession no. NP_001369711.1) were retrieved and saved in FASTA format from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih/Blast).

2. Identification and characterization of potential components of the vaccine

The MHC-I antigen presentation pathway serves a critical function in notifying the immune system about virally infected cells. MHC-I molecules are found on the surface of all nucleated cells and display peptide fragments obtained from intracellular proteins.[47] On the other hand, the role of MHC-II is largely limited to professional antigen-presenting cells displaying fragments of processed antigens, mainly derived from external sources to inform the adaptive arm of the immune system.[48] The two however, interact with and activate different subsets of T-cells. Directed priming of the adaptive immune system results in the generation of cytokines, antibodies, and CTL responses against the target cancer cells.

2.1: Prediction of MHC-I Binding Epitopes

T cell-based immune response depends on the presentation of peptide fragments derived from an antigen, as epitopes in the context of MHC-I and MHC-II molecules to $CD8^+$ and $CD4^+$ T cells respectively. The three main criteria set for the vaccine agent were: I. Should contain epitopes recognizable by T cell receptors II. Should be immunogenic but not allergenic and III. Should show strong binding to at least three HLA alleles

The IEDB database, Net-MHC 4.0 server, and RANKPEP server were utilized to predict all potential MHC-I binding epitopes for peptide antigens derived from GRB7, STAT3, and HER2 proteins.[49,50] From each protein, three 9-mer peptides, common to all three databases, were chosen according to their affinity score and percentile rank, as shown in **Table 1**. Subsequently, these peptide sequences were incorporated into the construction of a multi-epitope vaccine.

Three HLA alleles were selected. HLA-A*02:01 is associated with increased expression in cancer patients[51], whereas HLA-B*07:02 and HLA-A*03:01 are two of the most common HLA class-1 molecules present in populations worldwide.[52]

Table 1. Predicted MHC-I epitopes of the three target proteins generated on the IEDB and NetMHC 4.0 databases

Proteins	Start-end (IEDB)	Allele	Peptides (IEDB/NetMCH 4.0/Rankpep)	Affinity (nM) (NetMHC 4.0)	Percentile rank (IEDB)	Method (IEDB)
GRB7	494-502	HLA-A*02:01	**SMDDGQTRF**	4633.27	0.42	Netmhcpan_e1
	384-392	HLA-B*07:02	**VIENPREAL**	1282.31	0.53	Netmhcpan_e1
	427-435	HLA-A*03:01	**RTQLWFHGR**	1995.62	0.61	Netmhcpan_e1
STAT3	279-287	HLA-A*02:01	**QQIKKLEEL**	3166.38	0.55	Netmhcpan_e1
	599-607	HLA-B*07:02	**STKPPGTFL**	2442.37	0.3	Netmhcpan_e1
	340-348	HLA-A*03:01	**KTGVQFTTK**	127.94	0.06	Netmhcpan_e1
HER2	123-131	HLA-A*02:01	**VLIQRNPQL**	156.85	0.05	Netmhcpan_e1
	730-738	HLA-B*07:02	**SPKANKEIL**	97.55	0.04	Netmhcpan_e1
	746-754	HLA-A*03:01	**GVGSPYVSR**	2315.96	0.35	Netmhcpan_e1

Table 2. Predicted MHC-II epitopes of the three target proteins generated on the IEDB and NetMHC 4.0 databases

Protein	Start-end (IEDB)	Allele	Peptides (IEDB/NetMCH 4.0/Rankpep)	Affinity (nM) (NetMHC 4.0)	Percentile rank (IEDB)	IFN-γ (SVM method)	Antigenicity
GRB7	432-446	HLA-DRB1*01:01	**FHGRISREESQRLIG**	68.49	2.9	POSITIVE, 0.11409359	0.5666 (Probable ANTIGEN)
	380-394	HLA-DRB1*03:01	**HAGRVIENPREALSV**	527.54	2.3	POSITIVE,0.55319025	0.5806 (Probable ANTIGEN).
	374-388	HLA-DRB1*07:01	**AMDFSGHAGRVIENP**	105.42	0.89	POSITIVE,0.086354743	0.5989 (Probable ANTIGEN
STAT3	39-53	HLA-DRB1*01:01	**ESQDWAYAASKESHA**	108.47	2.4	POSITIVE, 0.20145532	0.8185 (Probable ANTIGEN)
	329-343	HLA-DRB1*03:01	**MPMHPDRPLVIKTGV**	147.23	2.5	POSITIVE,0.40237428	0.5598 (Probable ANTIGEN)
	40-54	HLA-DRB1*07:01	**SQDWAYAASKESHAT**	589.66	1.6	POSITIVE, 0.12561211	0.8301 (Probable ANTIGEN
HER2	394-408	HLA-DRB1*01:01	**VFQNLQVIRGRILHN**	5.89	2.1	POSITIVE ,0.054692853	0.1720 (Probable NON-ANTIGEN)
	51-65	HLA-DRB1*03:01	**QGYVLIAHNQVRQVP**	220.64	3.1	POSITIVE,0.052173022	0.1742 (Probable NON-ANTIGEN)
	317-331	HLA-DRB1*07:01	**MEHLREVRAVTSANI**	12.61	2.1	POSITIVE, 0.068670127	0.5200 (Probable ANTIGEN)

2.3: Prediction of CTL Epitopes and Binding Affinity

$CD8^+$ CTL are known to be directly involved in destroying cognate cells presenting antigenic peptides in the context of MHC molecules as well as being involved in the formation of long-term immunological memory.[54] Thus, to further ensure strong immunogenicity CTL epitopes of the three target proteins were included in the sequence of the vaccine. The NetCTL 1.2 server was employed to compute CTL epitopes for the specified proteins. From each protein (GRB7, STAT3, and HER2), three peptide epitopes were chosen, considering factors such as MHC binding affinity, C-terminal cleavage affinity, TAP transport efficiency, rescale binding affinity, and combined score.

These selected epitopes were then stored for further analysis, as illustrated in **Table 3**.

Our multi-pronged approach to incorporate MHC-I, MHC-II, CTL, and B cell epitopes was devised to ensure redundancy in terms of stimulating various pathways of the immune system. Adding to that was the incorporation of pathogen recognition receptors (PRR) directed adjuvants which are known to enhance vaccine immunogenicity.[55]

Table 3: List of predicted CTL Epitopes of the three target proteins generated on the CTLpred and NetCTL 1.2 databases

Proteins	Position	Peptides (CTLpred/NetCTL 1.2)	MHC binding Affinity (nM) (NetCTL 1.2)	C terminal cleavage affinity (NetCTL 1.2)	TAP transport efficiency (NetCTL 1.2)	Rescale binding affinity (NetCTL 1.2)	Combined score (NetCTL 1.2)
GRB7	91	GRKLREEER	0.0465	0.8032	1.5080	0.1976	0.3935
	513	HLQKVKHYL	0.0744	0.9711	0.8490	0.3160	0.5041
	317	DVNESNVYV	0.0719	0.9448	0.0410	0.3054	0.4491
STAT3	201	HLQDVRKRV	0.0704	0.9298	0.2090	0.2990	0.4489
	694	YTKQQLNNM	0.0913	0.7983	0.2590	0.3878	0.5205
	553	GTWDQVAEV	0.0678	0.9676	0.2840	0.2879	0.4473
HER2	934	ECRPRFREL	0.0471	0.8229	0.8180	0.1998	0.3641
	682	MRILKETEL	0.0613	0.9730	1.2750	0.2605	0.4702
	312	CYGLGMEHL	0.0523	0.9394	1.0130	0.2222	0.4138

2.4: Linear B-cell epitope prediction

Within the adaptive arm of the immune system, B cells recognize solvent-exposed antigens through their membrane-bound B cell receptors (BCR). Upon maturation, the B cells secrete soluble forms of these BCRs, which are known as antibodies.

Antibodies mediate humoral immunity through opsonization-mediated phagocytic uptake of antigens, as well as through their involvement in forming complement-mediated membrane attack complexes on target cells resulting in their destruction. Even though close to 90% of B cell epitopes are conformational in nature[56], it is common practice to include linear B cell epitopes in multiepitope vaccines to maximize antigenicity.[57,58]

Three peptide sequences, each representing a single peptide from the respective proteins, were identified as linear B cell epitopes via the IEDB server.

Subsequently, the selected peptide sequences underwent assessment for allergenicity and antigenicity using the AllergenFP v.1.0 server and VaxiJen 2.0 server, respectively, as outlined in **Table 4**.

Table 4: Predicted Linear B-cell epitopes of the three target proteins generated using the IEDB and Bepipred 2.0 databases

Protein	Start-end	Allergenicity	Antigenicity	Peptides (IEDB/Bepipred 2.0)
GRB7	5-99	Probable Non-Allergen	0.4417 (Probable Non-Antigen	LSPPHLSSSPEDLCPAPGTPPGTPRPPDTPLP EEVKRSQPLLIPTTGRKLREEERRATSLPSI PNPFPELCSPPSQSPILGGPSSARGLLPRDAS
STAT3	717-766	Probable Non-Allergen	0.7242 (Probable Antigen	TCSNTIDLPMSPRTLDSLMQFGNNGEGAEP SAGGQFESLTFDMELTSECA
HER2	649-695	Probable Non-Allergen	0.6788 (Probable Antigen).	QQKIRKYTMRRLLQETELVEPLTPSGAMP NQAQMRILKETELRKVKV

3. Mapping the vaccine construct

In the formulation of the final multi-epitope vaccine, the most promising epitopes were chosen from GRB7, STAT3, and HER2 proteins, along with three adjuvants: HBHA, MBP, and class C CPG ODNs. In the case of multiepitope sub-unit vaccines, traditional carriers and adjuvants have shown poor efficacy. Hence, the strategy to incorporate adjuvants within the primary sequence of the vaccine has been adopted.[59] Heparin-binding haemagglutinin (HBHA), a factor found in *Mycobacterium tuberculosis* has been identified as a virulence factor linked with agglutination of erythrocytes.[60] HBHA has been later shown to induce T cell-dependent production of HBHA-specific interferon-γ (IFN-γ) and thus its use as an adjuvant in vaccine formulations is common.[61] To further vaccine efficiency, two other adjuvants- maltose-binding protein[62] and large ribosomal subunit protein bL12 (rpIL)[63], of *Escherichia coli* and *Mycobacterium tuberculosis* origin respectively, were included in the primary sequence of the vaccine. An essential part of multi-epitope vaccines is to have the different epitopes fold independently and be presented to the immune system as such. Small polypeptide linkers achieve this function. Rigid EAAAK linkers were used to connect the adjuvant regions. Various linkers, including EAAAK, AYY, AK, KFER, and GPGPG, were utilized to bridge the selected epitopes and adjuvants in the vaccine construct. Subsequently, AYY linkers were employed to connect the MHC-I epitopes as this serves as a site for proteasomal cleavage[64], potentially avoiding the junctional immunogenicity of the epitopes. This was followed by attachment to MHC-II epitopes via GPGPG linkers. GPGPG linkers, a universal spacer motif, were also used between the MHC-II targeted epitopes as they are known to enhance helper T lymphocyte (HTL) responses.[65] AK linkers also facilitated the linkage to and among the CTL epitopes.[66] Additionally, KFER linkers were used to connect the linear B cell epitopes with the rest of the vaccine construct. Finally, the assembly of the final multi-epitope vaccine was completed, allowing for further analysis. (**Figure. 2**)

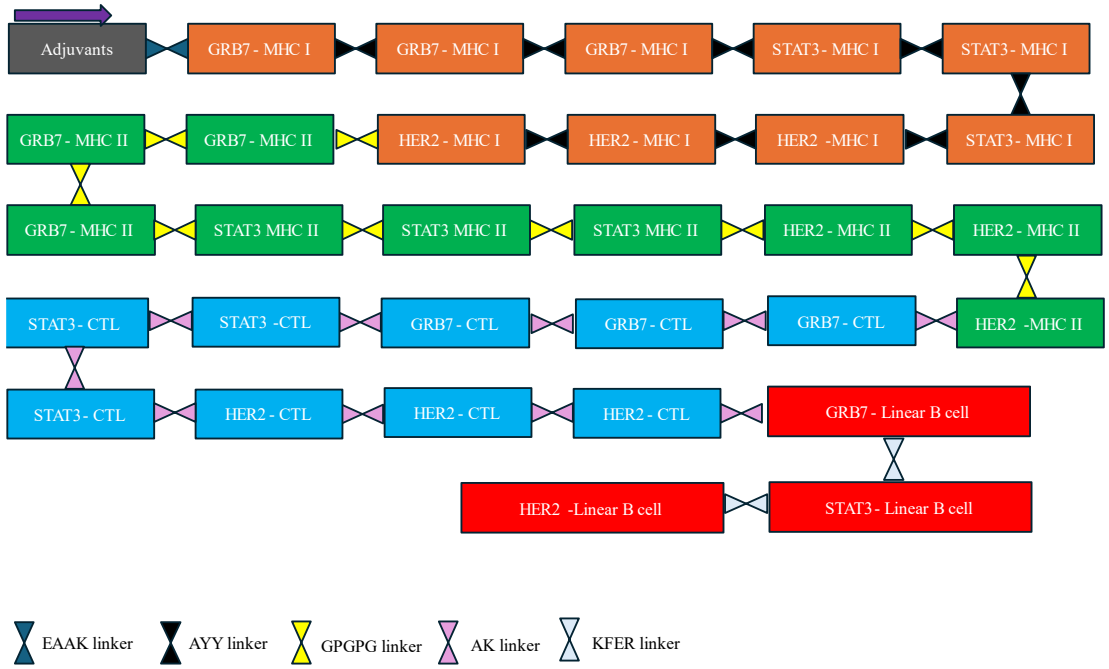


Figure 2: Vaccine was mapped by connecting adjuvants with EAAK linker. Subsequently, all three MHC-I epitopes for each target protein were linked together with the AYY linker. Following this, all the MHC-II epitopes for each target protein were joined with a GPGPG linker. Next, the CTL epitopes were bound together with an AK linker, and linear B cell epitopes for each target protein were linked using a KFER linker.

4. Assessment of Physicochemical Properties, Allergenicity, and Antigenicity

The ExPASy ProtPram server was utilized to evaluate various physicochemical characteristics of the vaccine. These properties are comprised of the total count of amino acids, Molecular weight, Theoretical isoelectric point (pI), etc. Furthermore, the Allergenicity and Antigenicity of the vaccine were assessed by the Vaxijen server the results indicate the vaccine antigen and non-allergen. Finally, the solubility of the vaccine is estimated by the SOLpro server and results showed the vaccine is soluble. All the Physicochemical Properties, Allergenicity, and Antigenicity are depicted in **Table 5.**

Table 5. Predicted physicochemical characteristics of the designed vaccine.

Number of amino acids	1181
Molecular weight	127336.79
Theoretical isoelectric point (pI)	8.64
Total number of negatively charged residues (Asp + Glu)	147
Total number of positively charged residues (Arg + Lys)	157
Predicted Probability of Antigenicity (Vaxijen)	0.736138
Antigenicity/Vaxijen	0.6234 (Probable ANTIGEN)
Allergenicity/AllergenFP v.1.0	Probable non-allergen
Allergenicity/AllerTOP v. 2.0	Probable non-allergen
Allergenicity/AlgPred	-1.1172982
Solubility/SOLpro	Solubility upon over-expression (0.997957)

5. Secondary structure prediction

The secondary structure of the multiepitope vaccine was forecasted using three servers: GOR4, SOPMA, and PSIPRED. GOR4 predicted 49.36% of the structure to fold into alpha helices, 9.82% into extended strands (beta sheet), and 40.81% into the random coil. (Figure 3A) On the other hand, the SOPMA server predicted a total of 45.47% of the structure to exist as alpha helices, 14.56% as extended strands, and 33.95% as random coils (Figure 3B). PSIPRED server was used to perform an amino acid characterization of the primary sequence. 19.1% of the residues were hydrophobic, 6.8% aromatic and cysteines, 33.9% were polar and 40.05% were small non-polar. (Figure 3C).

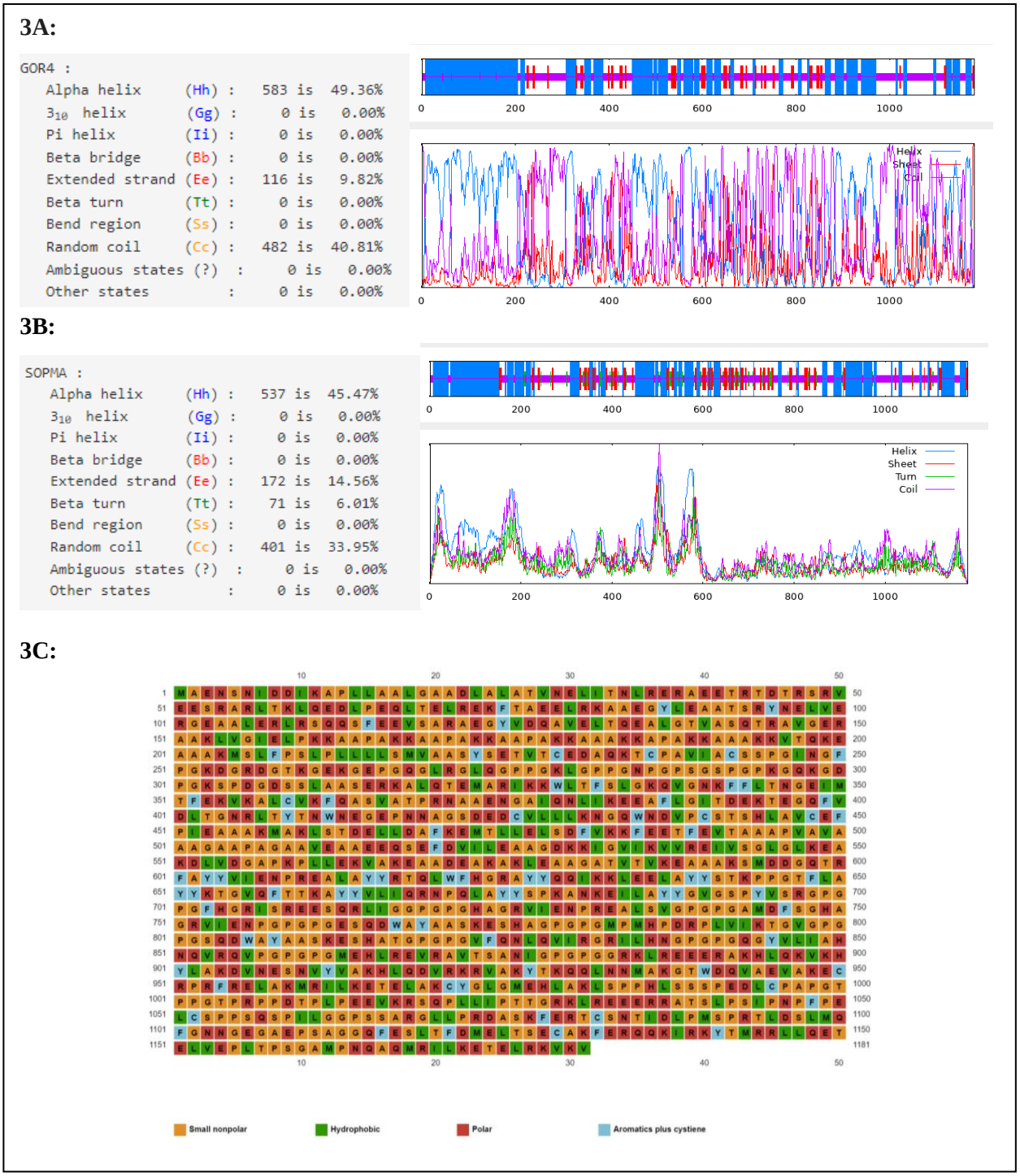


Figure 3. **The graphical representation of the secondary structure configuration of the multi-epitope vaccine as predicted by GOR4 (A) and SOPMA (B). Amino acid characterization within the primary sequence of the vaccine was estimated using the PSIPRED server (C).**

6. Tertiary structure prediction, energy minimization, and validation

The tertiary structure (3D) of the multiepitope vaccine was computed utilizing the Swiss-model server. **(Figure 4A)** Among all tertiary models, we selected the model with the best QMEANDisCo Global score of 0.77.and then this model was used for further refinement on the same server. The post-refinement QMEANDisCo Global score was 0.81 indicating an improvement of the predicted structure with a higher likelihood of resembling the actual 3D structure of the protein. **(Figure 4B)**

To validate the 3D model post-energy minimization, the Ramachandran plot was acquired from the SAVES v6.0 server (https://saves.mbi.ucla.edu/) for assessing the stereochemical quality of the predicted vaccine model. 90.9% of the residues were plotted in the core region, and 9.1% in allowed. This indicated a good overall structure for the vaccine. **(Table 6)**

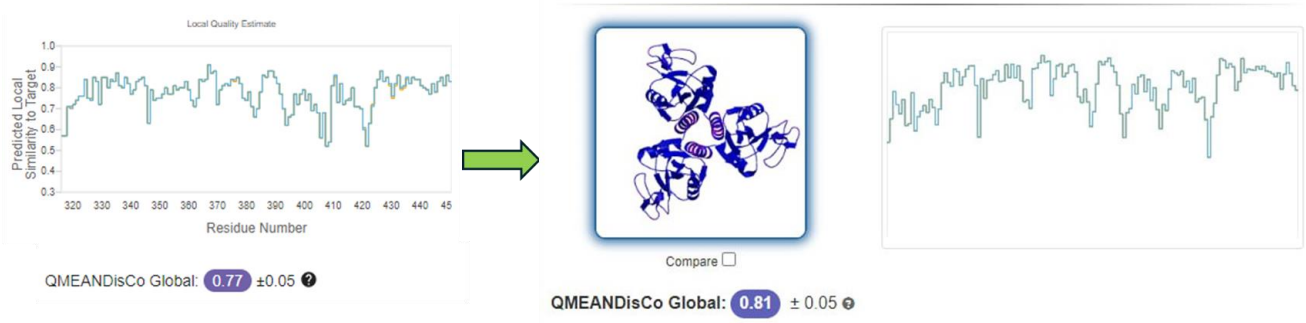


Figure 4: The QMEANDisCo Global score of the multiepitope vaccine before and after energy minimization on the Swiss model server.

```
+----------<<<  P  R  O  C  H  E  C  K     S  U  M  M  A  R  Y  >>>----------+
|                                                                            |
| /var/www/SAVES/Jobs/1662411/saves.pdb   1.5                  408 residues  |
|                                                                            |
| Ramachandran plot:   90.9% core    9.1% allow    0.0% gener    0.0% disall |
|                                                                            |
+| All Ramachandrans:    9 labelled residues (out of 402)                    |
| Chi1-chi2 plots:      0 labelled residues (out of 261)                     |
| Side-chain params:    5 better     0 inside      0 worse                   |
|                                                                            |
*| Residue properties: Max.deviation:     4.9             Bad contacts:    0 |
*|                     Bond len/angle:    5.6    Morris et al class:  1  1  2 |
+|     3 cis-peptides                                                         |
| G-factors           Dihedrals:  -0.17  Covalent:  -0.03    Overall:  -0.10 |
|                                                                            |
+| Planar groups:    94.2% within limits   5.8% highlighted                   |
|                                                                            |
 +--------------------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```

Table 6: Ramachandran plot after energy minimization of the vaccine generated using the UCLA SAVES v6.0 server.

7. Prediction of discontinuous B cell epitopes

Since the vaccine already had linear B cell epitopes built into the primary sequence, we were now interested in looking for possible 3D discontinuous B cell epitopes emerging out of the structure.[67] Using the 3D structure of the multiepitope vaccine as the input, the IEDB database was employed to identify these discontinuous epitopes. A total of four epitopes were predicted with the sequences outlined in **Table 7** and their 3D visualizations are shown in **Figure 5**.

Table 7: Table showing the possible discontinuous B cell epitopes of the vaccine investigated using Ellipro of the IEDB database.

No	Residues	Number of residues	Scores
1	A:R1, A:K2, A:A3, A:L4, A:Q5, A:T6, A:E7, A:M8, A:A9, B:R1, B:K2, B:A3, B:L4, B:Q5, B:T6, B:E7, B:M8, B:A9, C:R1, C:K2, C:A3, C:L4, C:Q5, C:T6, C:E7, C:M8, C:A9	27	0.706
2	A:M35, A:T36, A:F37, A:E38, A:K41, A:V51, A:T53, A:P54, A:R55, A:N56, A:A57, A:G74, A:I75, A:T76, A:D77, A:E78, A:K79, A:T80, A:E81, A:G82, A:Q83, A:F84, A:V85, A:D86, A:L87, A:T88, A:G89, A:N90, A:Y94, A:T95, A:N96, A:W97, A:N98, A:E99, A:G100, A:E101, A:P102, A:N103, A:N104, A:A105, A:G106, A:S107, A:D108, A:E109, A:D110, A:C111, A:V112, A:L115, A:N117, A:G118, A:Q119, A:W120, A:N121, A:D122, A:V123, A:P124, A:C125, A:S126, A:T127, A:S128, A:H129	61	0.697
3	B:M35, B:T36, B:F37, B:E38, B:K41, B:V51, B:T53, B:P54, B:R55, B:N56, B:A57, B:G74, B:I75, B:T76, B:D77, B:E78, B:K79, B:T80, B:E81, B:G82, B:Q83, B:F84, B:V85, B:D86, B:L87, B:T88, B:G89, B:N90, B:Y94, B:T95, B:N96, B:W97, B:N98, B:E99, B:G100, B:E101, B:P102, B:N103, B:N104, B:A105, B:G106, B:S107, B:D108, B:E109, B:D110, B:C111, B:V112, B:L115, B:N117, B:G118, B:Q119, B:W120, B:N121, B:D122, B:V123, B:P124, B:C125, B:S126, B:T127, B:S128, B:H129	61	0.697
4	C:M35, C:T36, C:F37, C:E38, C:K41, C:V51, C:T53, C:P54, C:R55, C:N56, C:A57, C:G74, C:I75, C:T76, C:D77, C:E78, C:K79, C:T80, C:E81, C:G82, C:Q83, C:F84, C:V85, C:D86, C:L87, C:T88, C:G89, C:N90, C:Y94, C:T95, C:N96, C:W97, C:N98, C:E99, C:G100, C:E101, C:P102, C:N103, C:N104, C:A105, C:G106, C:S107, C:D108, C:E109, C:D110, C:C111, C:V112, C:L115, C:N117, C:G118, C:Q119, C:W120, C:N121, C:D122, C:V123, C:P124, C:C125, C:S126, C:T127, C:S128, C:H129	61	0.697

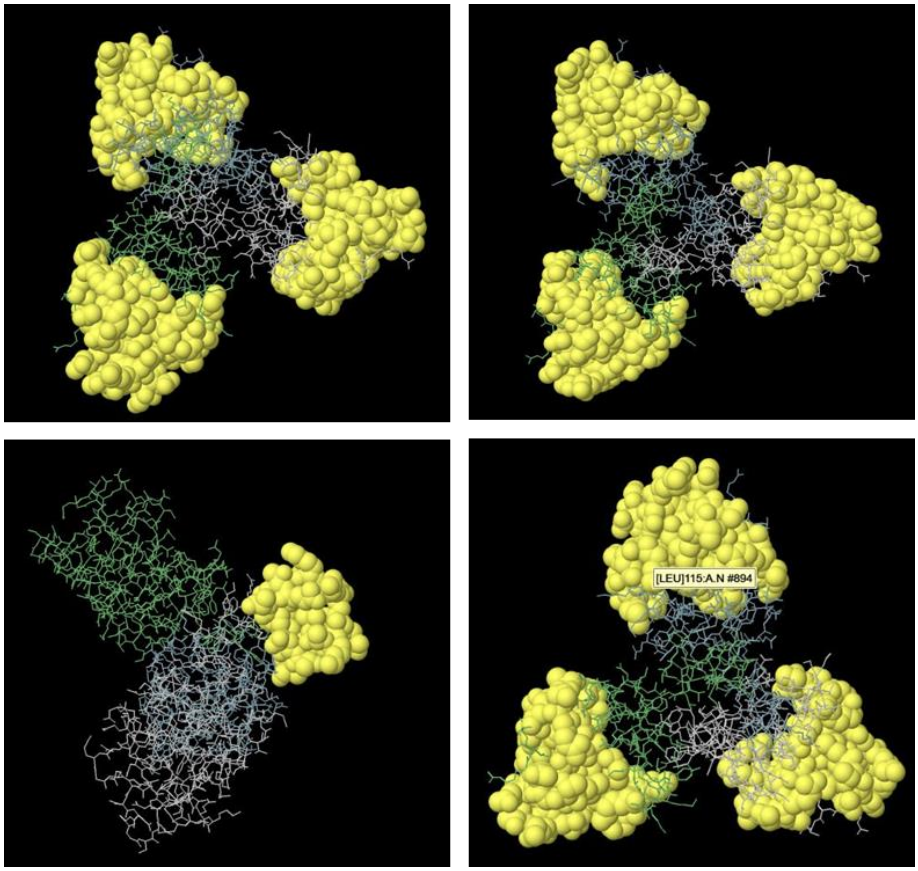


Figure 5: The discontinuous B cell epitopes of the vaccine structure using Ellipro of the IEDB database.

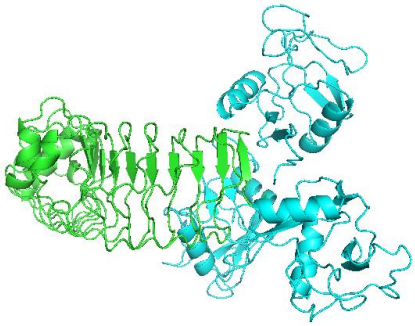
8. Molecular docking study

In order for the multiepitope vaccine to be able to execute its functions in vivo, the inbuilt adjuvant segment would need to interact with PRRs of the innate immune system to trigger the initial response of uptake and processing. Toll-like receptors (TLRs) have a key role in detecting and alerting the innate immune system to the presence of PRRs, which eventually leads to activation and orchestration of the adaptive response.[68] HBHA and MBP have been described in the literature as TLR4 and TLR2 binding PAMPS respectively.[69,70] Additionally, TLR2 has been reported to be a somewhat promiscuous PRR with a varied binding capability.[71] Therefore, the interactions between the inbuilt adjuvant section and TLR4 (**Figure 6A**) and TLR2 were studied using molecular docking analysis on the ClusPro server. Furthermore, the CTL and HTL restricted epitopes interact with MHC-I and MHC-II molecules respectively to activate adaptive immune responses.[72] Therefore, the nine 15-mer peptides were docked against a best representative MHC-II allele (**Figure 6B**), and the nine 9-mer peptides were docked against their respective MHC-I alleles (**Figures 6C and 6D**) on the ClusPro server. The center energy score represents the median value of the largest cluster of binding poses generated by ClusPro and on the other hand, the lowest energy score represents the most stable binding pose between the two interacting molecules. The least binding scores indicate the stronger interaction of the peptides with the target proteins. Most of the docked molecules showed good binding scores with 15-mer peptides generated from Her2 showing the strongest interaction with MHC-II (**Table 8)**

Table 8: The docking results obtained using the ClusPro 2.0 server. Docking complex numbers 2, 5, 16, and 18 are illustrated in Figure 6.

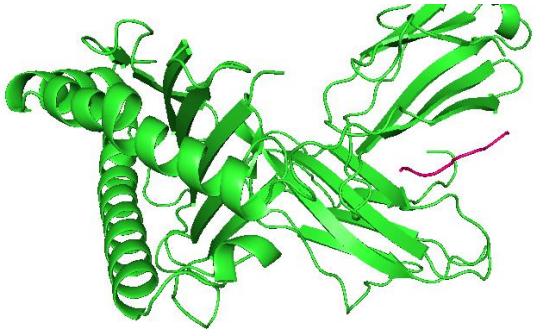
S. No.	Docking complex	Center energy score (kcal/mol)	Lowest energy score (kcal/mol)
1	Adjuvant segment- TLR2	-787.9	-882.4
2	Adjuvant segment- TLR4	-790.8	-914.1
3	GRB7- SMDDGQTRF- HLA-A*02:01- (MHC-I)	-816.9	-816.9
4	GRB7- VIENPREAL- HLA-B*07:02-(MHC-I)	-749.2	-749.2
5	GRB7- RTQLWFHGR- HLA-A*03:01-(MHC-I)	-1118.8	-1118.8
6	STAT3- QQIKKLEEL- HLA-A*02:01-(MHC-I)	-761.8	-774.7
7	STAT3- STKPPGTFL- HLA-B*07:02-(MHC-I)	-832.9	-832.9
8	STAT3-KTGVQFTTK- HLA-A*03:01-(MHC-I)	-835.4	-835.4
9	HER2- VLIQRNPQL- HLA-A*02:01-(MHC-I)	-979.1	-979.1
10	HER2-SPKANKEIL- HLA-B*07:02-(MHC-I)	-689.3	-689.3
11	HER2- GVGSPYVSR- HLA-A*03:01-(MHC-I)	-728.1	-823.5
12	GRB7- FHGRISREESQRLIG -HLA-DR1-(MHC-II)	-1016.2	-1057.9
13	GRB7- HAGRVIENPREALSV- HLA-DR1-(MHC-II)	-856.8	-1048.3
14	GRB7-AMDFSGHAGRVIENP- HLA-DR1-(MHC-II)	-1122.8	-1122.8
15	STAT3- ESQDWAYAASKESHA- HLA-DR1-(MHC-II)	-845.8	-845.8
16	STAT3- MPMHPDRPLVIKTGV- HLA-DR1-(MHC-II)	-1038.2	-1044.2
17	STAT3- SQDWAYAASKESHAT- HLA-DR1-(MHC-II)	-920.0	-920.0
18	HER2- VFQNLQVIRGRILHN- HLA-DR1-(MHC-II)	-1069.6	-1254.4
19	HER2- QGYVLIAHNQVRQVP- HLA-DR1-(MHC-II)	-1065.1	-1065.1
20	HER2- MEHLREVRAVTSANI- HLA-DR1-(MHC-II)	-1005.6	-1005.6

6A: Adjuvants- TLR4



6B: HER2- VFQNLQVIRGRILHN- HLA-DR1(MHC-II)

6C: GRB7- RTQLWFHGR- HLA-A*03:01-(MHC-I)



6D: STAT3- MPMHPDRPLVIKTGV- HLA-DR1-(MHCII)

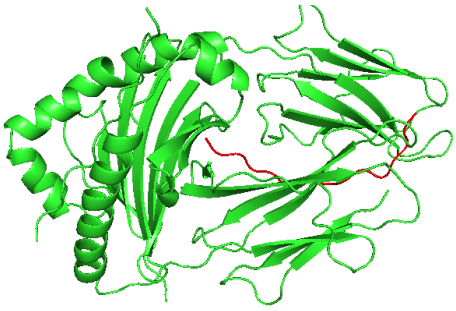


Figure 6: **Figure obtained from molecular docking results. (A) Inbuilt adjuvant section (turquoise) docked against TLR4 (green). (B) One of the three 15-mer HER2 fragments (red) docked against MHC-II (green). (C) One of the three 9-mer fragments (red) of GRB7 docked against an MHC-I (green) allele. (D) One of the three 15-mer fragments (red) of STAT3 docked against an MHC-II (green) allele.**

Discussion

Breast cancer has the highest mortality Rate among cancers in women in Pakistan therefore its earlier detection becomes crucial. Mammography is considered as gold standard in early detection and prevention of breast cancer however in a province like KPK which has meagre health care facilities, BSE should be considered as the go-to method for early detection of breast cancer. BSE is one of the easiest and safest methods to identify any breast abnormality, making it essential for all females to be aware of it and to practice it regularly.

In this study women with mean age of 36.5 years belonging to various districts of KPK having different education level and socioeconomic status participated. Among the participants only 42.7% had heard about BSE, which is relatively low compared to similar studies performed in Karachi, Pakistan[24] and Bangladesh[23], where 71.4% and 60.5% women had heard about BSE respectively. In our study 76.2% women were aware that BSE is important for early detection of breast cancer which is in line with the result in the study performed in Karachi where 74% of the participants considered BSE important for early detection of breast cancer.[24] 7.9% women knew the ideal time for performing BSE was after their monthly cycle which is quite low as compared a study in Iran where 2/3rd of the population were of the view that BSE should be performed after their monthly cycle[28].

Social media (26.2%) proved to be the best source of knowledge about BSE followed by home (25%) and Television (11%) in contrast to similar study conducted in Ethiopia where the most common source of knowledge about BSE was TV and Radio (38.1%) followed health professional (28.9%).[26], which shows the easy access and ever increasing use of social media in KpK.

Although 42.7% women had heard about BSE, Only 3.1% women were substantially aware and 12.9% were partially aware after their knowledge regarding BSE was thoroughly assessed through the 14 indicators in our questionnaire. This shows that the level of awareness about BSE among women of KPK is lower as compared to studies conducted in turkey (34.6%)[27] and Iran (79.8%)[28].

In our study 68.5% women had performed BSE at least once in their life while 41.5% had never performed BSE in their whole life. For the further assessment of these women regarding practice of BSE 7 indicators were used in the questionnaire. Out of 164 substantially aware women, only 26.8% had excellent BSE practice. (Among 164 study participants, only 44(26.8%) had excellent practice of BSE which is low as compared to a study in Ethiopia where 31% of the women had a good practice of BSE [26]

Although the awareness was heartening, the practice rates were disappointing, showing that only one-fourth of the participants performed excellent BSE, a result which was mimicked in the studies conducted in Karachi, Nigeria which had practice rate of 33% and 19% respectively.[24 and 29]

If we focus on enhancing the awareness of BSE further, it would help in increasing the percentage of the population that practices it and hence aids in the early detection of breast cancer. Keeping in mind how various studies have shown that media is the most commonly encountered source of awareness of BSE. Mass media coverage should be given high priority.

Regarding regular performance of BSE, in our finding only 14% of participants had performed BSE once in a month, which is a bit low in contrast to similar studies 31% in African American.[30] As we have addressed already, these differences might be attributed to lack of awareness, knowledge and cultural beliefs in society

A significant association was found between the level of education and awareness of BSE (p=0.032) was found .In our study, among the participants who had inadequate education only 1.2% were substantially aware most importantly, the participants who had done with their graduation 12.9% were substantially aware followed by partially aware(33.9%) in contrast to similar study conducted in Nigeria where only 0% of participants are aware about BSE with inadequate education, while 68.1% of the percipients who had done their graduation are substantially aware of BSE[31]

A significant association between awareness and practice of breast self-examination was also found (p<0.001). The participants who were not aware of BSE only 12.7% had regular practice, participants who were partially aware, 42% had regular practice, while participants who were substantially aware 83.3% had regular practice.

Conclusion

Bioinformatics methods which leverage computational tools are increasingly crucial in translational drug discovery, spanning academia and the pharmaceutical sector. Hence, this study endeavors to design a multi-epitope vaccine using such tools, focusing on key tumor-associated antigens such as GRB7, HER2, and STAT3, known for their roles in tumor proliferation and invasion. Because various alignments of amino acid sequences can result in distinct spatial protein structures, precise sequencing of specific epitopes is essential to yield a functional vaccine.[73,74]

We developed a multiepitope vaccine targeting lung cancer by incorporating three key antigens: GRB7, HER2, and STAT3. Predicted MHC-I and MHC-II binding epitopes, ranging from 9 to 15 amino acids in length, were derived from the GRB7, HER2, and STAT3 proteins. These peptides were tailored to bind specifically to five distinct alleles: HLA-A*02:01, HLA-B*07:02, HLA-A*03:01, DRB1*07:01, and HLA-DRB1*03:01. Furthermore, we assessed the potential for IFN-γ production and antigenicity of MHC-II binding epitopes, resulting in favorable findings. The B cell epitopes we predicted demonstrated high antigenicity while maintaining minimal allergenicity. Subsequently, we employed various linkers and adjuvants to assemble the ultimate multiepitope vaccine.

The designed vaccine exhibited notable antigenicity alongside minimal allergenicity. Additionally, it was estimated to possess a molecular weight (MW) of 55,112.85 Da. With a theoretical isoelectric point (pI) of 9.01, indicating neutrality at a pH of around 9, the vaccine's charge was clarified. According to the Ramachandran plot, the energy-minimized model revealed that 90.9% of amino acid residues resided in the most favored region, with 9.1% occupying the additional allowed region, while none fell within the generously allowed region.

The molecular docking analysis between the constructed vaccine and each of the three targeted proteins was performed using the ClusPro server. Visualizations of the interactions between the constructed vaccine and each protein were generated using PyMOL. The docking results revealed significantly the lowest energy scores (in Kcal/mol) for all targeted proteins.

Acknowledgments:

The authors would like to thank Dr Seetharama Jois, Department of Pathobiological Sciences, LSU School of Veterinary Medicine, LSU, Baton Rouge for his valuable suggestions in the molecular docking section of the study.

References

1. American Lung Association. https://www.lung.org/research/trends-in-lung-disease/lung-cancer-trends-brief/lung-cancer-prevalence-and-incidence.
2. American Cancer Society. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cff-special-section-lung-cancer.pdf.
3. Lahiri, A., Maji, A., Potdar, P.D., Singh, N., Parikh, P., Bisht, B., Mukherjee, A., and Paul, M.K. (2023). Lung cancer immunotherapy: progress, pitfalls, and promises. Mol Cancer *22*, 40. 10.1186/s12943-023-01740-y.
4. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians *68*, 394-424. https://doi.org/10.3322/caac.21492.
5. Sugimura, H., Nichols, F.C., Yang, P., Allen, M.S., Cassivi, S.D., Deschamps, C., Williams, B.A., and Pairolero, P.C. (2007). Survival After Recurrent Nonsmall-Cell Lung Cancer After Complete Pulmonary Resection. The Annals of Thoracic Surgery *83*, 409-418. 10.1016/j.athoracsur.2006.08.046.
6. Kelly, R.J., and Giaccone, G. (2011). Lung cancer vaccines. Cancer J *17*, 302-308. 10.1097/PPO.0b013e318233e6b4.
7. Shepherd, F.A., Dancey, J., Ramlau, R., Mattson, K., Gralla, R., O'Rourke, M., Levitan, N., Gressot, L., Vincent, M., Burkes, R., et al. (2000). Prospective Randomized Trial of Docetaxel Versus Best Supportive Care in Patients With Non–Small-Cell Lung Cancer Previously Treated With Platinum-Based Chemotherapy. Journal of Clinical Oncology *18*, 2095-2103. 10.1200/JCO.2000.18.10.2095.
8. Dobosz, P., and Dzieciątkowski, T. (2019). The Intriguing History of Cancer Immunotherapy. Front Immunol *10*, 2965. 10.3389/fimmu.2019.02965.
9. Dvorak, H.F. (1986). Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N Engl J Med *315*, 1650-1659. 10.1056/nejm198612253152606.
10. Dighe, A.S., Richards, E., Old, L.J., and Schreiber, R.D. (1994). Enhanced in vivo growth and resistance to rejection of tumor cells expressing dominant negative IFN gamma receptors. Immunity *1*, 447-456. 10.1016/1074-7613(94)90087-6.
11. Kaplan, D.H., Shankaran, V., Dighe, A.S., Stockert, E., Aguet, M., Old, L.J., and Schreiber, R.D. (1998). Demonstration of an interferon γ-dependent tumor surveillance system in immunocompetent mice. Proceedings of the National Academy of Sciences *95*, 7556-7561. 10.1073/pnas.95.13.7556.
12. Shankaran, V., Ikeda, H., Bruce, A.T., White, J.M., Swanson, P.E., Old, L.J., and Schreiber, R.D. (2001). IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature *410*, 1107-1111. 10.1038/35074122.
13. Dong, S., Guo, X., Han, F., He, Z., and Wang, Y. (2022). Emerging role of natural products in cancer immunotherapy. Acta Pharm Sin B *12*, 1163-1185. 10.1016/j.apsb.2021.08.020.
14. Fan, T., Zhang, M., Yang, J., Zhu, Z., Cao, W., and Dong, C. (2023). Therapeutic cancer vaccines: advancements, challenges, and prospects. Signal Transduction and Targeted Therapy *8*, 450. 10.1038/s41392-023-01674-3.
15. Jou, J., Harrington, K.J., Zocca, M.-B., Ehrnrooth, E., and Cohen, E.E.W. (2021). The Changing Landscape of Therapeutic Cancer Vaccines—Novel Platforms and Neoantigen Identification. Clinical Cancer Research *27*, 689-703. 10.1158/1078-0432.CCR-20-0245.
16. Yadav, M., Jhunjhunwala, S., Phung, Q.T., Lupardus, P., Tanguay, J., Bumbaca, S., Franci, C., Cheung, T.K., Fritsche, J., Weinschenk, T., et al. (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature *515*, 572-576. 10.1038/nature14001.
17. Riudavets, M., Sullivan, I., Abdayem, P., and Planchard, D. (2021). Targeting HER2 in non-small-cell lung cancer (NSCLC): a glimpse of hope? An updated review on therapeutic strategies in NSCLC harbouring HER2 alterations. ESMO Open *6*, 100260. 10.1016/j.esmoop.2021.100260.
18. Moasser, M.M. (2007). The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene *26*, 6469-6487. 10.1038/sj.onc.1210477.
19. Ferguson, K.M. (2008). Structure-based view of epidermal growth factor receptor regulation. Annu Rev Biophys *37*, 353-373. 10.1146/annurev.biophys.37.032807.125829.
20. Chu, P.-Y., Tai, Y.-L., Wang, M.-Y., Lee, H., Kuo, W.H., and Shen, T.-L. (2020). EGF-Activated Grb7 Confers to STAT3-Mediated EPHA4 Gene Expression in Regulating Lung Cancer Progression. bioRxiv, 2020.2010.2024.353268. 10.1101/2020.10.24.353268.
21. Heinrich, P.C., Behrmann, I., Haan, S., Hermanns, H.M., Müller-Newen, G., and Schaper, F. (2003). Principles of interleukin (IL)-6-type cytokine signalling and its regulation. Biochem J *374*, 1-20. 10.1042/bj20030407.
22. Parakh, S., Ernst, M., and Poh, A.R. (2021). Multicellular Effects of STAT3 in Non-small Cell Lung Cancer: Mechanistic Insights and Therapeutic Opportunities. Cancers *13*. 10.3390/cancers13246228.
23. Yu, H., Lee, H., Herrmann, A., Buettner, R., and Jove, R. (2014). Revisiting STAT3 signalling in cancer: new and unexpected biological functions. Nature Reviews Cancer *14*, 736-746. 10.1038/nrc3818.
24. Gao, S.P., Mark, K.G., Leslie, K., Pao, W., Motoi, N., Gerald, W.L., Travis, W.D., Bornmann, W., Veach, D., Clarkson, B., and Bromberg, J.F. (2007). Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas. J Clin Invest *117*, 3846-3856. 10.1172/jci31871.
25. Haura, E.B., Zheng, Z., Song, L., Cantor, A., and Bepler, G. (2005). Activated epidermal growth factor receptor-Stat-3 signaling promotes tumor survival in vivo in non-small cell lung cancer. Clin Cancer Res *11*, 8288-8294. 10.1158/1078-0432.Ccr-05-0827.
26. Yin, Z., Zhang, Y., Li, Y., Lv, T., Liu, J., and Wang, X. (2012). Prognostic significance of STAT3 expression and its correlation with chemoresistance of non-small cell lung cancer cells. Acta Histochem *114*, 151-158. 10.1016/j.acthis.2011.04.002.

27. Kim, S.M., Kwon, O.J., Hong, Y.K., Kim, J.H., Solca, F., Ha, S.J., Soo, R.A., Christensen, J.G., Lee, J.H., and Cho, B.C. (2012). Activation of IL-6R/JAK1/STAT3 signaling induces de novo resistance to irreversible EGFR inhibitors in non-small cell lung cancer with T790M resistance mutation. Mol Cancer Ther *11*, 2254-2264. 10.1158/1535-7163.Mct-12-0311.
28. Wang, X., and Chen, Z. (2015). MicroRNA-19a functions as an oncogenic microRNA in non-small cell lung cancer by targeting the suppressor of cytokine signaling 1 and mediating STAT3 activation. Int J Mol Med *35*, 839-846. 10.3892/ijmm.2015.2071.
29. Im, J.-Y., Kim, B.-K., Lee, K.-W., Chun, S.-Y., Kang, M.-J., and Won, M. (2020). DDIAS promotes STAT3 activation by preventing STAT3 recruitment to PTPRM in lung cancer cells. Oncogenesis *9*, 1. 10.1038/s41389-019-0187-2.
30. Schlessinger, K., and Levy, D.E. (2005). Malignant Transformation but not Normal Cell Growth Depends on Signal Transducer and Activator of Transcription 3. Cancer Research *65*, 5828-5834. 10.1158/0008-5472.CAN-05-0317.
31. Janes, M.E., Gottlieb, A.P., Park, K.S., Zhao, Z., and Mitragotri, S. (2024). Cancer vaccines in the clinic. Bioeng Transl Med *9*, e10588. 10.1002/btm2.10588.
32. Noguchi, M., Fujimoto, K., Arai, G., Uemura, H., Hashine, K., Matsumoto, H., Fukasawa, S., Kohjimoto, Y., Nakatsu, H., Takenaka, A., et al. (2021). A randomized phase III trial of personalized peptide vaccination for castration-resistant prostate cancer progressing after docetaxel. Oncol Rep *45*, 159-168. 10.3892/or.2020.7847.
33. Weller, M., Butowski, N., Tran, D.D., Recht, L.D., Lim, M., Hirte, H., Ashby, L., Mechtler, L., Goldlust, S.A., Iwamoto, F., et al. (2017). Rindopepimut with temozolomide for patients with newly diagnosed, EGFRvIII-expressing glioblastoma (ACT IV): a randomised, double-blind, international phase 3 trial. The Lancet Oncology *18*, 1373-1385. 10.1016/S1470-2045(17)30517-X.
34. Mitchell, P., Thatcher, N., Socinski, M.A., Wasilewska-Tesluk, E., Horwood, K., Szczesna, A., Martín, C., Ragulin, Y., Zukin, M., Helwig, C., et al. (2015). Tecemotide in unresectable stage III non-small-cell lung cancer in the phase III START study: updated overall survival and biomarker analyses. Ann Oncol *26*, 1134-1142. 10.1093/annonc/mdv104.
35. Narita, Y., Arakawa, Y., Yamasaki, F., Nishikawa, R., Aoki, T., Kanamori, M., Nagane, M., Kumabe, T., Hirose, Y., Ichikawa, T., et al. (2019). A randomized, double-blind, phase III trial of personalized peptide vaccination for recurrent glioblastoma. Neuro Oncol *21*, 348-359. 10.1093/neuonc/noy200.
36. Trevizani, R., Yan, Z., Greenbaum, J.A., Sette, A., Nielsen, M., and Peters, B. (2022). A comprehensive analysis of the IEDB MHC class-I automated benchmark. Brief Bioinform *23*. 10.1093/bib/bbac259.
37. Fleri, W., Paul, S., Dhanda, S.K., Mahajan, S., Xu, X., Peters, B., and Sette, A. (2017). The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. Frontiers in Immunology *8*.
38. Andreatta, M., and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics *32*, 511-517. 10.1093/bioinformatics/btv639.
39. Reche, P.A., Glutting, J.P., Zhang, H., and Reinherz, E.L. (2004). Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics *56*, 405-419. 10.1007/s00251-004-0709-7.
40. Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., and Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics *8*, 424. 10.1186/1471-2105-8-424.
41. Ras-Carmona, A., Lehmann, A.A., Lehmann, P.V., and Reche, P.A. (2022). Prediction of B cell epitopes in proteins using a novel sequence similarity-based method. Scientific Reports *12*, 13739. 10.1038/s41598-022-18021-1.
42. Angeletti, D., and Yewdell, J.W. (2018). Understanding and Manipulating Viral Immunity: Antibody Immunodominance Enters Center Stage. Trends in Immunology *39*, 549-561. 10.1016/j.it.2018.04.008.
43. Garnier, J., Gibrat, J.F., and Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol *266*, 540-553. 10.1016/s0076-6879(96)66034-0.
44. Geourjon, C., and Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci *11*, 681-684. 10.1093/bioinformatics/11.6.681.
45. Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res *31*, 3381-3385. 10.1093/nar/gkg520.
46. Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., and Vajda, S. (2017). The ClusPro web server for protein–protein docking. Nature Protocols *12*, 255-278. 10.1038/nprot.2016.169.
47. Hewitt, E.W. (2003). The MHC class I antigen presentation pathway: strategies for viral immune evasion. Immunology *110*, 163-169. 10.1046/j.1365-2567.2003.01738.x.
48. Rock, K.L., Reits, E., and Neefjes, J. (2016). Present Yourself! By MHC Class I and MHC Class II Molecules. Trends Immunol *37*, 724-737. 10.1016/j.it.2016.08.010.
49. Sanchez-Trincado, J.L., Gomez-Perosanz, M., and Reche, P.A. (2017). Fundamentals and Methods for T- and B-Cell Epitope Prediction. J Immunol Res *2017*, 2680160. 10.1155/2017/2680160.
50. Ahmed, R.K., and Maeurer, M.J. (2009). T-cell epitope mapping. Methods Mol Biol *524*, 427-438. 10.1007/978-1-59745-450-6_31.
51. Stokidis, S., Fortis, S.P., Kogionou, P., Anagnostou, T., Perez, S.A., and Baxevanis, C.N. (2020). HLA Class I Allele Expression and Clinical Outcome in De Novo Metastatic Prostate Cancer. Cancers (Basel) *12*. 10.3390/cancers12061623.
52. Gonzalez-Galarza, F.F., McCabe, A., Santos, E., Jones, J., Takeshita, L., Ortega-Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., et al. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res *48*, D783-d788. 10.1093/nar/gkz1029.
53. Hoof, I., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S., and Nielsen, M. (2009). NetMHCpan, a method for MHC class I binding prediction beyond humans. Immunogenetics *61*, 1-13. 10.1007/s00251-008-0341-z.

54. Doherty, P.C., Allan, W., Eichelberger, M., and Carding, S.R. (1992). Roles of alpha beta and gamma delta T cell subsets in viral immunity. Annu Rev Immunol *10*, 123-151. 10.1146/annurev.iy.10.040192.001011.
55. Pulendran, B., S. Arunachalam, P., and O'Hagan, D.T. (2021). Emerging concepts in the science of vaccine adjuvants. Nature Reviews Drug Discovery *20*, 454-475. 10.1038/s41573-021-00163-y.
56. Walter, G. (1986). Production and use of antibodies against synthetic peptides. J Immunol Methods *88*, 149-161. 10.1016/0022-1759(86)90001-3.
57. Martinelli, D.D. (2022). In silico vaccine design: A tutorial in immunoinformatics. Healthcare Analytics *2*, 100044. https://doi.org/10.1016/j.health.2022.100044.
58. Galanis, K.A., Nastou, K.C., Papandreou, N.C., Petichakis, G.N., Pigis, D.G., and Iconomidou, V.A. (2021). Linear B-Cell Epitope Prediction for In Silico Vaccine Design: A Performance Review of Methods Available via Command-Line Interface. Int J Mol Sci *22*. 10.3390/ijms22063210.
59. Lei, Y., Zhao, F., Shao, J., Li, Y., Li, S., Chang, H., and Zhang, Y. (2019). Application of built-in adjuvants for epitope-based vaccines. PeerJ *6*, e6185. 10.7717/peerj.6185.
60. Parra, M., Pickett, T., Delogu, G., Dheenadhayalan, V., Debrie, A.S., Locht, C., and Brennan, M.J. (2004). The mycobacterial heparin-binding hemagglutinin is a protective antigen in the mouse aerosol challenge model of tuberculosis. Infect Immun *72*, 6799-6805. 10.1128/iai.72.12.6799-6805.2004.
61. Temmerman, S., Pethe, K., Parra, M., Alonso, S., Rouanet, C., Pickett, T., Drowart, A., Debrie, A.-S., Delogu, G., Menozzi, F.D., et al. (2004). Methylation-dependent T cell immunity to Mycobacterium tuberculosis heparin-binding hemagglutinin. Nature Medicine *10*, 935-941. 10.1038/nm1090.
62. Jie, J., Zhang, Y., Zhou, H., Zhai, X., Zhang, N., Yuan, H., Ni, W., and Tai, G. (2018). CpG ODN1826 as a Promising Mucin1-Maltose-Binding Protein Vaccine Adjuvant Induced DC Maturation and Enhanced Antitumor Immunity. Int J Mol Sci *19*. 10.3390/ijms19030920.
63. Lee, S.J., Shin, S.J., Lee, M.H., Lee, M.-G., Kang, T.H., Park, W.S., Soh, B.Y., Park, J.H., Shin, Y.K., Kim, H.W., et al. (2014). A Potential Protein Adjuvant Derived from Mycobacterium tuberculosis Rv0652 Enhances Dendritic Cells-Based Tumor Immunotherapy. PLOS ONE *9*, e104351. 10.1371/journal.pone.0104351.
64. Bhatnager, R., Bhasin, M., Arora, J., and Dang, A.S. (2021). Epitope based peptide vaccine against SARS-COV2: an immune-informatics approach. J Biomol Struct Dyn *39*, 5690-5705. 10.1080/07391102.2020.1787227.
65. Livingston, B., Crimi, C., Newman, M., Higashimoto, Y., Appella, E., Sidney, J., and Sette, A. (2002). A Rational Strategy to Design Multiepitope Immunogens Based on Multiple Th Lymphocyte Epitopes1. The Journal of Immunology *168*, 5499-5506. 10.4049/jimmunol.168.11.5499.
66. Ayyagari, V.S., T, C.V., K, A.P., and Srirama, K. (2022). Design of a multi-epitope-based vaccine targeting M-protein of SARS-CoV2: an immunoinformatics approach. J Biomol Struct Dyn *40*, 2963-2977. 10.1080/07391102.2020.1850357.
67. Haste Andersen, P., Nielsen, M., and Lund, O. (2006). Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci *15*, 2558-2567. 10.1110/ps.062405906.
68. Fitzgerald, K.A., and Kagan, J.C. (2020). Toll-like Receptors and the Control of Immunity. Cell *180*, 1044-1066. 10.1016/j.cell.2020.02.041.
69. Lei, Y., Shao, J., Ma, F., Lei, C., Chang, H., and Zhang, Y. (2020). Enhanced efficacy of a multi-epitope vaccine for type A and O foot-and-mouth disease virus by fusing multiple epitopes with Mycobacterium tuberculosis heparin-binding hemagglutinin (HBHA), a novel TLR4 agonist. Molecular Immunology *121*, 118-126. https://doi.org/10.1016/j.molimm.2020.02.018.
70. Xiaoxia, Z., Weihua, N., Qingyong, Z., Fengli, W., Yingying, L., Xiaxia, S., Zhonghui, L., and Guixiang, T. (2011). Maltose-binding protein isolated from Escherichia coli induces Toll-like receptor 2-mediated viability in U937 cells. Clin Transl Oncol *13*, 509-518. 10.1007/s12094-011-0689-7.
71. de Oliviera Nascimento, L., Massari, P., and Wetzler, L.M. (2012). The Role of TLR2 in Infection and Immunity. Frontiers in Immunology *3*.
72. Kumar, N., Sood, D., Sharma, N., and Chandra, R. (2020). Multiepitope Subunit Vaccine to Evoke Immune Response against Acute Encephalitis. Journal of Chemical Information and Modeling *60*, 421-433. 10.1021/acs.jcim.9b01051.
73. Heidary, F., Tourani, M., Hejazi-Amiri, F., Khatami, S.H., Jamali, N., and Taheri-Anganeh, M. (2022). Design of a new multi-epitope peptide vaccine for non-small cell Lung cancer via vaccinology methods: an in silico study. Mol Biol Res Commun *11*, 55-66. 10.22099/mbrc.2022.42468.1697.
74. Tourani, M., Karkhah, A., and Najafi, A. (2017). Development of an epitope-based vaccine inhibiting immune cells rolling and migration against atherosclerosis using in silico approaches. Comput Biol Chem *70*, 156-163. 10.1016/j.compbiolchem.2017.08.016.