WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.453
Volume 14, Issue 16, 897-916. Research Article ISSN 2277–7105

# MOLECULAR DOCKING AND STRUCTURAL VALIDATION OF A MULTI-EPITOPE VACCINE CONSTRUCT TARGETING EPSTEIN-BARR VIRUS VIA COMPUTER-BASED METHODS

**[1*]Mr. Jayaseelan K., [2]Dr. Senthil Kumar SK., [3]Mr. Mani Raj S., [4]Ms. Udhaya Lakshmi R., [5]Mr. Vasanth Kumar R., [6]Mr. Vimal Raj M., [7]Mr. Vinoth J.**

[1*]Assistant Professor, Department of Pharmaceutical Analysis, Arunai College of Pharmacy, Velu Nagar, Thenmathur, Thiruvannamalai-606 603.

[2]Principal, Arunai College of Pharmacy, Velu Nagar, Thenmathur, Thiruvannamalai-606 603.

[3,4,5,6,7]B. Pharmacy Final Year Students, Arunai College of Pharmacy, Velu Nagar, Thenmathur, Thiruvannamalai-606 603.

Article Received on
01 July 2025,

Revised on 21 July 2025,
Accepted on 10 August 2025,

DOI: 10.20959/wjpr202516-37949

***Corresponding Author**
**Mr. Jayaseelan K.**
Assistant Professor, Department of Pharmaceutical Analysis, Arunai College of Pharmacy, Velu Nagar, Thenmathur, Thiruvannamalai-606 603.

## ABSTRACT

The Epstein-Barr Virus (EBV) is associated with various malignancies, including Burkitt's lymphoma and nasopharyngeal carcinoma. Due to this widespread impact, an effective messenger RNA (mRNA) vaccine is paramount to help curb its spread, further underscoring the need for its development. Cytotoxic T lymphocyte (CTL), helper T lymphocyte (HTL), and B-cell epitopes were predicted using online immunoinformatics tools. The most antigenic and non- allergenic epitopes were linked using appropriate linkers and fused with an adjuvant (e.g., β-defensin or TLR agonist) to enhance immunogenicity. The 3D structure of the vaccine construct was modeled, refined, and validated using tools such as **Swiss model, IEDB, Toxinpred,** and **uniport.** Molecular docking was conducted between the vaccine construct and immune receptors, particularly **Toll-like receptors (TLRs)** such as TLR2, TLR4, and TLR5, using **Autodock Vina**. The binding affinity, hydrogen bonding, and interaction energies were analyzed to assess the stability of vaccine-receptor complexes.

**KEYWORDS:** Epstein Barr Virus; multi-epitopes mRNA vaccine; molecular docking; allergic epitopes.

**Table of Contents**



## INTRODUCTION

EBV is further classified into two primary subgroups (type 1 and type 2); these two groups differ mainly in the EBNA-3 (nuclear immunogen-3 gene). Both subgroups are reported worldwide; however, in most of the populations, type 1 is dominant. Sore throat, swelling, fatigue, fever, swollen lymph nodes, and rashes on the skin are the clinical symptoms of infections with EBV. The virus is transmitted orally and is usually transmitted through contact with close family in childhood and infancy. Among the ways to prevent contracting

the virus is to reduce close contact with people suspected or infected with the virus and avoiding using an ordinary toothbrush, sharing food, or exchanging bodily fluids. Immunity against EBV has been studied extensively. Natural killer (NK) cells play an important role in the innate immune response, delaying or preventing the EBV transformation of B cells through the production of interferon gamma (IFN-g). Subsequently, the virus elicits strong adaptive immune responses, primarily mediated by cytotoxic CD8 T cells. CD8 T cell responses eliminate viral-infected cells upon recognition of EBV peptide antigens bound to MHC I molecules in the surface of target cells.

Although it affects about 90% of adults worldwide, inactive latent EBV usually does not cause severe health problems. In contrast, the virus can cause diseases such as Burkitt's lymphoma (BL), Hodgkin lymphoma (HL), nasopharyngeal carcinoma and gastric cancer.

This study employs a comprehensive computational approach to design a MEV construct targeting EBV. Key stages include epitope prediction, vaccine construct modeling, molecular docking with immune receptors (e.g., TLRs, MHC molecules), and structural validation using bioinformatics tools. Molecular docking is a pivotal step in assessing the binding affinity and stability of the vaccine- receptor complex, while structural validation ensures the conformational integrity and quality of the construct.

By leveraging computer-based methods, this study provides a framework for rapid, cost-effective, and rational vaccine design, offering a promising avenue toward combating EBV and its associated diseases.

**Docking**

Docking in the context of molecular biology and computational chemistry refers to a method used to predict the interaction between two molecules, typically

- A small molecule (ligand) and a macromolecule (usually a protein) Example: Drug binding to a viral protein

Molecular docking is a computational technique that predicts the preferred orientation and binding affinity of one molecule (e.g., ligand) when it binds to another (e.g., protein), forming a stable complex. Usually docking is used for

- Predict how drugs or peptides bind to **target proteins**
- Understand **binding sites**, **interaction residues**, and **binding energy**

- Aid in **drug design**, **vaccine development**, and **protein function prediction**

**In silico Vaccine design**

**In silico vaccine design** is the process of utilizing computational biology, bioinformatics algorithms, and computer simulations to predict and design potential vaccine candidates by identifying immunogenic epitopes, optimizing antigen structures, and simulating immune responses, thereby accelerating vaccine development and reducing the need for extensive laboratory testing in the early stages.

It is Used in the SARS-CoV-2 (COVID-19) epitope-based vaccines, Tuberculosis multi-epitope vaccines, Epstein Bar virus, HPV, HIV, Zika virus computational vaccine design.

**Protein Preparation and Control Ligand Selection**

The three-dimensional (3D) crystal structure of EBV gL with a resolution of 4.80 Å was retrieved from the RCSB Protein Data Bank (PDB) with PDB ID 6C5V. Since only gL was our target protein against which docking of candidate compounds needed to occur, we removed the glycoprotein H, glycoprotein 42, other protein chains, protein cofactors, water molecules, and metal ions. The protein cofactors and metal ions were removed to skip unwanted interference during docking simulations. Subsequently, we added polar hydrogen atoms and merged non-polar hydrogen atoms using BIOVIA Discovery Studio Visualizer.

Gasteiger charges were calculated using AutoDock tools. Finally, the control ligand, NAG (2-acetamido-2-deoxy-beta-D-glucopyranose), associated was downloaded from the PubChem database with PubChem.

**Antigen antibody preparation**

Antigen-antibody preparation refers to the experimental process in which specific antigens and their corresponding antibodies are combined under controlled laboratory conditions to form immune complexes. This preparation is essential for various immunological assays and diagnostic applications, including ELISA, immunoprecipitation, Western blotting, and immunohistochemistry. The procedure typically involves selecting high-purity antigens and antibodies, optimizing their concentrations, and incubating them under conditions that preserve specificity and binding affinity, ensuring accurate and reproducible results.

**Parameters for insilico vaccine**

1. **Epitope Prediction**: Identification of B-cell and T-cell epitopes based on antigenicity, binding affinity to MHC molecules, and immunogenic potential.
2. **Antigenicity**: Assessment of the candidate's ability to elicit an immune response, typically evaluated using prediction tools like VaxiJen or ANTIGENpro.
3. **Allergenicity and Toxicity**: Evaluation to ensure the vaccine candidate is non-allergenic and non-toxic using tools such as AllerTOP, AllergenFP, and ToxinPred.
4. **MHC Binding Affinity**: Prediction of strong binding to MHC class I and II alleles to ensure effective antigen presentation.
5. **Population Coverage**: Analysis of epitope-HLA allele binding across different ethnic populations to ensure broad immunogenic coverage.
6. **Conservancy Analysis**: Ensures selected epitopes are conserved across multiple strains or variants of the pathogen.
7. **Physicochemical Properties**: Evaluation of molecular weight, isoelectric point, stability index, aliphatic index, and hydropathicity to determine vaccine suitability.
8. **3D Structure Modeling and Docking**: Structural prediction and molecular docking simulations of vaccine constructs with immune receptors (e.g., TLRs, MHC) to assess interaction and binding efficiency.
9. **Codon Optimization and In Silico Cloning**: Optimization of gene sequences for expression in a suitable host (commonly *E. coli*) and simulated cloning into expression vectors.
10. **Immune Simulation**: Prediction of immune responses using computational models to assess the dynamics of antibody production, cytokine release, and memory cell formation.

**Binding Energy**

**Binding energy** is the energy required to **disassemble a system into its individual components**. In nuclear physics and chemistry, it's most often used in two main contexts.

**Nuclear energy**

It explains why mass is lost when a nucleus forms known as the **mass defect**. This "lost mass" is converted into binding energy via **Einstein's equation.** High binding energy is equal to more stable nucleus.

**In silico vaccine design refers to the use of computational tools to**

| Purpose | tools |
|---|---|
| Genome retrival<br>Antigenic Protein Prediction Prediction of B-cell Epitope Prediction of T-cell Epitope Allergenicity and Toxicity Screening Epitope Conservancy Analysis | Uniport Vaxijen v2.0 ABC pred IEDB tools Toxin pred<br><br>IEDB tools |
| 3D Structure modeling | Swiss model |
| Molecular dynamics simulation of the vaccine-receptor complexes | Swiss dock |
| Molecular docking | Autodock |

**Epitope preparation**

The insilico vaccine design, epitope preparation means identifying and selecting short, antigenic parts of a pathogen's proteins (called epitopes) that can trigger an immune response.

These epitopes are used to build multi-epitope vaccines.

Identify Epitopes

- T-cell epitopes ($CD8^+$ and $CD4^+$): predicted using IEDB tools.
- B-cell epitopes: predicted using ABCpred.

**Protein Selection**

The Epstein-Barr virus (EBV), a member of the Herpesviridae family, is known for its ability to establish persistent infections within host cells, primarily through a state of latency. During latency, EBV expresses a subset of proteins that maintain the viral genome, evade immune responses, and drive oncogenic transformation, making them ideal targets for a vaccine. Among these are EBNA1, which play pivotal roles in promoting Epstein barr virus transformations within infected cells.

EBNA1: EBNA1 is indispensable for maintaining the viral genome in host cells by tethering the episomal DNA to host chromosomes during cell division. It facilitates immune evasion by inhibiting antigen presentation through its Gly-Ala repeat domain, reducing recognition by cytotoxic T lymphocytes. Additionally, EBNA1 can modulate host cellular pathways to promote immune evasion and tumor progression. Its consistent expression in latency types I, II, and III, but absence in latency 0, makes it a prime target for therapeutic intervention. In NPC, which predominantly exhibits latency type II, EBNA1 plays a crucial role in sustaining

viral persistence and driving oncogenesis through immune evasion and tumor- supportive mechanisms.

**Steps involved in insilico vaccine development**

I. Genome retrieval
II. Antigenic protein prediction
III. Epitope mapping
IV. Prediction of Conformational B-Cell Epitope
V. Prediction of Conformational T-Cell Epitopes
VI. Assessment of Allergenicity and Toxicity screening
VII. Epitope conservancy analysis
VIII. Modeling of the 3D Structure
IX. Molecular dynamics simulation of vaccine-receptor complex
X. Refinement analysis
XI. Molecular Docking and Binding Affinity Analysis

**Methodology and result of insilico vaccine development**

**Step 1: Genome Retrieval**

Genome Retrieval in In Silico Vaccine Design is the *first and foundational step* in computational vaccine development. It involves obtaining the complete or partial genome (especially coding regions of interest) of a pathogen to identify potential vaccine targets. All sequences were extracted in FASTA format and will be used for epitope prediction, structural modeling, and immunogenicity assessments in the vaccine development pipeline. All sequences were extracted in FASTA format and will be used for epitope prediction, structural modeling, and immunogenicity assessments in the vaccine development pipeline.

```
>sp|P03211|EBNA1_EBVB9 Epstein-Barr nuclear antigen 1 OS=Epstein-Barr
virus (strain B95-8) OX=10377 GN=EBNA1 PE=1 SV=1
MSDEGPGTGPGNGLGEKGDTSGPEGSGGSGPQRRGGDNHGRGRGRGRGGGRPGAPGGS
GSGPRHRDGVRRPQKRPSCIGCKGTHGGTGAGAGAGGAGAGGAGAGGGAGAGGGAGGAGG
AGGAGAGGGAGAGGGAGGAGGAGAGGGAGAGGGAGGAGAGGGAGGAGAGGGAGAGGG
AGGAGAGGGAGGAGGAGAGGGAGAGGAGGAGAGGAGAGGGAGGAGGAGAGGAGAGGA
GAGGAGAGGAGGAGAGGAGGAGAGGAGGAGAGGGAGGAGAGGGAGGAGAGGAGGAGAGGA
GGAGAGGAGGAGAGGGAGAGGAGAGGGGRGRGGSGGRGRGGSGGRGRGGSGGRRGRGRER
ARGGSRERARGRGRGRGEKRPRSPSSQSSSSGSPPRRPPPGRRPFFHPVGEADYFEYHQE
GGPDGEPDVPPGAIEQGPADDPGEGPSTGPRGQGDGGRRKKGGWFGKHRGQGGSNPKFEN
IAEGLRALLARSHVERTTDEGTWVAGVFVYGGSKTSLYNLRRGTALAIPQCRLTPLSRLP
FGMAPGPGPQPGPLRESIVCYFMVFLQTHIFAEVLKDAIKDLVMTKPAPTCNIRVTVCSF
DDGVDLPPWFPPMVEGAAAEGDDGDDGDEGGDGDEGEEGQE
>sp|P12978|EBNA2_EBVB9 Epstein-Barr nuclear antigen 2 OS=Epstein-Barr
virus (strain B95-8) OX=10377 GN=EBNA2 PE=1 SV=1
MPTFYLALHGGQTYHLIVDTDSLGNPSLSVIPSNPYQEQLSDTPLIPLTIFVGENTGVPP
PLPPPPPPPPPPPPPPPPPPPPPPPPPSPPPPPPPPPPQRRDAWTQEPSPLDRDPLGY
DVGHGPLASAMRMLWMANYIVRQSRGDRGLILPQGPQTAPQARLVQPHVPPLRPTAPTIL
SPLSQPRLTPPQPLMMPPRPTPPTPLPPATLTVPPRPTRPTTLPPTPLLTVLQRPTELQP
TPSPPRMHLPVLHVPDQSMHPLTHQSTPNDPDSPEPRSPTVFYNIPPMPLPPSQLPPPAA
PAQPPPGVINDQQLHHLPSGPPWWPPICDPPQPSKTQGQSRGQSRGRGRGRGRGKGKS
RDKQRKPGGPWRPEPNTSSPSMPELSPVLGLHQGQGAGDSPTPGPSNAAPVCRNSHTATP
NVSPIHEPESHNSPEAPILFPDDWYPPSIDPADLDESWDYIFETTESPSSDEDYVEGPSK
RPRPSIQ
```

EBNA1 sequences were obtained from three different Epstein-Barr virus (EBV) strains with the following UniProt accession numbers: P03211 (strain B95-8), P12978 (strain GD1), and Q1HVF7 (strain AG876), each consisting of 641 amino acids.

**Step2: Antigenic Protein Prediction**

Protein sequences in this study were considered to screen out using VaxiJen web server for the identification of potent antigenic protein. The most potent antigenic protein having a maximum total prediction score of 0.5416. Here the threshold of 0.4 is considered as the potent antigenicity. This sequence was used for further analysis.


Bioinformatics tools help screen for outer membrane proteins and virulence factors.

**Step 3: Epitope Mapping**

Epitope mapping in Epstein-Barr Virus (EBV) is a crucial technique used to identify specific regions (epitopes) of EBV proteins that are recognized by the immune system, particularly by B cells (antibody responses) and T cells (CD4+ and CD8+ responses). This helps in understanding immune control of EBV, guiding vaccine development, and identifying targets for immunotherapy.

**Step 4: Prediction of B-cell Epitope**

To ABCpred server predicted 58 B-cell epitopes: five epitopes were for GP 42, eight for GP H, nineteen for GP B, one for GP L and GP N, five for GP M, and nineteen for GP 350. However, epitopes with scores above 0.99 were selected as the most potentially antigenic epitopes. Therefore, only one epitope each from GP42, GL, GM, GN and GH; four epitopes

from GB; and fifteen epitopes from GP350 were found to meet the threshold value.

## TABULAR RESULT

### Predicted B-cell epitope

**The predicted B cell epitopes are ranked according to their score obtained by trained recurrent neural network.**
**Higher score of the peptide means the higher probability to be as epitope.**
**All the peptides shown here are above the threshold value chosen.**

| Rank | Sequence | Start position | Score |
|---|---|---|---|
| 1 | EGSGGSGPQRRGGDNH | 86 | 0.95 |
| 1 | PGAIEQGPADDPGEGP | 493 | 0.95 |
| 1 | GGSGSGPRHRDGVRRP | 120 | 0.95 |
| 2 | HGRGRGRGRGRGGGRP | 101 | 0.94 |
| 3 | AAEGDDGDDGDEGGDG | 680 | 0.93 |
| 3 | EYHQEGGPDGEPDVPP | 478 | 0.93 |
| 4 | PGTGPGNGLGEKGDTS | 68 | 0.91 |
| 4 | PRGQGDGGRRKKGGWF | 512 | 0.91 |
| 5 | LVMTKPAPTCNIRVTV | 644 | 0.90 |
| 5 | KGGWFGKHRGQGGSNP | 523 | 0.90 |
| 5 | GGAGGAGAGGGAGGAG | 335 | 0.90 |
| 5 | GGAGGAGAGGGAGGAG | 326 | 0.90 |
| 5 | GAGGGAGGAGGAGAGG | 281 | 0.90 |
| 5 | GAGGGAGGAGGAGAGG | 247 | 0.90 |
| 5 | GGAGGAGAGGGAGGAG | 241 | 0.90 |
| 5 | GAGGGAGGAGGAGAGG | 220 | 0.90 |
| 5 | GGAGGAGAGGGAGGAG | 214 | 0.90 |
| 5 | GAGGGAGGAGGAGAGG | 193 | 0.90 |

The selected epitopes were further evaluated for antigenicity using VaxiJen 2.0 (threshold > 0.4), allergenicity using AllerTop 2.0, and toxicity using ToxinPred.

**Step 5: Prediction of T-cell epitope**

Epitope predictions for all seven proteins were conducted on Net CTL, an online epitope prediction server. MHC-I binding prediction using the SMM method resulted in many potential epitopes against one allele, HLA-A*24:02. The weight matrix and artificial neural network was used for the prediction of MHC-I binding and proteasome dependent C-terminal

MHC-I Binding Prediction Results

Method used: netmhcpan_el

| allele | seq_num | start | end | length | peptide | score | percentile_rank | |
|---|---|---|---|---|---|---|---|---|
| HLA-B*07:02 | 1 | 397 | 405 | 9 | RPPPGRRPF | 0.966589 | 0.02 | |
| HLA-B*35:01 | 1 | 407 | 415 | 9 | HPVGEADYF | 0.964565 | 0.02 | |
| HLA-A*03:01 | 1 | 578 | 586 | 9 | AIKDLVMTK | 0.958049 | 0.01 | |
| HLA-A*11:01 | 1 | 578 | 586 | 9 | AIKDLVMTK | 0.933216 | 0.01 | |
| HLA-B*51:01 | 1 | 606 | 614 | 9 | LPPWFPPMV | 0.916408 | 0.02 | |
| HLA-B*07:02 | 1 | 72 | 80 | 9 | RPQKRPSCI | 0.907598 | 0.04 | |
| HLA-A*30:01 | 1 | 578 | 586 | 9 | AIKDLVMTK | 0.901802 | 0.01 | |
| HLA-B*07:02 | 1 | 546 | 554 | 9 | GPGPQPGPL | 0.890273 | 0.05 | |
| HLA-B*53:01 | 1 | 407 | 415 | 9 | HPVGEADYF | 0.860752 | 0.03 | |
| HLA-B*07:02 | 1 | 380 | 388 | 9 | RPRSPSSQS | 0.853895 | 0.06 | |
| HLA-B*07:02 | 1 | 528 | 536 | 9 | IPQCRLTPL | 0.834495 | 0.07 | |
| HLA-B*08:01 | 1 | 518 | 526 | 9 | YNLRRGTAL | 0.822952 | 0.03 | |
| HLA-A*68:01 | 1 | 577 | 586 | 10 | DAIKDLVMTK | 0.817731 | 0.19 | |
| HLA-B*07:02 | 1 | 397 | 406 | 10 | RPPPGRRPFF | 0.804968 | 0.08 | |
| HLA-A*31:01 | 1 | 514 | 522 | 9 | KTSLYNLRR | 0.802294 | 0.06 | |
| HLA-B*07:02 | 1 | 550 | 558 | 9 | QPGPLRESI | 0.798453 | 0.08 | |
| HLA-B*07:02 | 1 | 396 | 405 | 10 | RRPPPGRRPF | 0.769944 | 0.09 | |
| HLA-B*35:01 | 1 | 563 | 571 | 9 | MVFLQTHIF | 0.759642 | 0.09 | |
| HLA-A*68:01 | 1 | 388 | 396 | 9 | SSSSGSPPR | 0.756656 | 0.25 | |
| HLA-A*68:01 | 1 | 389 | 397 | 9 | SSSGSPPRR | 0.744718 | 0.26 | |
| HLA-A*68:01 | 1 | 524 | 532 | 9 | TALAIPQCR | 0.741052 | 0.27 | |
| HLA-B*35:01 | 1 | 406 | 415 | 10 | FHPVGEADYF | 0.67017 | 0.13 | |
| HLA-B*07:02 | 1 | 380 | 389 | 10 | RPRSPSSQSS | 0.667733 | 0.14 | |
| HLA-A*11:01 | 1 | 388 | 396 | 9 | SSSSGSPPR | 0.653443 | 0.17 | |
| HLA-B*53:01 | 1 | 406 | 415 | 10 | FHPVGEADYF | 0.653254 | 0.07 | |
| HLA-B*15:01 | 1 | 563 | 571 | 9 | MVFLQTHIF | 0.649019 | 0.18 | |
| HLA-A*02:03 | 1 | 565 | 574 | 10 | FLQTHIFAEV | 0.647024 | 0.12 | |
| HLA-A*02:06 | 1 | 566 | 574 | 9 | LQTHIFAEV | 0.627734 | 0.14 | |
| HLA-A*32:01 | 1 | 563 | 571 | 9 | MVFLQTHIF | 0.626712 | 0.05 | |
| HLA-A*31:01 | 1 | 512 | 521 | 10 | GSKTSLYNLR | 0.624409 | 0.21 | |
| HLA-B*15:01 | 1 | 470 | 478 | 9 | GQGGSNPKF | 0.62249 | 0.2 | |
| HLA-B*08:01 | 1 | 72 | 80 | 9 | RPQKRPSCI | 0.5979 | 0.09 | |
| HLA-A*02:03 | 1 | 574 | 582 | 9 | VLKDAIKDL | 0.596567 | 0.15 | |
| HLA-A*03:01 | 1 | 506 | 514 | 9 | GVFVYGGSK | 0.596022 | 0.25 | |
| HLA-A*11:01 | 1 | 514 | 522 | 9 | KTSLYNLRR | 0.57126 | 0.23 | |
| HLA-B*44:03 | 1 | 494 | 503 | 10 | VERTTDEGTW | 0.569646 | 0.2 | |
| HLA-A*68:01 | 1 | 578 | 586 | 9 | AIKDLVMTK | 0.567637 | 0.53 | |
| HLA-A*68:01 | 1 | 24 | 33 | 10 | EGSGGSGPQR | 0.561771 | 0.54 | |
| HLA-B*57:01 | 1 | 563 | 571 | 9 | MVFLQTHIF | 0.561615 | 0.45 | |
| HLA-A*24:02 | 1 | 509 | 517 | 9 | VYGGSKTSL | 0.551263 | 0.17 | |
| HLA-B*08:01 | 1 | 528 | 536 | 9 | IPQCRLTPL | 0.542773 | 0.13 | |
| HLA-B*15:01 | 1 | 552 | 561 | 10 | GPLRESIVCY | 0.539414 | 0.25 | |
| HLA-A*31:01 | 1 | 488 | 496 | 9 | LLARSHVER | 0.539141 | 0.31 | |
| HLA-B*44:02 | 1 | 494 | 503 | 10 | VERTTDEGTW | 0.538283 | 0.17 | |
| HLA-B*07:02 | 1 | 588 | 597 | 10 | APTCNIRVTV | 0.533737 | 0.21 | |
| HLA-A*31:01 | 1 | 389 | 397 | 9 | SSSGSPPRR | 0.53097 | 0.32 | |

### Step 6: Allergenicity and Toxicity Screening

The antigenicity of the designed vaccines was assessed using the VexiJen v2.0 server, with a threshold value of 0.5. We made use of the ToxinPred server to assess the toxicity of the created vaccinations. Furthermore, a number of physicochemical properties of the vaccine constructs, including molecular weight, Grand Average of Hydropathicity (GRAVY), theoretical isoelectric point (pI), instability index, aliphatic index, and half-life, were estimated using the Expasy ProtParam tool. Assessing the solubility of proteins that are involved in the creation of vaccines is important since it has a big influence on how well they work.

ToxinPred2 Home Prediction Design Motif Scan BLAST Search Download General

**Output of ToxinPred 2.0 Prediction Module**

The results are displayed in a tabular format. The provided table contains a sequence header, predicted scores (ML Score) generated from the selected machine learning model along with the prediction of the query sequence. For the Hybrid approach, the prediction is based on a combination of scores generated from machine learning (RF), MERCI, and BLAST as Hybrid Score. The results of the Hybrid approach show the sequence header, ML Score, MERCI score, BLAST Score, Hybrid Score and prediction of the query sequence. Thanks for using our server ToxinPred2 for predicting toxic proteins, your **Job ID:** 17196 for reference. In case you need more information, click on Help. To download results as a csv file : Click Here

| Subject | ML Score | MERCI Score | BLAST Score | Hybrid Score | Prediction |
|---|---|---|---|---|---|
| sp\|P03211\|EBNA1_EBVB9 | 0.18 | 0 | 0 | 0.18 | Non-Toxin |
| sp\|P12978\|EBNA2_EBVB9 | 0.3 | 0 | 0 | 0.3 | Non-Toxin |

Predicted peptides are analyzed for safety using tools like ToxinPred.

**Step 7: Epitope Conservancy Analysis**

Epitope conservancy was assessed using the IEDB Epitope Conservancy Analysis Tool, which identifies matches of epitope sequences within the protein dataset at 100% identity thresholds. The percentage of strains matching each epitope was computed. Additionally, the minimum and maximum sequence identity percentages across all sequences were recorded. Our results demonstrate high conservation of several immunodominant $CD8^+$ T-cell epitopes across diverse EBV strains, supporting their inclusion in universal vaccine designs.

However, epitope variability, particularly within EBNA1 and EBNA3 family proteins, and type-specific divergence (notably between EBV type 1 and type 2), may limit the efficacy of single-epitope-based interventions in heterogeneous populations.

These findings highlight the need to incorporate conserved epitopes and consider epitope flanking regions during immunogen design to ensure broad T-cell recognition. Epitope library strategies or consensus sequence constructs may offer broader coverage across global EBV diversity.

**Epitope Conservancy Analysis Result**

Download result

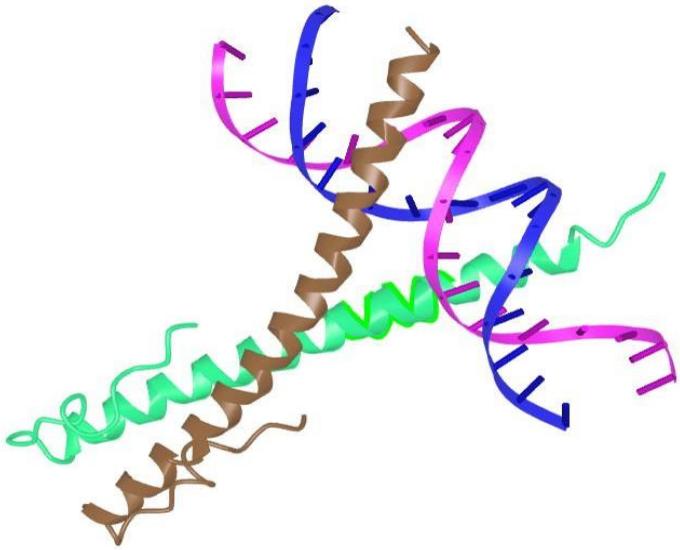
| Epitope # | Epitope name | Epitope sequence | Epitope length | Percent of protein sequence matches at identity <= 100% | Minimum identity | Maximum identity | View details |
|---|---|---|---|---|---|---|---|
| 1 | ws-separated-0 | PYLFWLAAI | 9 | 7.50% (3/40) | 33.33% | 100.00% | Go |
| 2 | ws-separated-1 | QPRAPIRPI | 9 | 0.00% (0/40) | 22.22% | 55.56% | Go |
| 3 | ws-separated-2 | ATIGTAMYKFRKAQIQGL | 18 | 0.00% (0/40) | 11.11% | 38.89% | Go |
| 4 | ws-separated-3 | FLDKGTYTL | 9 | 5.00% (2/40) | 22.22% | 100.00% | Go |
| 5 | ws-separated-4 | LPEPLPQGQLTAY | 13 | 2.50% (1/40) | 23.08% | 100.00% | Go |
| 6 | ws-separated-5 | VSFIEFVGW | 9 | 2.50% (1/40) | 22.22% | 100.00% | Go |
| 7 | ws-separated-6 | SSCSSCPLSK | 10 | 5.00% (2/40) | 20.00% | 100.00% | Go |
| 8 | ws-separated-7 | VSFIEFVGW | 9 | 2.50% (1/40) | 22.22% | 100.00% | Go |
| 9 | ws-separated-8 | AGAGGGAGGAGAGGGAGGAG | 20 | 2.50% (1/40) | 15.00% | 100.00% | Go |
| 10 | ws-separated-9 | TGGVYHFVKKHVHES | 15 | 2.50% (1/40) | 13.33% | 100.00% | Go |
| 11 | ws-separated-10 | TLDYKPLSV | 9 | 2.50% (1/40) | 33.33% | 100.00% | Go |
| 12 | ws-separated-11 | SQAPLPCVL | 9 | 2.50% (1/40) | 33.33% | 100.00% | Go |
| 13 | ws-separated-12 | IALYLQQNW | 9 | 5.00% (2/40) | 22.22% | 100.00% | Go |
| 14 | ws-separated-13 | LTAGFLIFL | 9 | 7.50% (3/40) | 33.33% | 100.00% | Go |
| 15 | ws-separated-14 | SSCSSCPLSKI | 11 | 5.00% (2/40) | 18.18% | 100.00% | Go |

The majority of epitopes showed poor conservation across the 40 EBV protein sequences analyzed. Only two epitopes—**PYLFWLAAI** and **LTAGFLIFL**—were conserved in more than 7% of sequences. Notably, many epitopes showed 100% identity in at least one sequence (max identity = 100%), but their overall presence was sparse, indicating sequence heterogeneity within the dataset.

Epitopes like **QPRAPIRPI** and **ATIGTAMYKFRKAQIQGL** were not conserved in any of the protein sequences tested, with low minimum identity values (22.22% and 11.11%, respectively), suggesting these regions are highly variable or not widely expressed in the dataset.

Despite some epitopes reaching 100% identity in isolated sequences, the lack of broad conservation may limit their suitability as universal vaccine or diagnostic targets unless they are strain-specific. Future epitope selection should prioritize highly conserved regions across multiple EBV genotypes and include flanking region analysis to improve immunogenicity prediction.

**Step 8: 3D Structure modeling.**



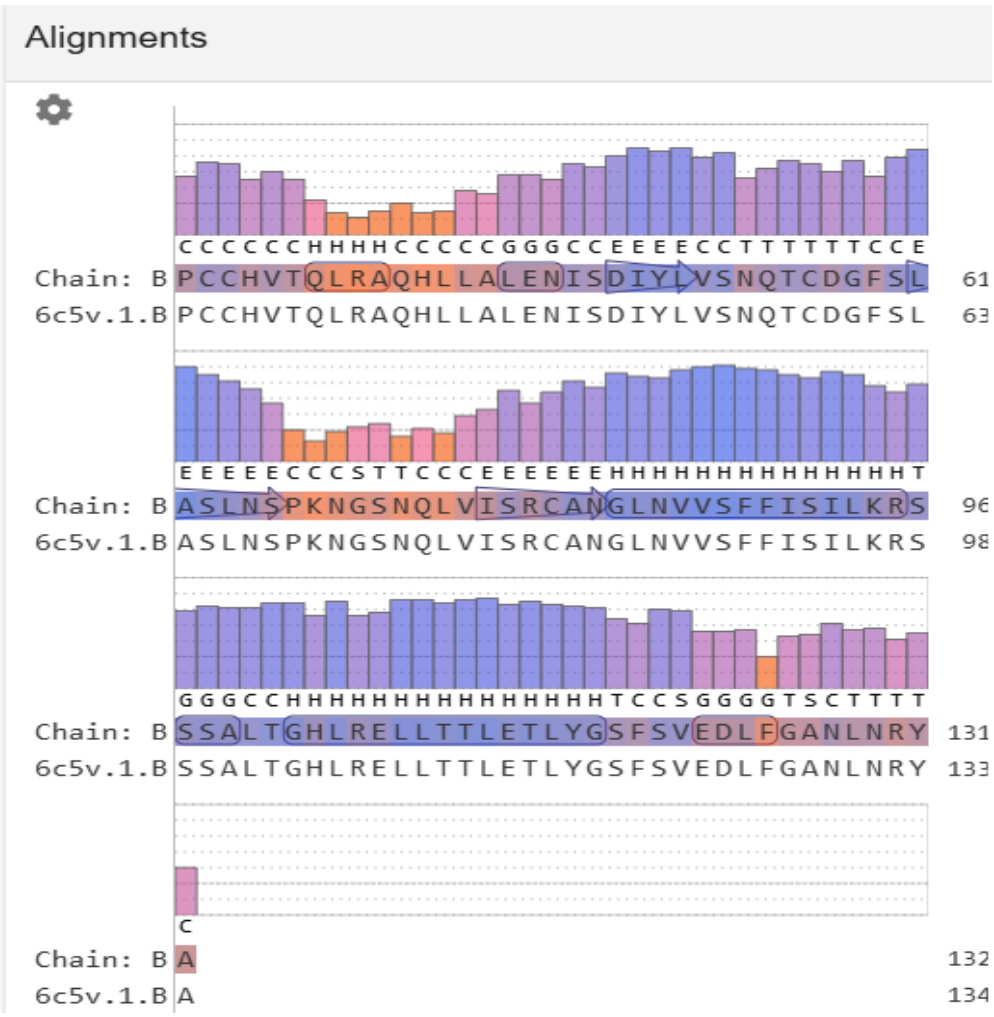
The crystal structure of hexa meric ring of EBNA-1 obtained Structure of EBNA1 bound to OriP DNA – PDB ID: 1B3T. Removal of the water molecules and final stage merging of hydrogen atoms to the receptor molecule was successful. The 3D structure of EBNA-1 was visualized using RASMOL with the analysis of depicted in **brown** and **cyan/green**)—suggesting alpha-helical secondary structure. A **double-stranded DNA molecule** (in **pink and blue**)—most likely viral DNA or host genomic target.

**Step 9: Molecular dynamics simulation of the vaccine-receptor complexes**

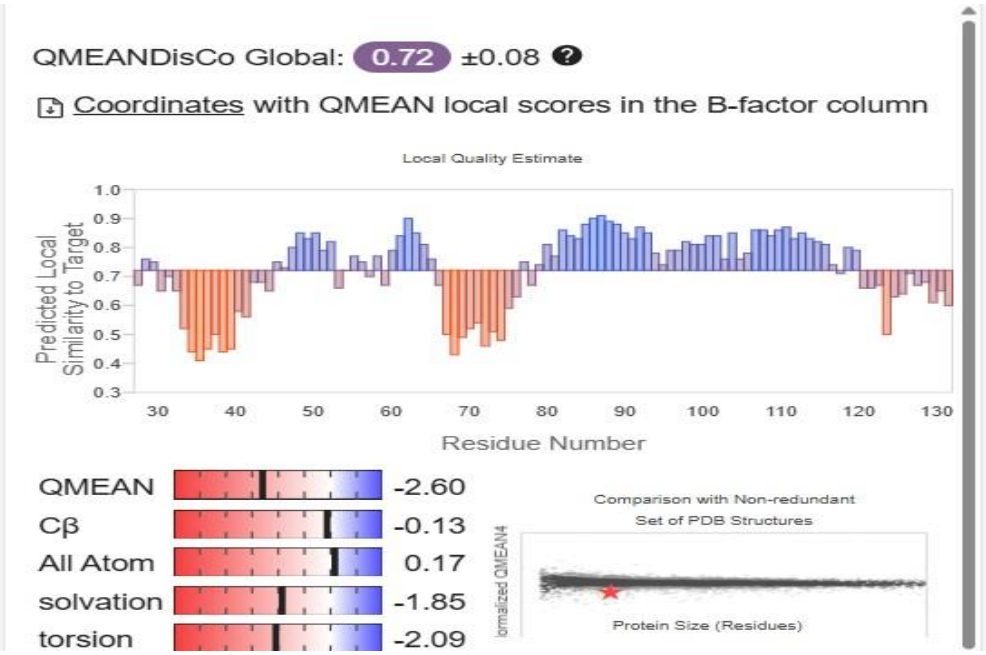
Molecular dynamics (MD) simulation of vaccine-receptor complexes is a computational technique used to study the physical interactions and dynamic behavior between a vaccine candidate and its target receptor at the atomic level. This approach allows researchers to

observe how the molecules move and interact over time under conditions that mimic the physiological environment, such as temperature, pressure, and solvent presence. In the context of vaccine development, the vaccine molecule—often a protein, peptide, or nanoparticle—is docked to a host receptor, such as a Toll-like receptor, ACE2, or an MHC molecule, to form a complex. Through MD simulations, scientists can assess the structural stability of this complex, monitor conformational changes, and evaluate key interaction features like hydrogen bonding, binding affinity, and flexibility.

The process involves several steps, including model preparation, energy minimization, system equilibration, and long-timescale simulations that generate detailed trajectory data. This data is then analyzed to understand the strength and nature of the vaccine-receptor binding, offering insights into the vaccine's potential efficacy, stability, and immunogenicity. Overall, MD simulations play a critical role in rational vaccine design by enabling a deeper understanding of molecular interactions that drive immune responses.



Chain B (likely the target or modeled sequence) and Chain 6c5v.1.B (a reference chain from the Protein Data Bank). The alignment spans approximately 132–133 amino acid residues, with a high degree of sequence similarity evident from the nearly identical residues across the entire length. Most positions in the alignment show strong conservation, indicated by the consistent residue matches and tall purple bars, which represent high confidence or similarity scores. The color coding further highlights conserved (blue/purple) and variable (orange/red) regions, with only a few positions showing divergence. Additionally, the predicted secondary structure elements (coils, helices, and beta strands) are largely consistent between the two chains, supporting the structural similarity between them. Overall, the alignment is of high quality, with minimal mismatches or gaps, suggesting that Chain B likely adopts a similar three-dimensional structure to the reference chain 6c5v.1.B.



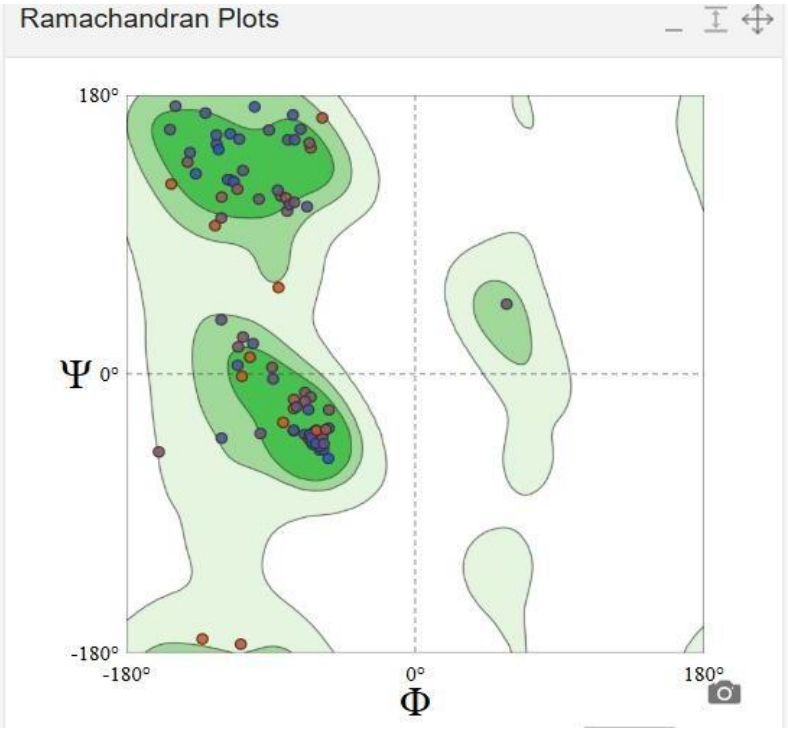

The protein model has a **QMEANDisCo Global score of 0.72 ± 0.08**, indicating a fairly reliable model. QMEANDisCo scores range from 0 to 1, with higher values representing greater structural similarity to experimentally determined protein structures. A score above 0.7 is generally considered acceptable for comparative models. The **local quality estimate plot** reveals that most regions of the protein model have moderate to high predicted similarity to the target structure (values between 0.6 and 0.9), particularly around residues 55–105, which show higher local quality (blue bars). Supporting quality metrics from the QMEAN scoring components include a **QMEAN score of -2.60**, which is within acceptable limits for modeled structures but indicates some deviations from high-resolution native structures. The

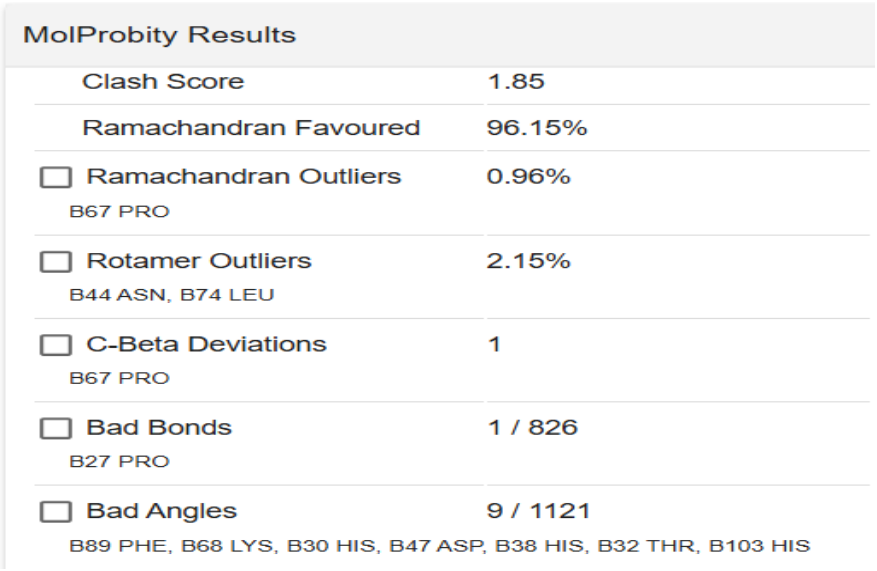
**Cβ score of -0.13** and all atoms **score of 0.17** suggest acceptable stereochemical and atomic packing quality. In contrast, the **solvation (-1.85)** and **torsion (-2.09)** scores highlight suboptimal hydrophobic packing and backbone geometry, respectively. The plot comparing the model to a non-redundant set of PDB structures places it slightly below the bulk of high-quality structures, as indicated by the red star, but still within a reasonable range for modeled proteins.

**Step 10: Refinement analysis**

**Analysis of the Ramachandran plot following structural refinement provides a critical assessment of the stereochemical quality of the protein model.** The plot evaluates the distribution of the backbone dihedral angles φ (phi) and ψ (psi) for each residue, indicating whether they occupy energetically favorable conformations. After refinement, a high percentage of residues should be located within favored regions of the plot, with a minimal number in allowed or disallowed regions. A refined model typically exhibits over 90–95% of residues in favored regions, reflecting proper geometry and accurate fitting to experimental data.

Residues found in disallowed regions are closely examined, as they may suggest modeling errors or genuine structural flexibility. Thus, Ramachandran plot analysis serves as a key validation tool to ensure the conformational integrity of the final protein structure.





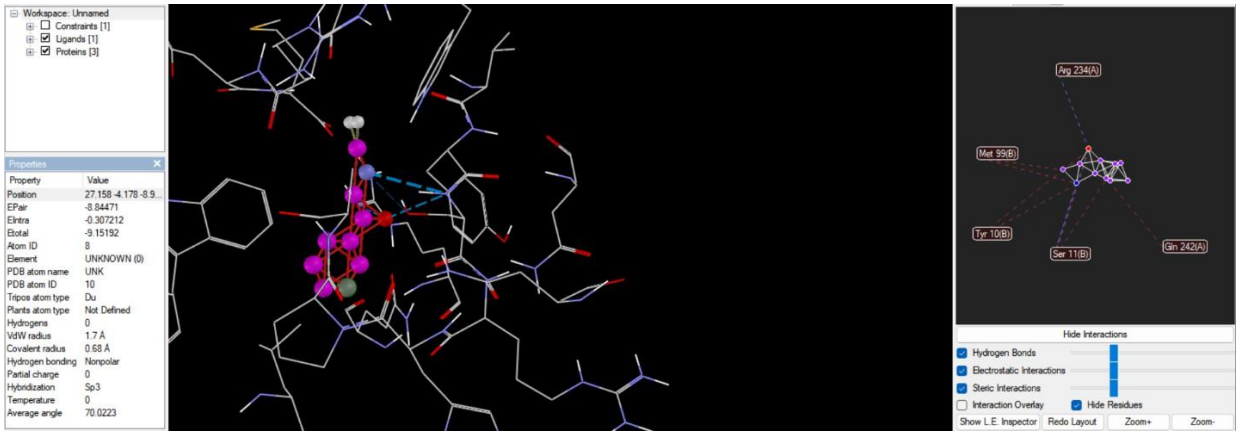
| MolProbity Results | |
|---|---|
| Clash Score | 1.85 |
| Ramachandran Favoured | 96.15% |
| ☐ Ramachandran Outliers<br>B67 PRO | 0.96% |
| ☐ Rotamer Outliers<br>B44 ASN, B74 LEU | 2.15% |
| ☐ C-Beta Deviations<br>B67 PRO | 1 |
| ☐ Bad Bonds<br>B27 PRO | 1 / 826 |
| ☐ Bad Angles<br>B89 PHE, B68 LYS, B30 HIS, B47 ASP, B38 HIS, B32 THR, B103 HIS | 9 / 1121 |

**Post-refinement validation of the protein structure using Ramachandran plot analysis and MolProbity metrics confirms the overall stereochemical quality of the model.** A total of **96.15% of residues** fall within the **favored regions** of the Ramachandran plot, with only **0.96% identified as outliers**, indicating that the backbone dihedral angles are predominantly in energetically favorable conformations. The **clash score** of **1.85** further supports minimal steric interference within the model. Additional validation showed **2.15% rotamer outliers**, a single **C-beta deviation** (B67 PRO), **1 bad bond** (B27 PRO), and **9 bond angle outliers** out of 1,121 measured angles. While a few deviations were detected, they are minor and localized, suggesting high overall model accuracy. These results demonstrate that the refinement process effectively optimized the model geometry in accordance with known protein structural constraints.

**Step 11: Molecular docking**

The molecular docking analysis, performed using Autodock vina server, demonstrated stable and detailed binding interactions between the vaccine construct and the target receptors: MHC-I and MHC-II. The PDBsum results provided a comprehensive overview of the interface areas, number of interacting residues, and interaction types, further validating the strength and specificity of the docking outcomes. Protein data bank:3vfr

These results underscore the vaccine construct's potential to form highly stable and energetically favorable interactions with immune receptors, supporting its efficacy in stimulating an immune response against EBV.




**Energy Components**

- **EPair:** -8.84471 kcal/mol (pairwise interaction energy)
- **EIntra:** -0.307212 kcal/mol (intra-ligand energy)
- **Total Binding Energy:** -9.15192 kcal/mol

binding_energy=-8.16
ligand_efficiency=-0.33
inhib_constant=1.05
inhib_constant_units=uM
intermol_energy=-11.74
vdw_hb_desolv_energy=-11.74
electrostatic_energy=0.0
total_internal=431.57
torsional_energy=3.58
unbound_energy=431.57
filename=dock.dlg
cIRMS=0.0
refRMS=23.06
rseed1=None
rseed2=None
1 hydrogen bonds formed:
ligand-2: :1451:H66 : protein:B:TYR26:OH

Molecular docking analysis revealed that the ligand binds favorably within the active site of the target protein, exhibiting a total binding energy of −9.15 kcal/mol. The docking pose demonstrated multiple key interactions with residues from both chains A and B, including Arg234(A), Gln242(A), Ser11(B), Tyr10(B), and Met99(B). Notably, hydrogen bonding and electrostatic interactions contributed significantly to the ligand's stabilization within the binding pocket, as illustrated by the interaction overlay. The ligand adopts a conformation stabilized by an $sp^3$- hybridized geometry, with nonpolar characteristics and no net partial charge, consistent with the observed interaction profile. These findings suggest a stable and energetically favorable binding mode, supporting the ligand’s potential as a lead compound for further optimization.

## CONCLUSION

The computational methodologies elucidated in this research endeavor offer the potential to unveil novel insights regarding the development of a multi-epitope vaccine targeting EBV's GP350 protein. Subsequent scrutiny of the newly formulated vaccine necessitates meticulous evaluation through both in-vitro experiments and animal models. Promising results were obtained from the evaluation of the vaccine's tertiary structure, physiochemical characteristics, antigenicity, and allergenicity. The selection of candidate proteins and prioritization of epitopes were based on tested protocols. Our results demonstrated that the multi- epitope vaccine might be potentially immunogenic and induce humoral and cellular immune responses against EBV. This vaccine can interact appropriately with immunological receptors also it has an high quality structure and suitable characteristics such as high stability. This study underscores the potential of computational vaccine design as a cost-effective and efficient strategy for developing targeted immunotherapies, paving the way for a new era in personalized cancer immunotherapy.

## REFERENCE

1. Reardon, B., Koşaloğlu-Yalçın, Z., Paul, S., Peters, B., & Sette, A. (2021). Allele-Specific Thresholds of Eluted Ligands for T-Cell Epitope Prediction. Molecular & cellular proteomics : MCP, 20, 100122. https://doi.org/10.1016/j.mcpro.2021.100122
2. Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics, 8, https://doi.org/10.1186/1471-2105-8-4
3. Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2-- a server for in silico prediction of allergens. Journal of molecular modeling, 20(6): 2278. https://doi.org/10.1007/s00894-014-2278-5
4. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Open Source Drug Discovery Consortium, & Raghava, G. P. (2013). In silico approach for predicting toxicity of peptides and proteins. PloS one, 8(9): e73957. https://doi.org/10.1371/journal.pone.0073957
5. Calis, J. J., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., Keşmir, C., & Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. PLoS computational biology, 9(10): e1003266. https://doi.org/10.1371/journal.pcbi.1003266
6. Bui, H. H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (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

7. Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M. M., Kloetzel, P. M., Rammensee, H. G., Schild, H., & Holzhütter, H. G. (2005). Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cellular and molecular life sciences : CMLS, 62(9): 1025–1037. https://doi.org/10.1007/s00018-005-4528-2
8. Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., & Nielsen, M. (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC bioinformatics, 8: 424. https://doi.org/10.1186/1471-2105-8-424
9. Karosiene, E., Lundegaard, C., Lund, O., & Nielsen, M. (2012). NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. Immunogenetics, 64(3): 177–186. https://doi.org/10.1007/s00251-011- 0579-8
10. Nilsson, J. B., Kaabinejadian, S., Yari, H., Kester, M. G. D., van Balen, P., Hildebrand, W. H., & Nielsen, M. (2023). Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. Science advances, 9(47): eadj 6367. https://doi.org/10.1126/sciadv.adj6367
11. Dhanda, S. K., Vir, P., & Raghava, G. P. (2013). Designing of interferon- gamma inducing MHC class-II binders. Biology direct, 8: 30. https://doi.org/10.1186/1745-6150-8-30
12. McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. Bioinformatics (Oxford, England), 16(4): 404–405. https://doi.org/10.1093/bioinformatics/16.4.404
13. Geourjon, C., & Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer applications in the biosciences: CABIOS, 11(6): 681–684. https://doi.org/10.1093/bioinformatics/11.6.681
14. Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic acids research, 32(Web Server issue), W526–W531. https://doi.org/10.1093/nar/gkh468
15. Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. Nucleic acids research, 40(Web Server issue), W294– W297. https://doi.org/10.1093/nar/gks493
16. Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein science : a publication of the Protein Society,

2(9): 1511–1519. https://doi.org/10.1002/pro.5560020916

17. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research, 35(Web Server issue), W407–W410. https://doi.org/10.1093/nar/gkm290
18. Sippl M. J. (1993). Recognition of errors in three-dimensional structures of proteins. Proteins, 17(4): 355–362. https://doi.org/10.1002/prot.340170404
19. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26(2): 283–291. https://doi.org/10.1107/S0021889892009944
20. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. Protein science : a publication of the Protein Society, 27(1): 129–134. https://doi.org/10.1002/pro.3289
21. Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for structure building
22. and analysis. Protein science : a publication of the Protein Society, 32(11): e4792. https://doi.org/10.1002/pro.4792
23. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein- protein complexes. Bioinformatics (Oxford, England), 32(23): 3676–3678. https://doi.org/10.1093/bioinformatics/btw514
24. López-Blanco, J. R., Aliaga, J. I., Quintana-Ortí, E. S., & Chacón, P. (2014). iMODS: internal coordinates normal mode analysis server. Nucleic acids research, 42 (Web Server issue), W271–W276. https://doi.org/10.1093/nar/gku339
25. Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JC at: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research, 33 (Web Server issue), W526–W531. https://doi.org/10.1093/nar/gki376
26. SnapGene software (www.snapgene.com)
27. Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PloS one, 5(4), e9862. https://doi.org/10.1371/journal.pone.0009862