International Journal of Botany Studies
www.botanyjournals.com
ISSN: 2455-541X
Received: 07-06-2026, Accepted: 07-07-2026, Published: 22-07-2026
Volume 11, Issue 3, 2026, Page No. 140-149

An Immunoinformatics approach to design a neoantigen-driven Multi-Epitope Vaccine against Leukemia

Hemang Bhindi[1], Dharmendrasinh F Rao[1], Arav Sharma[2], Saumya K Patel[1]
[1] Department of Botany, Bioinformatics and Climate Change Impacts Management, School of Sciences, Gujarat University, Ahmedabad, Gujarat, India
[2] Sardar Patel Vidyalaya, New Delhi, Delhi, India
DOI: https://doi.org/10.66856/ijbs.2026.11.3.11187

Abstract
Leukemia remains a major hematological malignancy with high relapse rates and limited therapeutic options, highlighting the need for effective immunotherapeutic strategies. This study employed an integrated immunoinformatics and structural vaccinology approach to design a multi-epitope therapeutic vaccine targeting leukemia-associated neoantigens from acute myeloid leukemia (AML), chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL), and chronic lymphocytic leukemia (CLL). A total of 81 neoantigens retrieved from the Cancer Epitope Database and Analysis Resource (CEDAR) were screened for antigenicity, allergenicity, toxicity, and immunogenicity, resulting in the selection of 7 CTL and 9 HTL epitopes for vaccine construction. The vaccine incorporated β-defensin-3, the PADRE peptide, suitable linkers, and a His-tag. Physicochemical analysis indicated favorable stability and solubility, while structural validation showed 95.1% of residues in the most favored regions of the Ramachandran plot. Population coverage analysis predicted 99.95% global HLA coverage. Molecular docking and a 200 ns molecular dynamics simulation confirmed stable binding to TLR2, and MM-GBSA analysis demonstrated an increase in binding free energy from −55.54 kcal/mol to −248.13 ± 37.95 kcal/mol following the simulation. Codon optimization yielded a CAI of 1.0 and a GC content of 53.57%, indicating efficient expression in *Escherichia coli*. Furthermore, *in silico* immune simulation predicted rapid antigen clearance, strong antibody production, and robust IFN-γ-mediated cellular immunity. These findings suggest that the designed multi-epitope vaccine is a promising therapeutic candidate against leukemia and warrants further *in vitro* and *in vivo* validation.

Keywords: Leukemia, Neoantigen vaccine, Multi-epitope vaccine, Immunoinformatics, Structural vaccinology

Introduction

Leukemia is a diverse group of hematological malignancies characterized by the uncontrolled proliferation of abnormal leukocytes originating from the bone marrow. It comprises four major clinical subtypes—acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), chronic myeloid leukemia (CML), and chronic lymphocytic leukemia (CLL)—each differing in cellular origin, disease progression, and therapeutic response [1, 2]. Despite remarkable advances in chemotherapy, hematopoietic stem cell transplantation, targeted tyrosine kinase inhibitors, monoclonal antibodies, and chimeric antigen receptor (CAR)-T cell therapies, long-term disease management remains challenging due to relapse, drug resistance, treatment-associated toxicities, and the genetic heterogeneity of leukemic cells. Furthermore, many current therapeutic approaches are associated with high costs, limited accessibility, and immune escape mechanisms that reduce their long-term efficacy. These limitations highlight the need for innovative immunotherapeutic strategies capable of inducing durable and highly specific anti-leukemic immune responses [3, 4].

Cancer vaccines have emerged as a promising immunotherapeutic approach because they stimulate the host immune system to recognize and eliminate malignant cells while minimizing damage to normal tissues. In particular, therapeutic vaccines based on tumor-specific neoantigens offer significant advantages over conventional tumor-associated antigens, as neoantigens originate from somatic mutations that are absent in healthy cells, thereby reducing the likelihood of immune tolerance and off-target effects [3]. Advances in high-throughput sequencing, cancer epitope databases, and computational immunology have accelerated the identification of immunogenic neoantigens suitable for vaccine development. Multi-epitope vaccines further enhance this strategy by incorporating multiple cytotoxic T-lymphocyte (CTL), helper T-lymphocyte (HTL), and B-cell epitopes into a single construct, enabling the simultaneous activation of cellular and humoral immune responses while improving population coverage across diverse human leukocyte antigen (HLA) genotypes [5, 6].

Recent developments in immunoinformatics have transformed vaccine design from a largely experimental process into a rational, data-driven workflow. Computational tools can rapidly screen thousands of candidate antigens for antigenicity, allergenicity, toxicity, cytokine induction potential, and HLA-binding affinity, substantially reducing the time and cost associated with vaccine discovery. These predictions can be complemented by structural bioinformatics approaches, including peptide modeling, molecular docking, molecular dynamics simulations, and binding free-energy calculations, to evaluate the stability and biological plausibility of vaccine–receptor interactions before experimental validation. In addition, *in silico* immune simulations and codon optimization provide valuable insights into the immunogenic potential and expression feasibility of vaccine constructs, allowing promising candidates to be prioritized for laboratory studies [7–9].

In the present study, an integrated immunoinformatics and structural vaccinology pipeline was employed to design a therapeutic multi-epitope vaccine targeting leukemia-associated neoantigens identified across AML, ALL, CML, and CLL. Candidate neoantigens were systematically screened to identify safe and highly immunogenic epitopes, followed by the prediction of CTL, HTL, and B-cell epitopes for incorporation into a chimeric vaccine construct. The designed vaccine was comprehensively evaluated through physicochemical characterization, three-dimensional structure prediction, molecular docking with HLA molecules and Toll-like receptor 2 (TLR2), molecular dynamics simulations, MM-GBSA binding free-energy analysis, immune response simulation, and codon optimization for recombinant expression. This comprehensive computational framework provides a strong foundation for the development of a broadly immunogenic therapeutic vaccine against leukemia and offers a promising candidate for future experimental and clinical investigation.

Materials and Methods

1. Retrieval and Screening of Leukemia Neoantigens

Leukemia-associated neoantigens were retrieved from the Cancer Epitope Database and Analysis Resource (CEDAR), which provides experimentally validated and computationally curated cancer epitopes [10]. Neoantigens reported for acute myeloid leukemia (AML), chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL), and chronic lymphocytic leukemia (CLL) were collected for further analysis. Candidate neoantigens were initially filtered to remove redundant peptide sequences before subsequent immunological screening.

The antigenicity of each peptide was evaluated using the VaxiJen v2.0 server with the tumor model and a threshold value of 0.5 [11]. Peptides predicted to be probable antigens were retained. The allergenic potential of the selected peptides was assessed with AllerTOP v2.0, and peptide toxicity was evaluated with ToxinPred [12, 13]. In addition, the IFNepitope server was used to predict interferon-γ-inducing peptides, as IFN-γ plays a critical role in antitumor immunity [14]. Only peptides predicted to be antigenic, non-allergenic, non-toxic, and capable of inducing favorable immune responses were selected for epitope prediction.

2. Prediction of Cytotoxic T-Lymphocyte (CTL) Epitopes

Cytotoxic T-lymphocyte epitopes were identified using the Immune Epitope Database (IEDB) MHC-I binding prediction tool [15]. Candidate peptides were analyzed against common human leukocyte antigen (HLA) class I alleles using the recommended prediction method. Peptides exhibiting high binding affinity to multiple HLA class I molecules were selected. Predicted CTL epitopes were further evaluated for antigenicity, immunogenicity, toxicity, and allergenicity to ensure their suitability for vaccine development. The corresponding HLA alleles with the strongest predicted binding affinity were selected for molecular docking studies.

3. Prediction of Helper T-Lymphocyte (HTL) Epitopes

Helper T-lymphocyte epitopes were predicted using the IEDB MHC-II binding prediction tool against a panel of frequently occurring HLA-II alleles. Candidate peptides demonstrating strong binding affinity were selected based on percentile rank and predicted binding scores [16]. Since helper T cells play a central role in coordinating adaptive immune responses, selected HTL epitopes were further screened for antigenicity, toxicity, allergenicity, and potential to induce cytokines. High-confidence HTL epitopes were subsequently included in the vaccine construct.

4. Population Coverage Analysis

The global population coverage of the selected CTL and HTL epitopes was estimated using the IEDB Population Coverage tool. The analysis examined the frequency distributions of HLA class I and class II alleles across different ethnic and geographic populations [17]. Combined population coverage was calculated to estimate the proportion of individuals expected to mount an immune response against the designed vaccine.

5. Construction of the Multi-Epitope Vaccine

The final vaccine construct was designed by combining the selected CTL and HTL epitopes using appropriate peptide linkers to preserve independent epitope processing and presentation. An immunostimulatory adjuvant was incorporated at the N-terminal region of the construct to enhance innate immune activation through Toll-like receptor 2 (TLR2). Flexible and rigid linkers were introduced between different functional components to improve structural stability and reduce steric interference. The complete amino acid sequence was subsequently subjected to physicochemical and structural analyses [18, 19].

6. Physicochemical Characterization and Structure Prediction

The ProtParam server was used to determine the physicochemical characteristics of the vaccine construct, including molecular weight, theoretical isoelectric point, instability index, aliphatic index, estimated half-life, and grand average of hydropathicity (GRAVY) [20]. The three-dimensional structure of the vaccine was predicted using trRosetta. Structural quality was assessed using the Ramachandran plot analysis. To further improve structural stability, disulfide engineering was performed using the Disulfide by Design 2 server [21].

7. Molecular Docking with Toll-Like Receptor 2

The interaction between the vaccine construct and human Toll-like receptor 2 (TLR2) was investigated using the HADDOCK 2.4 docking platform. The crystal structure of TLR2 was obtained from the Protein Data Bank and prepared by removing unnecessary molecules before docking. Docking results were ranked according to HADDOCK scoring functions, and the highest-ranked complex was selected for subsequent molecular dynamics simulation [22]. Protein–protein interaction networks were analyzed using PDBsum [23].

8. Molecular Dynamics Simulation and Binding Free Energy Analysis

The structural stability of the vaccine–TLR2 complex was evaluated by molecular dynamics (MD) simulations using the Desmond simulation package implemented in the Schrödinger Suite. The system was solvated with the TIP3P water model in an orthorhombic simulation box, neutralized with counterions, and simulated with the OPLS force field

under periodic boundary conditions. After energy minimization and equilibration, production simulations were performed under constant temperature and pressure. Structural stability was assessed using root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), hydrogen-bond analysis, and trajectory visualization [24, 25]. The binding affinity of the docked complex was further estimated using the Prime MM-GBSA module of the Schrödinger Suite, which calculates the binding free energy based on molecular mechanics energies combined with generalized Born and surface-area solvation terms.

9. Codon Optimization and *In Silico* Cloning

To facilitate recombinant expression in *Escherichia coli*, the vaccine amino acid sequence was reverse-translated and codon-optimized using the Java Codon Adaptation Tool (JCat). Codon Adaptation Index (CAI) and GC content were evaluated to estimate expression efficiency. The optimized nucleotide sequence was subsequently inserted into the pET-28a(+) expression vector using SnapGene software to simulate recombinant cloning [26, 27].

10. *In Silico* Immune Simulation

The immunogenic potential of the designed vaccine was evaluated using the C-ImmSim server. A prime-boost vaccination strategy consisting of three simulated injections at predefined intervals was employed to mimic routine immunization. The simulation predicted antigen clearance, antibody production, B-cell and T-cell responses, cytokine secretion, macrophage activation, and the formation of immunological memory, thereby providing an overall assessment of the vaccine's ability to induce protective immune responses [28].

Results

1. Identification and Screening of Leukemia Neoantigens

Neoantigens associated with four major leukemia subtypes, acute myeloid leukemia (AML), chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL), and chronic lymphocytic leukemia (CLL), were retrieved from the Cancer Epitope Database and Analysis Resource (CEDAR). The database identified a total of 81 candidate neoantigens, comprising 8 neoantigens for AML, 9 for CML, 3 for CLL, and 61 for ALL. The markedly higher number of predicted neoantigens in ALL suggests a broader repertoire of potential immunogenic targets for subsequent epitope screening. These candidate neoantigens served as the initial dataset for further filtering based on antigenicity, allergenicity, toxicity, and immunogenic potential to identify suitable candidates for multi-epitope vaccine design.

2. Prediction of CTL Epitopes

A total of seven CTL epitopes satisfying the predefined selection criteria, including antigenicity (VaxiJen score > 0.5), non-allergenicity, and non-toxicity, were incorporated into the final multi-epitope vaccine construct (Table 1). The selected epitopes originated from CML, CLL, and ALL neoantigens and demonstrated strong predicted binding affinity toward multiple HLA class I alleles, as indicated by their low elution percentile scores. Among the identified CTL epitopes, YLINKEEAL and TAHLNCNLF exhibited particularly high antigenicity scores of 0.8591 and 1.0161, respectively, while IRMKHCVMV showed the highest antigenicity score (1.0478) among all selected CTL epitopes. Most epitopes showed elution percentages below 1.0, suggesting efficient presentation by MHC class I molecules and a high probability of recognition by cytotoxic T lymphocytes. In addition, several epitopes were predicted to interact with multiple HLA alleles in the MHC restriction assay, indicating broad HLA promiscuity and the potential to enhance vaccine coverage across genetically diverse populations.

Table 1: Final CTL epitopes incorporated into the multi-epitope vaccine construct

Neoantigen	Sequence	HLA Allele(s)	Elution Percentile	Antigenicity Score	Allergenicity
CML	YLINKEEAL	HLA-A*02:01	0.03	0.8591	Non-allergen
CML	YLKALQRPV	HLA-A*02:03; HLA-A*02:01*	0.16	0.6293	Non-allergen
CLL	MPIEPGDIGC	HLA-B*35:01	1.20	0.5597	Non-allergen
ALL	IRMKHCVMV	HLA-C*07:02*	0.21	1.0478	Non-allergen
ALL	KLQFHCMPV	HLA-A*02:03; HLA-A*02:01*	0.71	0.5684	Non-allergen
ALL	LRNFIRMKHCVMVLV (Core: FIRMKHCVM)	HLA-B*08:01; HLA-A*24:02, HLA-A*02:07, HLA-B*58:01, HLA-B*40:01, HLA-C*03:02, HLA-C*07:02*	0.04	0.7295	Non-allergen
ALL	TAHLNCNLF	HLA-B*35:01	0.43	1.0161	Non-allergen

3. HTL Epitope Prediction

Similarly, nine HTL epitopes meeting the antigenicity, non-allergenicity, and non-toxicity criteria were selected for incorporation into the vaccine construct (Table 2). These epitopes included two peptides derived from CML and seven from ALL, each demonstrating favorable binding to common HLA class II alleles with acceptable elution percentile scores. The CML-derived epitope TVHSIPLTINKEEALQRPVASDFEP exhibited the highest antigenicity among the selected HTL epitopes (0.6850), whereas VQNILDEVVSFGERI showed the highest antigenicity among the ALL-derived peptides (0.9049). Notably, VYFFQNNQTTSIQEG displayed the lowest elution percentile (0.11), indicating a high likelihood of presentation by HLA-DRB3*02:02 molecules. The inclusion of multiple HTL epitopes that bind diverse HLA class II alleles is expected to promote robust $CD4^+$ T-cell activation, enhance cytokine secretion, provide effective helper functions for cytotoxic T-cell responses, and support long-term immunological memory, thereby strengthening the overall immunogenicity of the designed multi-epitope vaccine.

Table 2: Final HTL epitopes incorporated into the multi-epitope vaccine construct.

Neoantigen	Sequence	HLA Allele	Elution Percentile	Antigenicity Score	Allergenicity
CML	Ivhsatgfkqsskalqrpvasdfep	HLA-DRB1*11:01	0.22	0.5583	Non-allergen
CML	Tvhsipltinkeealqrpvasdfep	HLA-DRB5*01:01	1.50	0.6850	Non-allergen
ALL	Gqtrehallaymlgv	HLA-DPA1*02:01/DPB1*14:01	7.00	0.5614	Non-allergen
ALL	Ifklqfhcmpvsnev	HLA-DRB1*04:05	2.20	0.5338	Non-allergen
ALL	Stsspmekaislatd	HLA-DRB1*09:01	8.40	0.7346	Non-allergen
ALL	Vmvngdippwlkksa	HLA-DRB3*01:01	9.00	0.5022	Non-allergen
ALL	Vqnildevvsfgeri	HLA-DRB4*01:01	7.00	0.9049	Non-allergen
ALL	Vyffqnnqttsiqeg	HLA-DRB3*02:02	0.11	0.7077	Non-allergen
ALL	Gvslaskksrlfsss	HLA-DRB1*12:01	3.10	0.5602	Non-allergen

4. Population Coverage Analysis

Hypothetical population coverage of MHC class I (CTL) and MHC Class II (HTL) epitopes was predicted by the IEDB Population Coverage Analysis tool. A total of 16 epitopes that interact with 124 HLA alleles across MHC I and MHC II were combined and submitted for population coverage prediction. The study shows that our selected MHC class I (CTL) and MHC Class II (HTL) epitopes represent 99.95% global coverage of HLA alleles. Furthermore, the regions demonstrating the highest HLA allele population coverage amounted to 19, achieving an optimal score of 100% The regions includes United State, United State Amerindian, Sweden, Sweden Caucasoid, Spain, Spain Caucasoid, Russia, Russia Siberian, Papua New Guinea, Papua New Guinea Melanesian, North America, Mexico, Mexico Amerindian, France, France Caucasoid, Europe, Cook Islands, Cook Islands Polynesian, Brazil. The Asian Region, encompassing India, East Asia, China, Northeast Asia, and Southeast Asia, faces a considerable burden of leukemia. The calculated HLA allele coverage for the regions was 99.9%, 99.92%, 99.82%, 99.83%, and 99.56%, respectively (Fig 1).

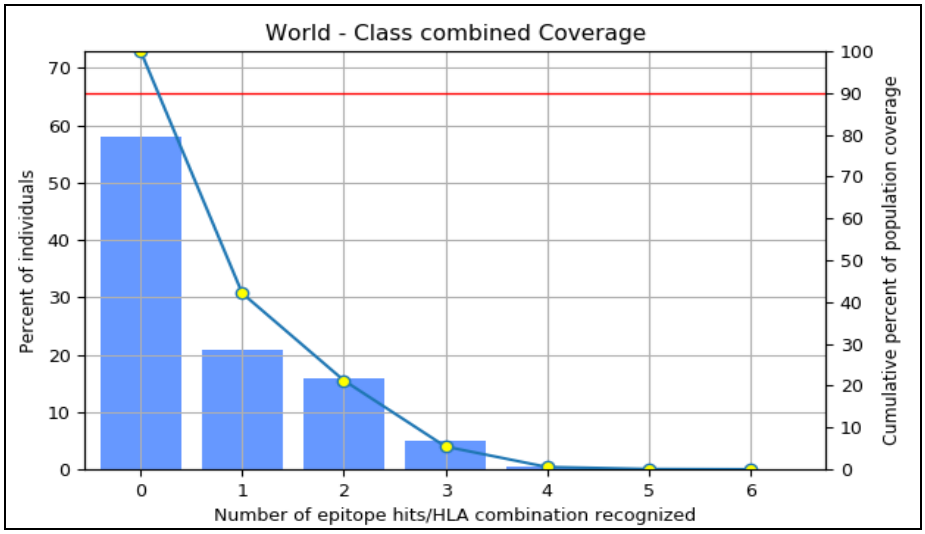

Fig 1: Combined population coverage of MHC Class I- and Class II-restricted epitopes.

5. Vaccine Construction

Following the sequential screening for antigenicity, allergenicity, toxicity, and HLA-binding affinity, the selected CTL and HTL epitopes were assembled into a multi-epitope vaccine construct. The construct comprised human β-defensin-3 as an N-terminal adjuvant to enhance innate immune activation, followed by the PADRE (Pan HLA-DR-binding epitope) sequence to promote universal $CD4^+$ T-cell responses. The adjuvant and PADRE peptide were connected using rigid EAAAK linkers to maintain structural independence, while individual CTL epitopes were joined by AAY linkers to facilitate efficient proteasomal processing and MHC class I presentation. Similarly, GPGPG linkers were used to separate HTL epitopes, preserving their independent immunogenicity and promoting optimal MHC class II presentation. A C-terminal 6×His tag was incorporated to facilitate recombinant protein purification. The final vaccine construct consisted of seven CTL epitopes and nine HTL epitopes arranged in a rational architecture to maximize immunogenic potential (Fig 2).

6. Physicochemical Properties Analysis of the Vaccine

Physicochemical characterization of the construct was performed using the ProtParam server, which predicted a molecular formula of $C_{1705}H_{2689}N_{485}O_{475}S_{22}$ with an estimated half-life of 30 h in mammalian reticulocytes (*in vitro*), 20 h in yeast (*in vivo*), and 10 h in *Escherichia coli* (*in vivo*), indicating acceptable stability across different biological systems. The instability index was 41.09, suggesting that the construct is marginally unstable, whereas the aliphatic index of 75.32 indicated good thermostability. Furthermore, the GRAVY value of −0.168 suggested a slightly hydrophilic nature, implying favorable aqueous solubility and the potential for efficient interaction with immune receptors.

Collectively, these physicochemical characteristics support the suitability of the designed construct as a promising recombinant multi-epitope vaccine candidate for subsequent structural and immunological evaluation.

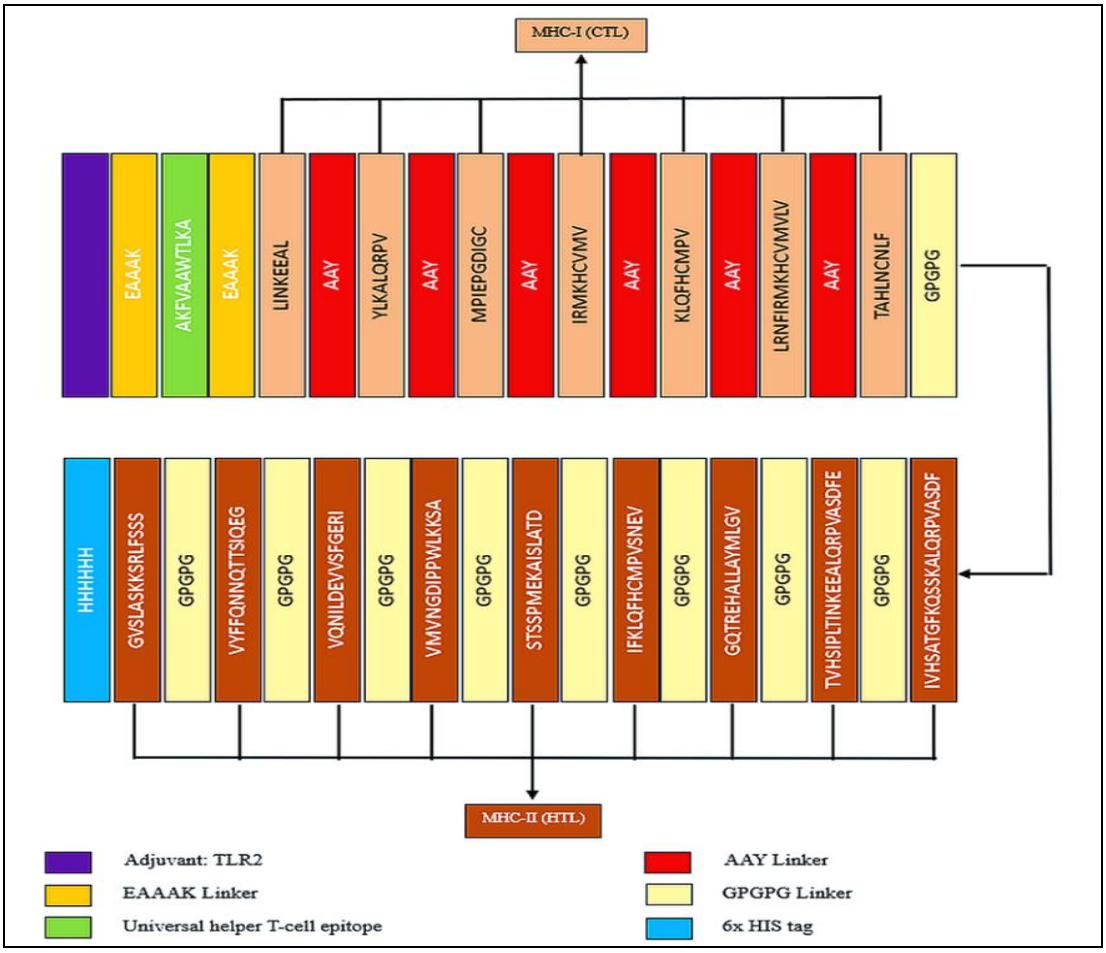

Fig 2: Depiction of epitope arrangement in the multi-epitope vaccine construct.

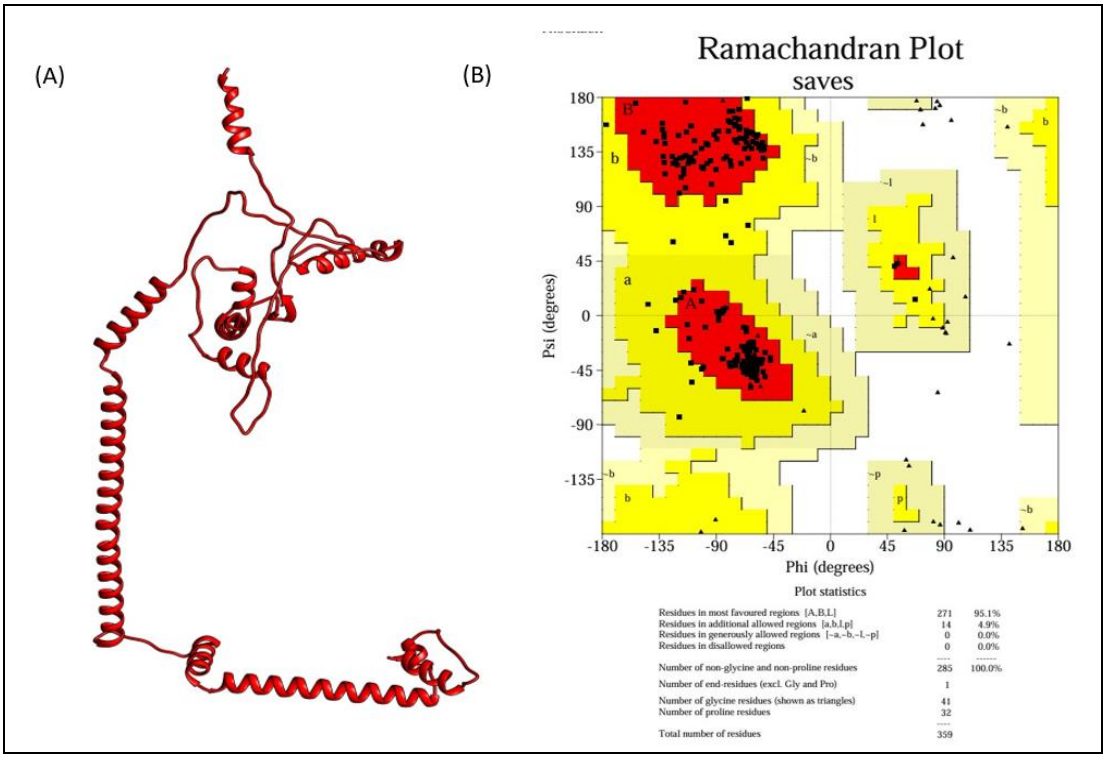

Fig 3: (A) Vaccine 3D structure, (B) Ramachandran plot of the vaccine.

7. Vaccine 3D Structure Prediction and Validation

The three-dimensional structure of the designed multi-epitope vaccine was predicted using the trRosetta server, which employs a deep learning-based algorithm to model protein structures by integrating inter-residue distance distributions, residue orientations, and geometric constraints to generate accurate tertiary structures. The predicted model was subsequently validated using the SAVES v6.0 server

through Ramachandran plot analysis to assess its stereochemical quality. The Ramachandran plot demonstrated that 95.1% of the residues were located in the most favored regions, while the remaining 4.9% were distributed within additionally allowed regions. Importantly, no residues were observed in generously allowed or disallowed regions, indicating the absence of significant steric clashes and confirming the high stereochemical quality of the predicted model. The structure comprised 285 non-glycine/non-proline residues, 41 glycine residues, 32 proline residues, and one terminal residue, with glycine residues exhibiting greater conformational flexibility and proline residues displaying the expected restricted conformations. Furthermore, the presence of distinct clusters corresponding to α-helical and β-sheet regions demonstrated that the vaccine construct adopted a well-defined and structurally stable tertiary conformation, supporting its suitability for subsequent molecular docking and molecular dynamics simulations (Fig 3).

8. Disulfide Engineering of the Vaccine

The structural stability of the vaccine construct was further enhanced through disulfide engineering using the Disulfide by Design v2.0 server. A total of 16 potential residue pairs were identified as candidates for disulfide bond formation based on geometric constraints. Among these, Cys11–Cys18 (energy: 1.83 kcal/mol, χ3: +73.62°) and Ser34–Gly37 (energy: 0.43 kcal/mol, χ3: +92.74°) were selected owing to their favorable energy and χ3 values. Additionally, these residues exhibited relatively low B-factor values, indicating reduced atomic flexibility and a well-ordered local structure. The introduction of these disulfide bonds is expected to enhance the structural rigidity and overall stability of the vaccine construct without significantly altering its native conformation.

9. Molecular Docking with Immune Receptor

Molecular docking demonstrated that the designed multi-epitope vaccine bound favorably within the TLR2 binding interface (PDB ID: 2Z7X) with a score of -61.23 +- 15.96, indicating a stable and specific vaccine–receptor complex. The vaccine adopted a complementary conformation toward the receptor, suggesting its potential to activate TLR2-mediated innate immune signaling. PDBsum analysis revealed that the interaction occurred between chain A and chain B, involving 32 and 24 interface residues, respectively, with interface areas of 1262 Å² and 1385 Å². The complex was stabilized by 12 hydrogen bonds, 9 salt bridges, and 139 non-bonded interactions, reflecting strong binding specificity and structural stability, while no intermolecular disulfide bonds were observed. Furthermore, the predominance of α-helical regions within the vaccine construct is likely to contribute to efficient receptor recognition and stable protein–protein interactions. Overall, these findings indicate that the vaccine construct forms a robust and biologically relevant interaction with TLR2, supporting its potential to initiate innate immune responses.

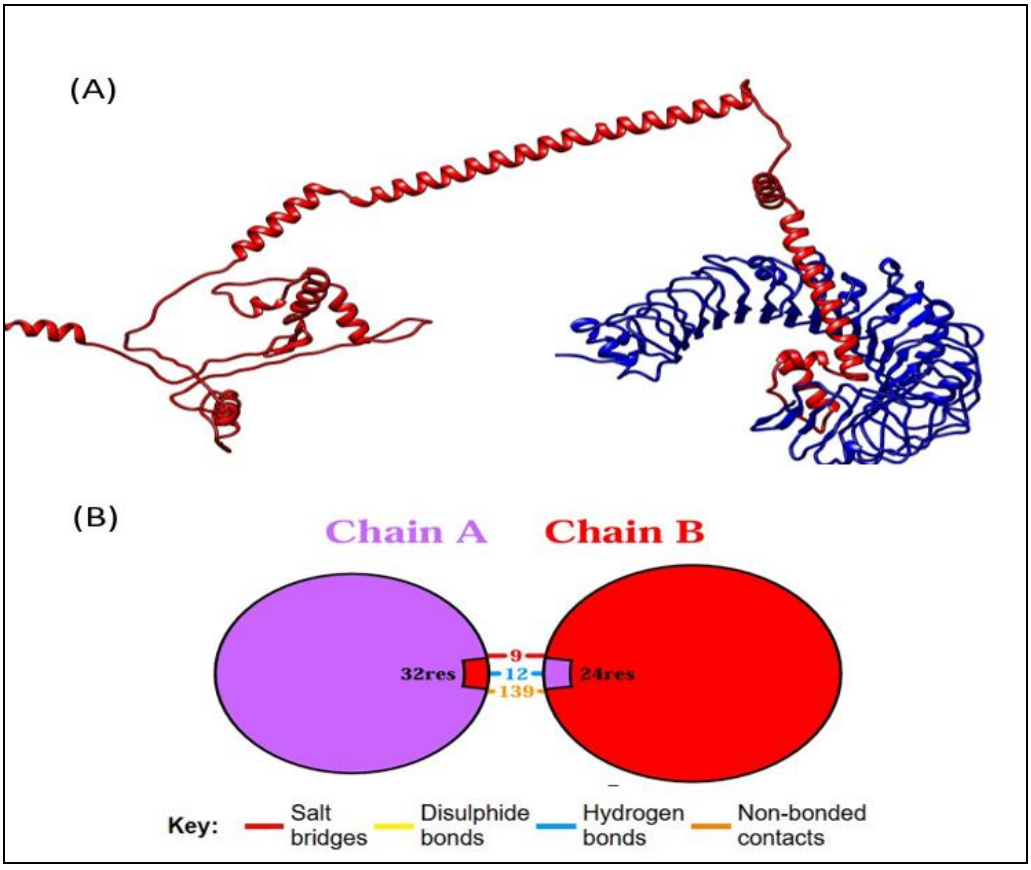

Fig 4: (A) Docked complex of vaccine and TLR2 receptor, (B) No. of interactions between vaccine and TLR2.

10. Molecular Dynamics Simulation of Docked Complex

The structural stability of the vaccine–TLR2 complex was evaluated through a 200 ns molecular dynamics (MD) simulation. The protein backbone RMSD initially increased during the equilibration phase and stabilized after approximately 30 ns, fluctuating between ~2.8 and 4.0 Å for the remainder of the simulation, indicating that the receptor

maintained a stable conformation. Although the ligand RMSD exhibited larger conformational changes relative to its initial docked pose, it gradually converged and remained stable after the early simulation period, suggesting that the vaccine adjusted within the TLR2 binding pocket before attaining a favorable binding orientation. RMSF analysis further demonstrated that most residues exhibited moderate fluctuations of 4-10 Å, with greater flexibility in the N-terminal region and in a few loop segments around residues 250–300 and the C-terminus. Importantly, the residues constituting the receptor-binding interface remained comparatively less flexible, indicating that the vaccine–TLR2 complex retained its structural integrity throughout the simulation. Overall, the MD simulation confirmed the dynamic stability of the docked complex and supported the robustness of the predicted vaccine–receptor interaction (Fig 5).

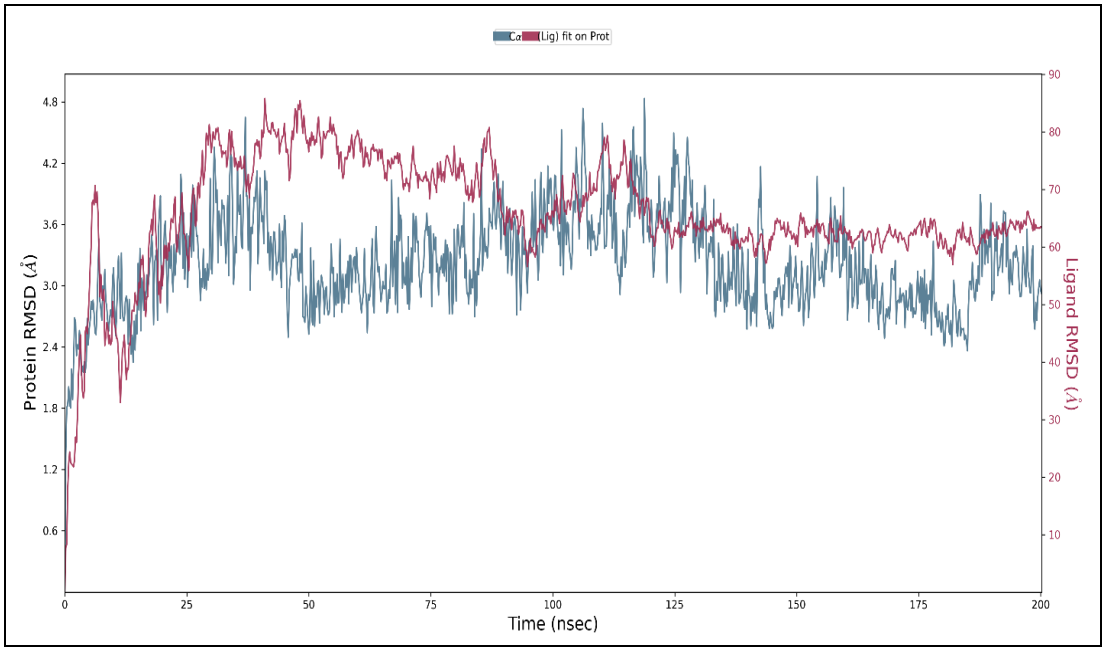

Fig 5: RMSD plot of vaccine-TLR2 complex

11. MM-GBSA Binding Free Energy Analysis

The binding affinity of the vaccine–TLR2 complex was further quantified using the Prime MM-GBSA method. The pre-MD MM-GBSA analysis predicted a binding free energy of −55.54 kcal/mol, indicating a favorable interaction immediately after molecular docking. Following the 200 ns MD simulation, the average binding free energy increased significantly to −248.13 ± 37.95 kcal/mol, indicating a substantial increase in binding strength after structural relaxation and dynamic equilibration. This marked decrease in binding free energy suggests that the vaccine adopted a more energetically favorable conformation within the TLR2 binding site during the simulation. Collectively, the docking, molecular dynamics, and MM-GBSA results indicate that the designed multi-epitope vaccine forms a stable and energetically favorable complex with TLR2, supporting its potential to effectively engage innate immune signaling.

12. Codon Optimization and *In Silico* Cloning

To evaluate the expression potential of the designed multi-epitope vaccine, the amino acid sequence was reverse-translated and codon-optimized using the Java Codon Adaptation Tool (JCat) for expression in *Escherichia coli* K-12. The optimized nucleotide sequence achieved a Codon Adaptation Index (CAI) of 1.0, indicating excellent codon usage compatibility with the host organism, while the GC content was 53.57%, falling within the optimal range (40–60%) for efficient transcription and translation. The optimized gene was subsequently cloned *in silico* into the pET-28a(+) expression vector using the HindIII and MluI restriction sites through SnapGene software. The resulting recombinant plasmid measured 5490 bp, with the inserted vaccine sequence correctly positioned within the multiple cloning site (MCS), confirming proper orientation and successful integration. These findings demonstrate the suitability of the vaccine construct for recombinant expression and facilitate its potential experimental production in *E. coli* (Fig 6).

13. Immune Simulation

The C-ImmSim immune simulation predicted a strong immune response following vaccination, characterized by rapid antigen clearance and robust antibody production. The antigen concentration reached approximately 6.7×10^5 counts/mL immediately after administration and declined to nearly zero by day 10, indicating efficient immune-mediated clearance. The combined IgM + IgG antibody response peaked at approximately 6.6×10^3, while IgM reached about 4.3×10^3, followed by sustained IgG1 and IgG1 + IgG2 responses of approximately $2.1–2.3 \times 10^3$, demonstrating effective class switching and humoral immunity. Cytokine profiling revealed a pronounced IFN-γ response, peaking at approximately 4.2×10^5 ng/mL, accompanied by elevated levels of IL-2 (~2.4×10^5 ng/mL), TGF-β (~7×10^4 ng/mL), and IL-10 (~4×10^4 ng/mL). These findings suggest that the vaccine can induce strong cellular and humoral immune responses, supporting its potential to confer effective, long-lasting protective immunity.

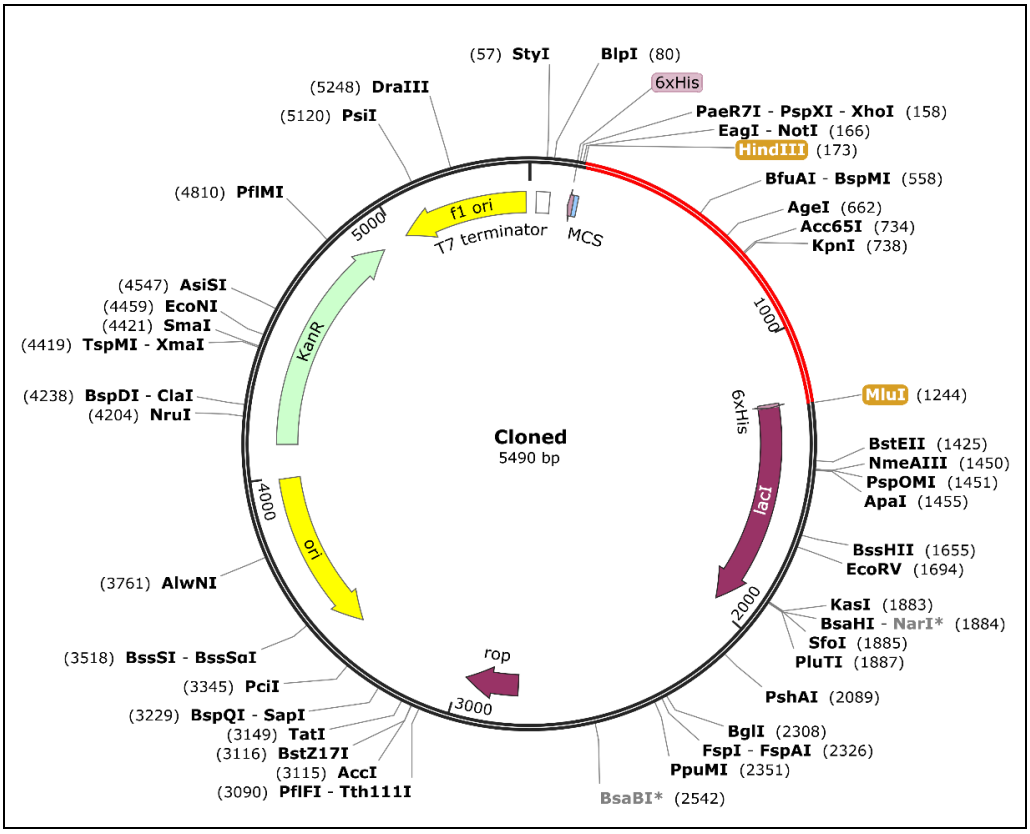

Fig 6: *In silico* cloning of the vaccine construct

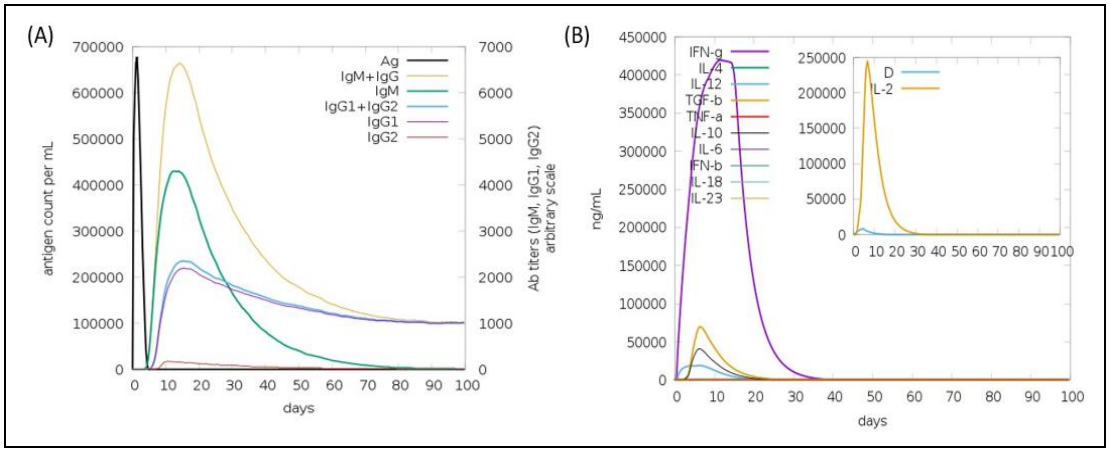

Fig 7: (A) Antigen count throughout the immune simulation, (B) Cytokine levels during the simulation

Conclusion

This study successfully employed an integrated immunoinformatics and structural vaccinology approach to design a novel multi-epitope therapeutic vaccine targeting leukemia-associated neoantigens. The selected CTL and HTL epitopes exhibited high antigenicity, non-allergenicity, and non-toxicity, making them suitable candidates for vaccine development. The final vaccine construct demonstrated favorable physicochemical properties, good structural quality, and enhanced stability following disulfide engineering. Molecular docking, molecular dynamics simulations, and MM-GBSA analyses confirmed stable, energetically favorable interactions with TLR2, while immune simulation predicted strong humoral and cellular immune responses. Furthermore, codon optimization and *in silico* cloning suggested efficient recombinant expression in *Escherichia coli*. Collectively, these findings indicate that the designed vaccine is a promising therapeutic candidate against leukemia; however, *in vitro* and *in vivo* experimental validation will be essential to confirm its safety, immunogenicity, and therapeutic efficacy before clinical application.

Declarations

Ethical Approval and Consent to Participate

The manuscript does not contain experiments using animals and humans.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Availability of supporting data

The data used and analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
Hemang Bhindi: Conceptualization, Software, Methodology, Data Analysis, Writing – original draft. **Dharmendrasinh F. Rao:** Conceptualization, Software, Methodology, Data Analysis, Writing – original draft, Writing – review & editing. Arav Sharma: Data Analysis, Writing- original draft. Saumya K. Patel: Visualization, Supervision, Investigation, Funding acquisition, Writing – review & editing.

Acknowledgments
We acknowledge the support of GUJCOST, DST, Government of Gujarat, for providing the Supercomputing facility, and GSBTM, DST, Government of Gujarat, for providing the BIN-Node Facility to the department.

References
1. Huang J, Chan SC, Lok V, Zhang L, Lucero-Prisno DE, Xu W, *et al*. The epidemiological landscape of multiple myeloma: a global cancer registry estimate of disease burden, risk factors, and temporal trends. Lancet Haematology, 2022, 9. https://doi.org/10.1016/S2352-3026(22)00165-X.
2. Zhang R, Guo C, Liang Y, Qing S, Liang Z, Wang Y, *et al*. The burden dynamics of leukemia in China from 1990 to 2021: an epidemiological analysis of trends, risk factors, and projections to 2036. Preventive Medicine Reports,2025:57. https://doi.org/10.1016/j.pmedr.2025.103191.
3. Richard G, Princiotta MF, Bridon D, Martin WD, Steinberg GD, De Groot AS. Neoantigen-based personalized cancer vaccines: the emergence of precision cancer immunotherapy. Expert Review of Vaccines,2022:21. https://doi.org/10.1080/14760584.2022.2012456.
4. Jabbour E, Kantarjian H. Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. American Journal of Hematology,2020:95. https://doi.org/10.1002/ajh.25792.
5. Singh P, Khatib MN, R R, Kaur M, Srivastava M, Barwal A, *et al*. Advancements and challenges in personalized neoantigen-based cancer vaccines. Oncology Reviews,2025:19. https://doi.org/10.3389/or.2025.1541326.
6. Katsikis PD, Ishii KJ, Schliehe C. Challenges in developing personalized neoantigen cancer vaccines. Nature Reviews Immunology,2024:24. https://doi.org/10.1038/s41577-023-00937-y.
7. Yu J, Sun H, Cao W, Song Y, Jiang Z. Research progress on dendritic cell vaccines in cancer immunotherapy. Experimental Hematology and Oncology,2022:11. https://doi.org/10.1186/s40164-022-00257-2.
8. Rezvani K. Peptide vaccine therapy for leukemia. International Journal of Hematology,2011:93. https://doi.org/10.1007/s12185-011-0781-3.
9. Zhao X, Wei L, Zhang X. Computational methods and data resources for predicting tumor neoantigens. Briefings in Bioinformatics,2025:26. https://doi.org/10.1093/bib/bbaf302.
10. Koşaloğlu-Yalçın Z, Blazeska N, Vita R, Carter H, Nielsen M, Schoenberger S, *et al*. The Cancer Epitope Database and Analysis Resource (CEDAR). Nucleic Acids Research,2023:51. https://doi.org/10.1093/nar/gkac902.
11. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics,2007:8. https://doi.org/10.1186/1471-2105-8-4.
12. Dimitrov I, Bangov I, Flower DR, Doytchinova I. AllerTOP v.2 - A server for *in silico* prediction of allergens. Journal of Molecular Modeling,2014:20. https://doi.org/10.1007/s00894-014-2278-5.
13. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. Peptide toxicity prediction. Methods in Molecular Biology,2015:1268. https://doi.org/10.1007/978-1-4939-2285-7_7.
14. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. 2013. http://crdd.osdd.net/.
15. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, *et al*. NetMHCpan, a method for MHC class i binding prediction beyond humans. Immunogenetics,2009:61. https://doi.org/10.1007/s00251-008-0341-z.
16. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Research,2021:48. https://doi.org/10.1093/NAR/GKAA379.
17. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics, 2006, 7. https://doi.org/10.1186/1471-2105-7-153.
18. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Engineering,2001:14. https://doi.org/10.1093/protein/14.8.529.
19. Chen X, Zaro JL, Shen WC. Fusion protein linkers: Property, design and functionality. Advanced Drug Delivery Reviews,2013:65. https://doi.org/10.1016/j.addr.2012.09.039.
20. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al*. Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server. n.d. http://www.expasy.org/tools/.
21. Craig DB, Dombkowski AA. Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. BMC Bioinformatics,2013:14. https://doi.org/10.1186/1471-2105-14-346.
22. Honorato RV, Trellet ME, Jiménez-García B, Schaarschmidt JJ, Giulini M, Reys V, *et al*. The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. Nature Protocols,2024:19:3219-3241. https://doi.org/10.1038/s41596-024-01011-0.
23. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB

entries. Protein Science,2018:27:129-134. https://doi.org/10.1002/pro.3289.
24. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron,2018:99. https://doi.org/10.1016/j.neuron.2018.08.011.
25. Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field. Journal of Chemical Theory and Computation,2010:6:1509-1519. https://doi.org/10.1021/ct900587b.
26. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, *et al*. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Research,2005:33. https://doi.org/10.1093/nar/gki376.
27. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends in Biotechnology,2004:22. https://doi.org/10.1016/j.tibtech.2004.04.006.
28. Castiglione F, Duca K, Jarrah A, Laubenbacher R, Hochberg D, Thorley-Lawson D. Simulating Epstein-Barr virus infection with C-ImmSim. Bioinformatics,2007:23. https://doi.org/10.1093/bioinformatics/btm044.