



OPEN ACCESS

EDITED BY

Li Ang,
First Affiliated Hospital of Zhengzhou
University, China

REVIEWED BY

Denise Cecil,
University of Washington, United States
Paula Dobosz,
Poznan University of Medical Sciences, Poland

*CORRESPONDENCE

Jianping Lin
✉ jianpinglin@nankai.edu.cn

RECEIVED 18 June 2025

ACCEPTED 07 October 2025

PUBLISHED 23 October 2025

CITATION

Wang L, Zhou X, Wei Y and Lin J (2025) A
novel multi-epitope mRNA vaccine against
colorectal cancer: in silico design and
immune efficacy profiling.
Front. Immunol. 16:1649091.
doi: 10.3389/fimmu.2025.1649091

COPYRIGHT

© 2025 Wang, Zhou, Wei and Lin. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A novel multi-epitope mRNA vaccine against colorectal cancer: in silico design and immune efficacy profiling

Lin Wang^{1,2}, Xiaofei Zhou³, Yu Wei³ and Jianping Lin^{1,2*}

¹State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin, China, ²College of Pharmacy and Tianjin Key Laboratory of Molecular Drug Research, Nankai University, Tianjin, China, ³Tianjin BioAI-global Technology Co., Ltd, Tianjin, China

Background: Colorectal cancer (CRC) has emerged as a growing global health challenge, while immunotherapy, particularly mRNA-based cancer vaccines, has emerged as a promising approach due to its ability to induce targeted immune responses with minimal systemic toxicity. This study aimed to design a multi-epitope mRNA vaccine targeting tumor-specific antigens (TSAs) as a cancer therapeutic regimen.

Results: We chose six CRC-specific TSAs and selected their appropriate epitopes with immunoinformatic tools. In order to enhance the vaccine stability, we subsequently optimized the open reading frame (ORF) sequences, which demonstrated the highest structural stability among all evaluated approaches. Furthermore, we built a CNN model combined with RNA large language model (RNA-FM) embeddings to screen 212 candidate 5'UTR sequences and identify variants that boost the vaccine's translational efficiency. Finally, in silico immune simulations confirmed the vaccine's ability to elicit robust humoral and cellular immune responses.

Conclusion: This study presents an in silico designed mRNA vaccine against colorectal cancer (CRC). Immune simulations demonstrated that this mRNA vaccine can elicit strong antitumor immune responses, indicating it is an effective and promising candidate that warrants further *in vitro* and *in vivo* investigations. Additionally, this work highlights the potential of in silico approaches in vaccine design and provides valuable insights for the development of effective vaccines targeting CRC.

KEYWORDS

colorectal cancer, mRNA vaccine, epitopes, codon optimization, MHC allele

1 Introduction

Colorectal Cancer (CRC) ranks as the third most prevalent malignant tumor worldwide, accounting for approximately 52,980 deaths in 2021 (1). According to global cancer research projections, an estimated 3.2 million new cases and 1.6 million deaths are anticipated by 2040, representing a 63% and 73.4% increase from 2020 levels, respectively (2). Notably, the incidence of early-onset CRC (diagnosed in individuals under 50 years of age) has risen significantly over recent decades (3). Furthermore, CRC often progresses asymptotically during early stages, with clinical manifestations typically emerging only at advanced disease phases. These factors underscore the urgent need for novel therapeutic strategies to combat CRC (4).

Current standard CRC treatments, including surgical intervention, chemotherapy (5), and radiotherapy (6), exhibit substantial limitations. Surgery may not be feasible for all patients, particularly those with advanced-stage or metastatic disease. Chemotherapeutic agents frequently induce severe adverse effects such as nausea, vomiting, alopecia, and immunosuppression, significantly compromising patients' quality of life (7). Moreover, the development of chemoresistance further diminishes therapeutic efficacy. Radiotherapy, while beneficial, risks damaging adjacent healthy tissues near tumor sites (8).

Recent advances in immunotherapy have positioned it as a promising alternative for cancer management (9). Among immunotherapeutic strategies, cancer vaccines have garnered considerable attention for their potential to stimulate targeted immune responses against malignant cells (10) while minimizing systemic toxicity. However, conventional vaccine platforms, including peptide-based (11), protein-based (12), and whole-cell vaccines (13), face critical challenges such as low immunogenicity, complex manufacturing processes (13), and limited capacity to simultaneously target multiple antigens (14).

The global COVID-19 pandemic has accelerated mRNA vaccine technology into a revolutionary paradigm for vaccine development. Compared to traditional approaches, mRNA vaccines offer distinct advantages: 1) Rapid design and cell-free production capabilities, crucial for addressing emerging pathogens or cancer subtypes (15); 2) Capacity to encode multiple antigens, enabling broad-spectrum immune activation against heterogeneous tumor populations (16); 3) Non-integrative nature, eliminating the risks of insertional mutagenesis associated with DNA-based vaccines (17).

When designing effective mRNA vaccines targeting CRC, genes abnormally overexpressed in the cancer tissues of CRC patients can serve as ideal tumor-specific antigens (TSAs). According to the study by Liu et al. (18), CRC patients with high expression of six genes, *THBS2*, *FSTL3*, *TNNT1*, *BGN*, *CTHRC1*, and *NOX4*, exhibit shorter overall survival (OS) and shorter recurrence-free survival (RFS). Additionally, the expression levels of these six genes are associated with the infiltration levels of antigen-presenting cells (APCs) and T lymphocytes. Thus, in this study, we selected these six genes as ideal CRC-TSAs for mRNA vaccine development. Mechanistically, *THBS2* promotes CRC metastasis by modulating the WNT/ β -catenin signaling pathway while suppressing antitumor

immunity through HIF1A/lactate/GPR132 axis interactions (19). *FSTL3* drives epithelial-mesenchymal transition (EMT) via fibronectin1/ α 5 β 1 interactions and serves as a biomarker for extracellular matrix remodeling in CRC diagnosis (20). *TNNT1* regulates EMT processes through miR-873-mediated mechanisms and functions as a prognostic indicator for colon adenocarcinoma (21). *BGN* has been reported to exert a significant impact on CRC cell proliferation, cell cycle progression, apoptosis, invasion, and migration (22). Previous studies have shown that knockout of *BGN* can inhibit the proliferation and migration of CRC cells (23). *CTHRC1* enhances the proliferation and invasiveness of human CRC cells by activating the Wnt/planar cell polarity (PCP) signaling pathway (24). *NOX4* regulates the expression of genes related to cancer cell biological behaviors, thereby promoting CRC cell proliferation, inhibiting apoptosis, and enhancing cell migration and invasion (25). The schematic diagram illustrating how the six aforementioned genes regulate the progression of colorectal cancer (CRC) is shown in Figure 1.

This study aims to develop and evaluate a novel mRNA vaccine targeting CRC-specific TSAs to enhance antitumor immunity, potentially offering a therapeutic alternative for CRC patients. Through rational antigen selection and advanced bioengineering approaches, we seek to address current limitations in CRC management while leveraging the unique advantages of mRNA vaccine technology.

2 Material and methods

2.1 Retrieval of tumor-specific protein sequences

The UniProt database (<https://www.uniprot.org/>) was utilized to acquire amino acid sequences of the six target proteins using their respective accession numbers (1): *THBS2* (P35442), (2) *FSTL3* (P95633), (3) *TNNT1* (P13805), (4) *BGN* (P21810), (5) *CTHRC1* (Q96CG8), and (6) *NOX4* (Q9NPH5).

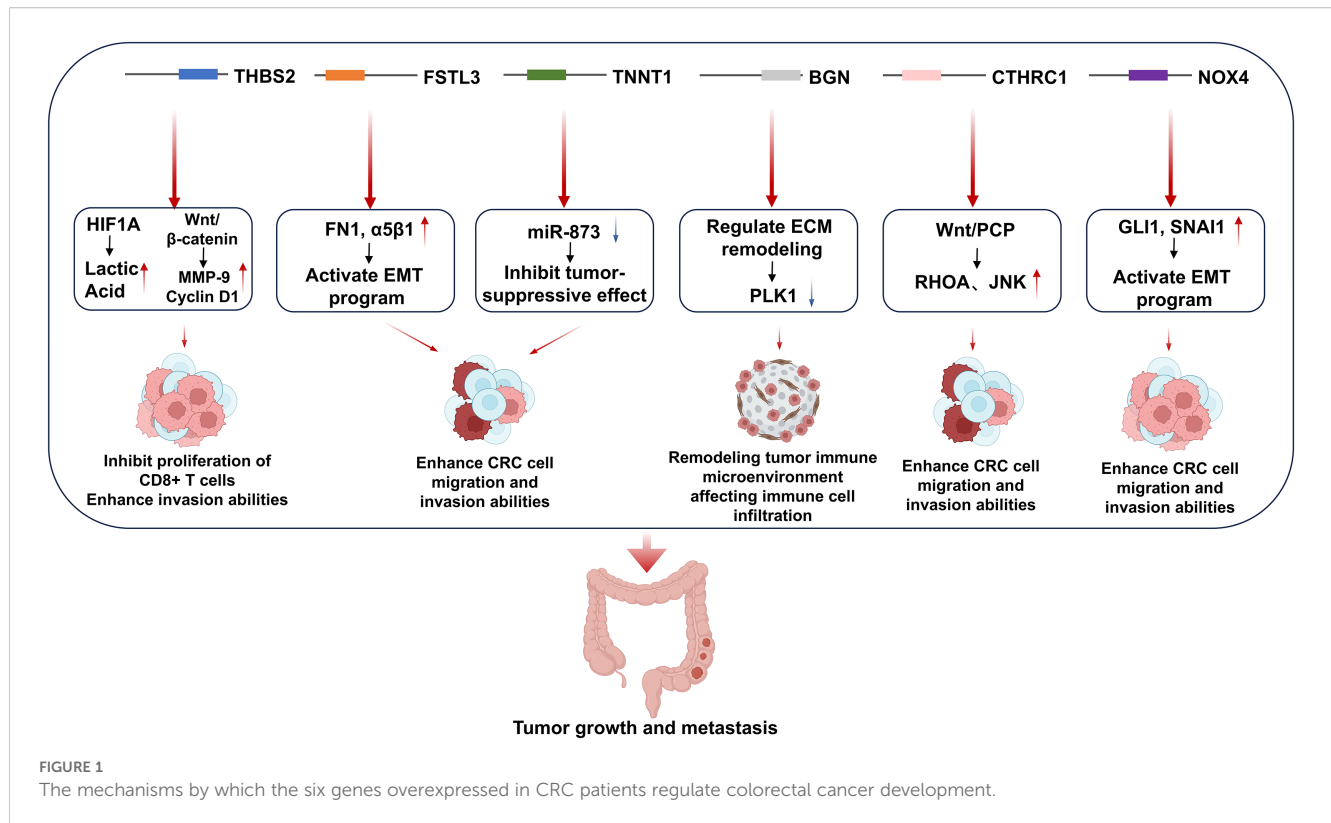
2.2 Prediction of immune cell epitopes

2.2.1 B-cell epitope prediction

B-cell epitopes were predicted using the ABCPred web server (<https://webs.iitd.edu.in/raghava/abcpred/>), a machine learning platform trained on linear B-cell epitope data from the BCIPEP database. Parameters were configured with an epitope length of 16 amino acid residues, a prediction threshold of 0.5, and an activated overlap filter to ensure sequence exclusivity (26).

2.2.2 Cytotoxic T lymphocyte epitope prediction

HLA class I-restricted epitopes were identified through the NetMHCpan 4.1 EL algorithm on the IEDB Analysis Resource (<http://tools.iedb.org/mhci/>). Predictions employed default reference sets of HLA alleles, with epitopes ranked by percentile binding scores.



2.2.3 Helper T lymphocyte epitope prediction

HLA class II-binding epitopes were determined using the NetMHCIIpan 4.1 EL method via the IEDB server (<http://tools.iedb.org/mhcii/>). Epitope candidates were similarly prioritized based on predicted binding affinities (27).

2.3 Prediction of epitopes' antigenicity, allergenicity and toxicity

Following the prediction of linear B-cell epitopes (LBL), cytotoxic T lymphocyte (CTL), and helper T lymphocyte (HTL) epitopes, we systematically evaluated the immunobiological properties of candidate epitopes. Antigenicity was assessed using the VaxiJen v3.0 web server (<http://www.ddg-pharmfac.net/vaxijen/>), configured with the tumor protein antigen model at a threshold score of 0.5. This machine learning platform employs alignment-free algorithms trained on bacterial, viral, and tumor antigen datasets to predict whole-protein immunogenicity. Allergenicity screening was performed through AllerTop v2.1 (https://www.ddg-pharmfac.net/allertop_test/), which utilizes amino acid propensity scales and auto-cross covariance transformation for epitope safety evaluation. Toxicity predictions were conducted via ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred/>), employing support vector machine models trained on toxic/non-toxic peptide datasets. Epitopes demonstrating strong antigenicity (VaxiJen score ≥ 0.5), non-allergenicity, and non-toxicity were retained for downstream analyses.

2.4 Population coverage of epitopes

Geographical and environmental factors drive substantial diversity in human leukocyte antigen (HLA) alleles across global populations. To evaluate the epidemiological relevance of our vaccine candidate, we conducted global population coverage analysis using the Immune Epitope Database (IEDB) Population Coverage Calculation Tool (<http://tools.iedb.org/population/>). This platform calculates cumulative coverage probabilities for selected HLA class I (CTL) and class II (HTL) epitopes based on their binding affinities to region-specific HLA allele distributions across 16 geographical regions. The analysis integrated frequency data of HLA-A, -B, -C, and -DRB1 alleles from ethnically diverse populations to estimate potential vaccine efficacy thresholds.

2.5 Molecular docking between T-lymphocyte epitopes and MHC alleles

To evaluate the binding affinity between selected T lymphocyte epitopes and their corresponding major histocompatibility complex (MHC) class I alleles, we performed systematic molecular docking simulations. MHC-I crystal structures were retrieved from the RCSB Protein Data Bank (PDB). Structural preprocessing involved removal of water molecules and non-essential ligands using PyMOL v2.5.7 (28). Epitope sequences underwent 3D conformational modeling via PEP-FOLD 3.5 server (<https://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/>), followed

by complex assembly with MHC-I proteins. Energy minimization was conducted using Rosetta's Relax protocol (v3.13) to optimize molecular geometries (29). Finally, epitope-MHC docking simulations were executed in local refinement mode through RosettaDock, employing Monte Carlo minimization algorithms to sample low-energy binding conformations.

2.6 Screening of 5'UTR

The 5'UTR sequence in particular is a major determinant of translation efficiency and thus an intriguing target for engineering (30). The ribosomal 43 S pre-initiation complex (PIC) scans the 5'UTR in the 5'-to-3' direction until a start codon is found. Therefore, 5'UTRs can affect translation by capturing PICs prematurely via upstream start codons (uAUGs) and ORFs (uORFs) (31). To identify 5'-UTR variants with enhanced translational efficiency, we employed RNA-FM (32), a pretrained large language model, to generate embeddings for 83,919 synthetic human 5'-UTRs of 75 distinct lengths from the combinatorial library generated by Sample et al. (33). The convolutional neural network (CNN) architecture was constructed based on Sample et al. to screen candidate 5'-UTRs. The CNN model was constructed as three 1D convolutional layers with 120 filters and a ReLU activation for each layer. The third convolution layer output one channel, which was fed into two fully-connected layers with one output node as the final prediction. The inputs of the model were RNA-FM embeddings (640 dim). In this study, the data was partitioned into a training set and a validation set with a ratio of 8:2. This partitioning strategy aimed to ensure sufficient data for model training while enabling effective validation of the model's generalization ability.

Our 5'-UTR candidate pool integrates four distinct sources: 1) 212 variable-length sequences from Chu et al. (34), 2) 8 Kozak sequence-containing variants from Li et al. (35), 3) The 5'-UTR of human α -globin genes, and 4) the 5'-UTR of BNT162b2(SARS-CoV-2 vaccine, from Pfizer/BioNTech) (36).

2.7 Rational design of mRNA vaccine construct

A potent mRNA vaccine construct requires systematic integration of five critical components (1): an open reading frame (ORF) encoding antigenic elements (2), the Untranslated Regions (UTRs) flanking the coding regions, (3) a Kozak sequence incorporating the start codon (GCCACCAUGG) to enhance translational initiation (37), (4) functionally optimized linkers, and (5) regulatory termination signals. The proposed construct features a 5'→3' architecture comprising a modified 5'm⁷GCap structure, followed by an optimized 5' untranslated region (5'UTR) and Kozak sequence to maximize ribosomal engagement. The ORF initiates with a signal peptide for intracellular trafficking, connected via an EAAAK rigid linker to helper T lymphocyte (HTL) epitopes interconnected through GP GPG spacers that maintain domain autonomy. This HTL cluster transitions via KK linkers to linear B-cell epitopes (LBL), followed by AAY-linked

cytotoxic T lymphocyte (CTL) epitopes strategically positioned upstream of an MHC class I trafficking domain (MITD). The MITD sequence enhances immunogenicity through dual mechanisms: 1) optimizing antigen presentation efficiency via endosomal targeting motifs, and 2) directing vaccine components to antigen-processing compartments. Epitope segregation through GP GPG, KK, and AAY linkers ensures proper conformational folding while preserving immunological functionality. The construct terminates with a UAA stop codon, a stabilizing 3' untranslated region (3'UTR), and a 120-nucleotide poly(A) tail to ensure mRNA stability and translational fidelity. This multi-layered design leverages structural bioengineering principles to balance epitope accessibility, intracellular trafficking efficiency, and immune activation capacity.

2.8 Prediction of antigenicity, allergenicity, toxicity and physicochemical properties of the vaccine construct

Following vaccine sequence assembly, systematic bioinformatic validation was performed to assess four critical properties: (1) antigenicity, (2) allergenicity, (3) toxicity, and (4) physicochemical stability. Antigenicity prediction employed dual machine learning platforms—VaxiJen v3.0 (threshold: 0.5; tumor antigen model) and ANTIGENpro (SCRATCH Protein Predictor suite)—using the translated amino acid sequence excluding trafficking/processing elements (tPA and MITD domains). Allergenicity screening was conducted through AllerTop v2.1 (https://www.ddg-pharmfac.net/allertop_test/) using default parameters. Toxicity profiling utilized ToxinPred (<https://webs.iitd.edu.in/raghava/toxinpred/>), implementing SVM models trained on experimentally validated toxic peptides. Physicochemical characterization was performed via the ProtParam tool (<https://web.expasy.org/protparam/>), quantifying six essential parameters: amino acid composition, molecular weight, theoretical isoelectric point (pI), instability index (II), aliphatic index (AI), and grand average of hydropathicity (GRAVY). This multi-platform validation strategy ensures structural integrity while confirming immunological safety and biophysical stability of the vaccine candidate.

2.9 In silico immune simulation

To predict the immunogenic potential of the mRNA vaccine construct in humans, we conducted in silico immune simulations using the C-IMMSIM immunoinformatics platform (<https://kraken.iac.rm.cnr.it/C-IMMSIM/>). The vaccination regimen was configured to follow the recommended dosage schedule of current vaccines: three 1,000-unit doses administered at time-steps 1, 84 (3 weeks), and 168 (6 weeks). All parameters were maintained at default settings. Dynamic simulation outputs quantified key immunological metrics: 1) Antigen-specific lymphocyte proliferation rates, 2) Cytokine production profiles, 3) Memory cell differentiation kinetics, and 4) Antibody titer trajectories. This computational framework enables systematic evaluation of the

vaccine's capacity to elicit coordinated humoral and cellular immune responses while predicting long-term immunological memory formation.

2.10 Codon optimization and sequence refinement

Leveraging the degeneracy of the genetic code, we implemented a multi-algorithm optimization strategy to enhance translational efficiency of the mRNA vaccine construct. Five codon optimization methods, including LinearDesign (38), were systematically applied to resolve synonymous codon conflicts while maximizing expression potential. Post-optimization sequences underwent rigorous biophysical characterization through three key metrics: 1) Minimum free energy (MFE) of mRNA secondary structures predicted by RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>), 2) GC content analysis (optimal range: 45–55%), and 3) Codon Adaptation Index (CAI) calculations relative to human codon usage tables. This combinatorial approach achieves dual objectives: (a) elimination of cryptic splice sites and ribosomal drop-off sequences through codon bias correction, and (b) stabilization of mRNA architecture via thermodynamic optimization of folding patterns.

2.11 Secondary structure prediction of the designed mRNA vaccine

The RNAfold web server was employed to predict the secondary structure of the mRNA vaccine construct using McCaskill's partition function algorithm (39). This computational framework calculates the minimum free energy (MFE) conformation through dynamic programming-based thermodynamic modeling of RNA folding pathways. The analysis yielded two critical outputs: 1) the MFE-optimized secondary structure visualization, and 2) quantitative thermodynamic stability metrics (ΔG in kcal/mol). Subsequent evaluation of mRNA structural features focused on identifying persistent stem-loop formations and regions of high base-pairing potential that might impede ribosomal scanning. This characterization enabled rational design optimization to balance thermodynamic stability ($\Delta G < -300$ kcal/mol) with translational efficiency through strategic codon rearrangement in unstable regions (MFE > -150 kcal/mol). The refined mRNA architecture demonstrates enhanced resistance to endonucleolytic degradation while maintaining optimal ribosomal accessibility-critical parameters for ensuring structural integrity and sustained antigen expression in human physiological conditions.

2.12 Molecular docking of the designed vaccine

To investigate the vaccine construct's innate immune activation potential, we performed protein-protein docking between predicted

vaccine epitopes and human Toll-like receptor 3 (TLR3; PDB ID 1ZIW) and Toll-like receptor 4 (TLR4; PDB ID 3FXI) using the ClusPro 2.0 server (<https://cluspro.org/>). This rigid-body docking platform employs a hierarchical protocol combining fast Fourier transform (FFT) correlation approaches with Monte Carlo minimization to sample $>10^9$ possible binding conformations.

3 Results

3.1 B-cell epitope prediction and selection

B-cell epitopes were predicted for the amino acid sequences of THBS2, FSTL3, TNNT1, BGN, CTHRC1, and NOX4 using the ABCPred web server. For each protein, the top five scoring epitopes were initially retained based on prediction rank. Subsequent refinement employed a tripartite filtering strategy: 1) Antigenicity validation via VaxiJen v3.0 (threshold: 0.5; tumor antigen model), 2) Allergenicity screening using AllerTop v2.1, and 3) Toxicity profiling through ToxinPred. A total of six epitopes demonstrating strong antigenicity (VaxiJen score ≥ 0.5), non-allergenicity, and non-toxicity were selected for final inclusion (Supplementary Table S1). This stringent selection protocol ensures exclusive retention of epitopes with optimal immunogenic potential while mitigating risks of hypersensitivity or cytotoxic responses.

3.2 Prediction and estimation of the CTL epitopes

Potential cytotoxic T lymphocyte (CTL) epitopes across the six target proteins were predicted using the MHC-I Binding Predictions tool within the IEDB Analysis Resource (<http://tools.iedb.org/mhci/>). Predicted epitopes were subsequently filtered through a triaxial validation pipeline: 1) Antigenicity scoring via VaxiJen v3.0 (threshold ≥ 0.5), 2) Allergenicity assessment using AllerTop v2.1, and 3) Toxicity profiling through ToxinPred. Epitopes were prioritized based on combined metrics: percentile binding rank (<1.0), antigenicity score (>0.7), and evolutionary conservation index (>0.8 in ConSurf analysis). Twelve high-affinity epitopes residing in phylogenetically conserved regions were ultimately selected for vaccine incorporation (Supplementary Table S1), ensuring broad HLA coverage and variant-resistant immunogenicity.

3.3 Prediction and estimation of the HTL epitopes

Potential helper T lymphocyte (HTL) epitopes were identified through comprehensive analysis of the six target proteins using the IEDB MHC-II Binding Predictions tool (<http://tools.iedb.org/mhcii/>) with NetMHCIIpan 4.1 algorithm. Following rigorous triaxial screening—antigenicity assessment (VaxiJen v3.0 score ≥ 0.5), allergenicity profiling (AllerTop v2.1), and toxicity

evaluation (ToxinPred), 18 epitopes residing in phylogenetically conserved domains (ConSurf conservation score >0.85) were prioritized for vaccine inclusion (Supplementary Table S1). After calculation and filtering, 18 selected HTL epitopes could induce Th1/Th2-polarized cytokine responses (IFN- γ , IL-4, IL-10) through TCR-MHC II complex stabilization, thereby amplifying both humoral and cellular arms of vaccine-induced immunity via cognate T-B lymphocyte collaboration.

3.4 Population coverage analysis

The worldwide epidemiological relevance of the vaccine construct was evaluated through HLA allele population coverage analysis using the IEDB Population Coverage Calculation Tool (<http://tools.iedb.org/population/>). The 30 incorporated HLA alleles demonstrated differential coverage across 16 geographical regions, with peak efficacy observed in Europe (89.13%) and North America (84.69%) (Table 1). Comparatively reduced coverage rates were identified in Central America (15.29%), South America (52.23%), and Northeast Asia (52.38%), reflecting regional disparities in HLA allele distribution patterns (details in Supplementary Figure S1).

3.5 Molecular docking between HLA alleles and the selected T-lymphocyte epitopes

Through the previous prediction and screening steps, 30 lymphocyte epitopes and their corresponding HLA alleles were identified (Supplementary Table S2), and six of these epitope-allele pairs underwent molecular docking simulations using Rosetta’s LocalDock module. Crystal structures of selected HLA class I molecules (PDB IDs: 8EMF, 7LG0, 7L1C, 4LNR, 5VUD, 6MPP) were energy-minimized to remove steric clashes (<0.3 Å RMSD deviation). Docking metrics (Supplementary Table S3) revealed two critical energy parameters: (1) Total_Score, predominantly reflecting monomeric folding energy, and (2) Interface_Score (I_{sc}), quantifying interaction energy across the binding interface. As emphasized in Rosetta documentation, I_{sc} provides superior predictive value for epitope-HLA binding stability. The MSDTEEQEY epitope demonstrated optimal binding with HLA-A*01:01, achieving the lowest I_{sc} value (-115.51 REU) and complete structural accommodation within the HLA binding groove (Figure 2F).

3.6 Screening of 5’-UTR

Our CNN model was trained and validated over 50-epoch using 83,919 human 5’-UTR sequences (spanning 75 distinct lengths) from Sample et al.’s library. Optimal regression performance was achieved at epoch 15, yielding an R² value of 0.844 with corresponding error metrics (mean squared error [MSE] = 0.307, mean absolute error [MAE] = 0.397, root mean squared error [RMSE] = 0.554). External

prediction conducted on 222 literature-curated 5’-UTR sequences identified 10 top-performing candidates, with complete rankings detailed in Table 2. The highest-ranked variant (sequence: GGGATCTTATTCCACCTTCTGAAGCTTCTGTGCGAACCAAGTTGTAAGGAGA) was ultimately selected as the 5’-UTR component for our CRC mRNA vaccine construct.

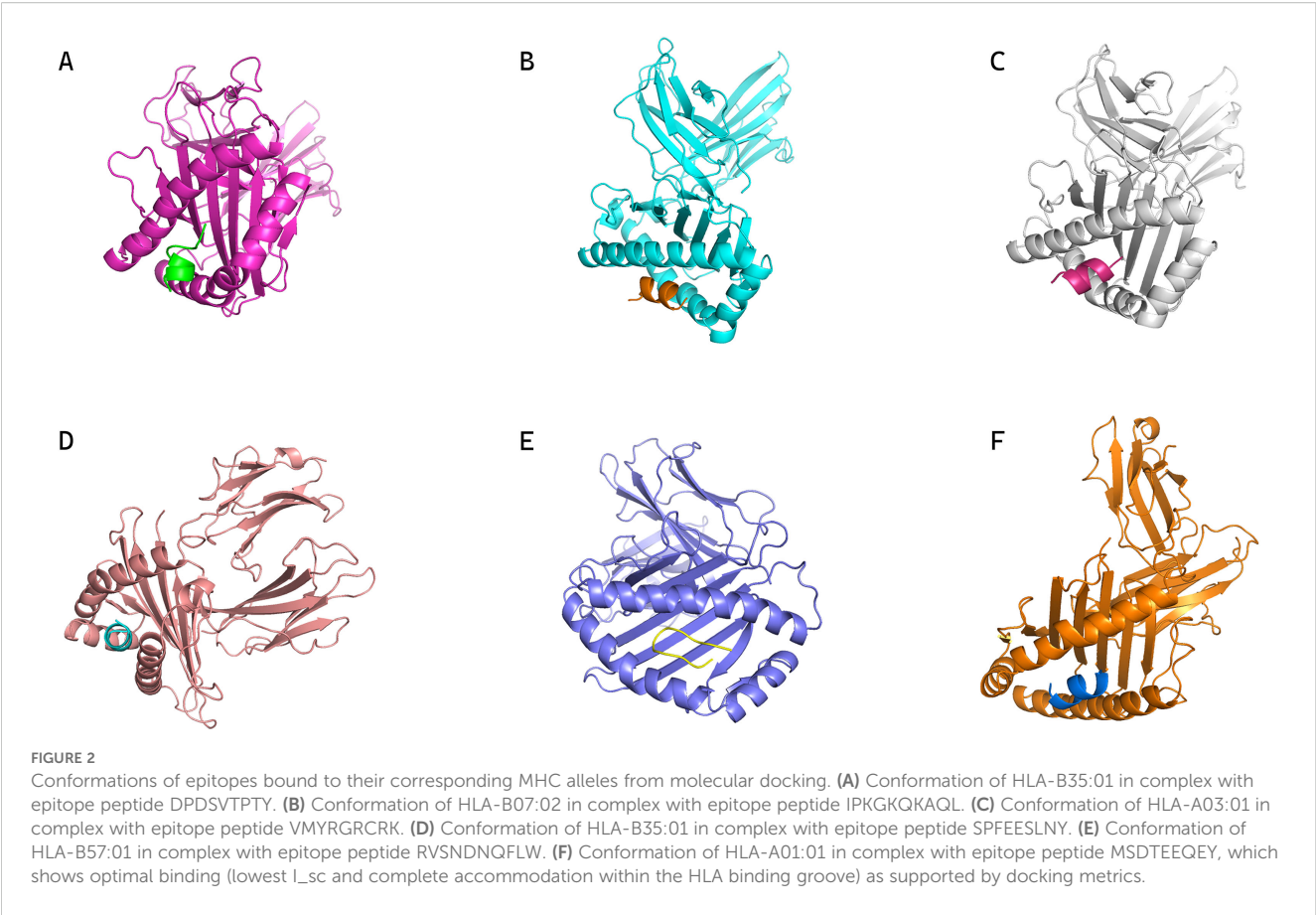
3.7 Vaccine construct design

The mRNA vaccine construct was proposed to be arranged from the N to C terminus in the following order: 5’^m7GCap, 5’UTR, Kozak sequence, Signal peptide (tPA)–GPGPG Linker–LLQVVYLHSNNITKV–GPGPG Linker–TTLLDLQNNDISELR–GPGPG Linker–KISKIHEKAFSPLRK–GPGPG Linker–RDGFKGEKGECLRES–GPGPG Linker–GRDGFKEKGECLRE–GPGPG Linker–MNSTINIHRHTSSVEG–GPGPG Linker–FVSSMGSGNPAPGGV–GPGPG Linker–GFVSSMGSGNPAPGG–GPGPG Linker–VSSMGSGNPAPGGVC–GPGPG Linker–IQKIIGEKYHALNSR–GPGPG Linker–KPAEFTQHKFVKICM–GPGPG Linker–RWKLLFDEIAKYNRG–GPGPG Linker–NCPYVHNPAQIDTDN–GPGPG Linker–DNCPIYSNANQADHD–GPGPG Linker–TAQLKQDGKSRGTLL–GPGPG Linker–SEKFDLMAKLKQQKY–GPGPG

TABLE 1 The population coverage of each epitope and its corresponding HLA allele in the mRNA vaccine.

population/area	coverage	average_hit	pc90
Central Africa	63.33%	1.32	0.27
Central America	15.29%	0.22	0.12
East Africa	70.10%	1.49	0.33
East Asia	63.42%	1.32	0.27
Europe	89.13%	2.94	0.92
North Africa	69.31%	1.66	0.33
North America	84.69%	2.43	0.65
Northeast Asia	52.38%	0.91	0.21
Oceania	60.96%	1.04	0.26
South Africa	55.74%	1.32	0.23
South America	52.23%	1.02	0.21
South Asia	71.86%	1.8	0.36
Southeast Asia	54.27%	0.86	0.22
Southwest Asia	55.18%	1.18	0.22
West Africa	67.90%	1.54	0.31
West Indies	77.19%	1.99	0.44
Average	62.69	1.44	0.33
Standard deviation	16.29	0.63	0.19

Average hit means the average number of epitope hits/HLA combinations recognized by the population, and the pc90 means minimum number of epitope hits/HLA combinations recognized by 90% of the population.



Linker-FDLMAKCLKQKYEIN-GPGPG Linker-QQFRTEKERERQAK-KK Linker-DEEASGADTSGVLPD-KK Linker-ACCQRWYFTFNGAECS-KK Linker-PQSCVVDQTGSAHCVV-KK Linker-IGRPRWKLLFDEIAKY-KK Linker-SGTQQRGRSCDVTSNT-KK Linker-YNRISHAQKFRKGAGK-AAY Linker-DPDSVTPTY-AAY Linker-KLQKLYISK-AAY Linker-KQKAQLRQR-AAY Linker-IPKGKQKAL-AAY Linker-MRPGAPGLW-AAY Linker-VMYRGRCRK-AAY Linker-SPFEESLNY-AAY Linker-QANFPQTWLW-AAY Linker-SVDFSGTFY-AAY Linker-RVSNDNQFLW-AAY Linker-MSDTEEQY-AAY Linker-KEEELVAL-AAY Linker-MITD sequence-Stop codon-3'UTR-Poly(A) tail. The schematic diagram of the vaccine construct is shown in [Figure 3](#), where key components (signal peptide, HTL/

TABLE 2 Predictions of relative ribosome load for 222 literature-curated 5'-untranslated region (5'-UTR) sequences using the RNA-FM embedding combined with convolutional neural network (CNN) model.

5'-UTR	length	predict_mrl
GGGATCTTATTCCACCTTCTGAAGCTTCTGTCTGAACCAAGTTGTAAGGAGA	50	0.985138357
CGTGAAGGCAAAGAGAACACGCTGCAAAAGGCTTTCCAAGAATCCTCGAC	50	0.953278959
GAATTATCAGAAATACTTTATAGTTATCAAAAATTCTAAAGAAAAAGGCC	50	0.953139484
GAGAGATCCAGCCTCTCAAACATCCAGAGAGACCATAGGCTGCTGCA	50	0.927385569
ACACTTTCTTCTGACATAACAGTGTTCAGTACGCAACCTCAAACAGACACC	50	0.927229881
CTGGAGTCTCCGCGGCAGATCTCATATTTGGATTCTGGATATATTATA	50	0.927137017
GGTAGTTCGGATTACTTCTTTAAGTCTCTTTTCTCTTTTTCGCGCAAAA	50	0.922530532
GCAGTTGGGCAGCGGTTTACCTCCATTTTGAGACCAGACAAGTGGACTC	50	0.920656741
ACATTTGCTTCTGACACAAGTGTTCAGTACGCAACCTCAAACAGACACC	50	0.884883046
GAAGCCAACAAGAATTTGAGAAGTGTAAATACCAAGCCTTGAAAGGGACC	50	0.88337332

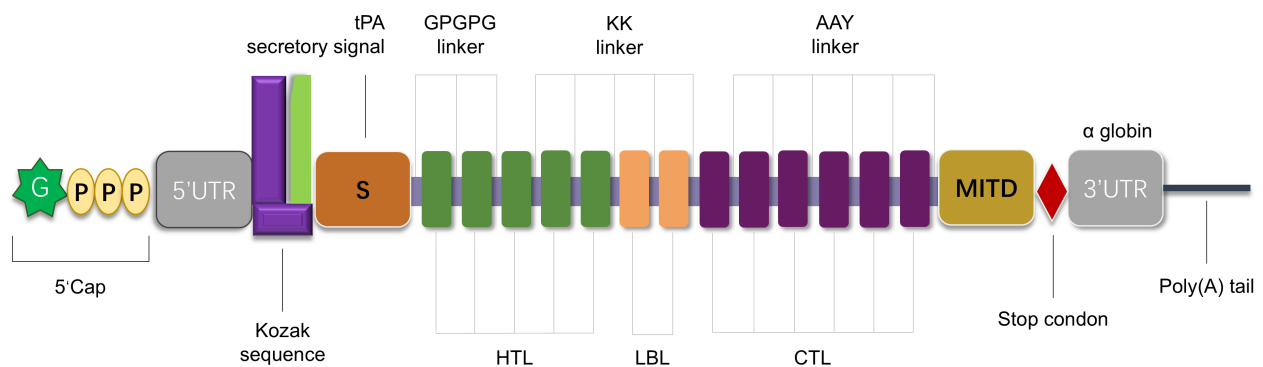


FIGURE 3

Schematic representation of the designed mRNA vaccine construct. The construct is organized from the 5' to 3' terminus, including: 5' m⁷G cap; 5' untranslated region (5'UTR) and Kozak sequence; tissue plasminogen activator (tPA) secretory signal peptide; multiple epitopes [helper T cell epitopes (HTL, green rectangles), potential B cell epitopes (LBL, orange rectangles), and cytotoxic T lymphocyte epitopes (CTL, purple rectangles)] connected by GPGPG linkers, KK linkers, or AAY linkers to ensure appropriate spatial distribution and folding; MITD sequence; stop codon; 3' untranslated region (3'UTR); Poly (A) tail.

LBL/CTL epitope clusters, linkers, and regulatory elements) are color-coded to clarify the 5'→3' architecture.

3.8 Evaluation of antigenicity, allergenicity, toxicity and physicochemical properties of the vaccine construct

Following assembly of the complete mRNA vaccine construct, we systematically evaluated the translated polypeptide sequence through three critical safety parameters: (1) antigenicity prediction using VaxiJen v3.0 (threshold ≥ 0.5 ; tumor antigen model) and ANTIGENpro (SCRATCH suite), (2) allergenicity screening via AllerTop v2.1, and (3) toxicity profiling through ToxinPred. Physicochemical characterization was performed using the ExPASy ProtParam server (<https://web.expasy.org/protparam/>), quantifying six biostability metrics: molecular weight (54603.7), theoretical pI (9.27), instability index (47.94), aliphatic index (60.57), grand average of hydropathicity (GRAVY: -0.776), and thermal stability (*in vitro* half-life >30 hours at 50 °C). As summarized in Table 3, the construct demonstrated strong antigenicity (VaxiJen: 0.7284; ANTIGENpro: 0.93), non-allergenicity, and non-toxicity.

3.9 In silico immune response simulation against the vaccine

Three-dose immunization simulations (1,000 vaccine units per dose) conducted via the C-IMMSIM platform revealed coordinated adaptive immune responses (Figure 4). Primary analysis demonstrated predominant IgM over IgG titers post-initial vaccination, with significant immunoglobulin amplification following booster doses (Figure 4A). Sustained antibody elevation post-antigen clearance suggests established immunological memory, enabling rapid anamnestic responses upon antigen re-

exposure. T lymphocyte profiling showed progressive expansion of both activated and resting helper T cell populations, stabilizing at elevated levels within 30 days post-immunization (Figures 4D, E). Peak T helper cell proliferation occurred at day 10, coinciding with macrophage activation maxima (days 5-10). Resting macrophage populations initially declined (days 0-2) before rebounding through monocyte differentiation (day 3 onward). These data indicate successful antigen presentation and lymphocyte priming. The simulation data collectively confirm the mRNA vaccine's capacity to orchestrate robust humoral and cellular immunity against CRC.

3.10 Codon optimization of the mRNA vaccine construct

To maximize translational efficiency of the mRNA vaccine construct, we implemented a comparative codon optimization strategy employing five algorithms: LinearDesign, JCat, OPTIMIZER, Gensmart, and ExpOptimizer. The optimized nucleotide sequences and corresponding metrics—Codon Adaptation Index (CAI), GC content, and minimum free energy (MFE)—are detailed in Table 4.

All algorithms except LinearDesign achieved comparable CAI values of approximately 0.8 (range: 0.81-0.96), while maintaining optimized GC content within the preferred human codon usage optimum (52.25-67.43%). Among conventional approaches, JCat demonstrated superior performance in both CAI (0.96) and GC content (67.43%). However, LinearDesign exhibited exceptional thermodynamic stability with an MFE of -1193.70 kcal/mol, exceeding other algorithms by more than 2.1-fold. Considering the critical importance of mRNA structural stability *in vivo*, we selected LinearDesign-optimized sequence for further development, thereby ensuring sustained antigen presentation and robust immunogenicity in human hosts.

TABLE 3 Antigenic, allergenic, toxic, and physicochemical assessments of the protein translated from the mRNA vaccine-encoded peptide.

Property	Measurement	Indication
Total number of amino acids	488	Appropriate
Molecular weight	54603.7	Appropriate
Formula	C2393H3758N694O728S22	—
Theoretical pI	9.27	Acidic
Total number of negatively charged residues (Asp + Glu)	52	—
Total number of positively charged residues (Arg + Lys)	70	—
Total number of atoms	7595	—
Instability Index (II)	47.94	Stable
Aliphatic Index (A.I)	60.57	Thermostable
Grand average of hydropathicity (GRAVY)	-0.776	Hydrophilic
estimated half-life	30 hours (mammalian reticulocytes, <i>in vitro</i>)	Stable
Antigenicity (using VaxiJen)	0.7284	Antigenic
Antigenicity (using ANTIGENpro)	0.933854	Antigenic
Allergenicity (using AllerTop 2.0)		Non-allergenic
Toxicity (ToxinPred)		Non-toxic

3.11 Secondary structure validation and stability analysis of mRNA vaccine

The mRNA vaccine construct’s secondary structure was computationally validated through a dual-platform approach: 1) LinearDesign’s constraint programming framework predicted MFE conformations, and 2) RNAfold’s partition function algorithm (ViennaRNA Package 2.6) provided comparative thermodynamic profiling (Figure 4). LinearDesign yielded an MFE of -1193.70 kcal/mol, while RNAfold analysis revealed enhanced stability (MFE = -1233.87 kcal/mol). The subtle differences in the secondary structures of the sequences obtained by the two optimization methods can be seen in Figure 5. This convergence between distinct computational methodologies (<3.4% MFE variance) confirms exceptional structural stability. The robust thermodynamic profile correlates with enhanced vaccine efficacy through three mechanisms: 1) Reduced secondary structure interference with translational initiation complexes, 2) Increased nuclease resistance via compact folding motifs, and 3) Maintenance of epitope codon optimality under physiological temperature fluctuations (37 ± 2°C). These features collectively ensure sustained antigen expression critical for eliciting durable immune responses.

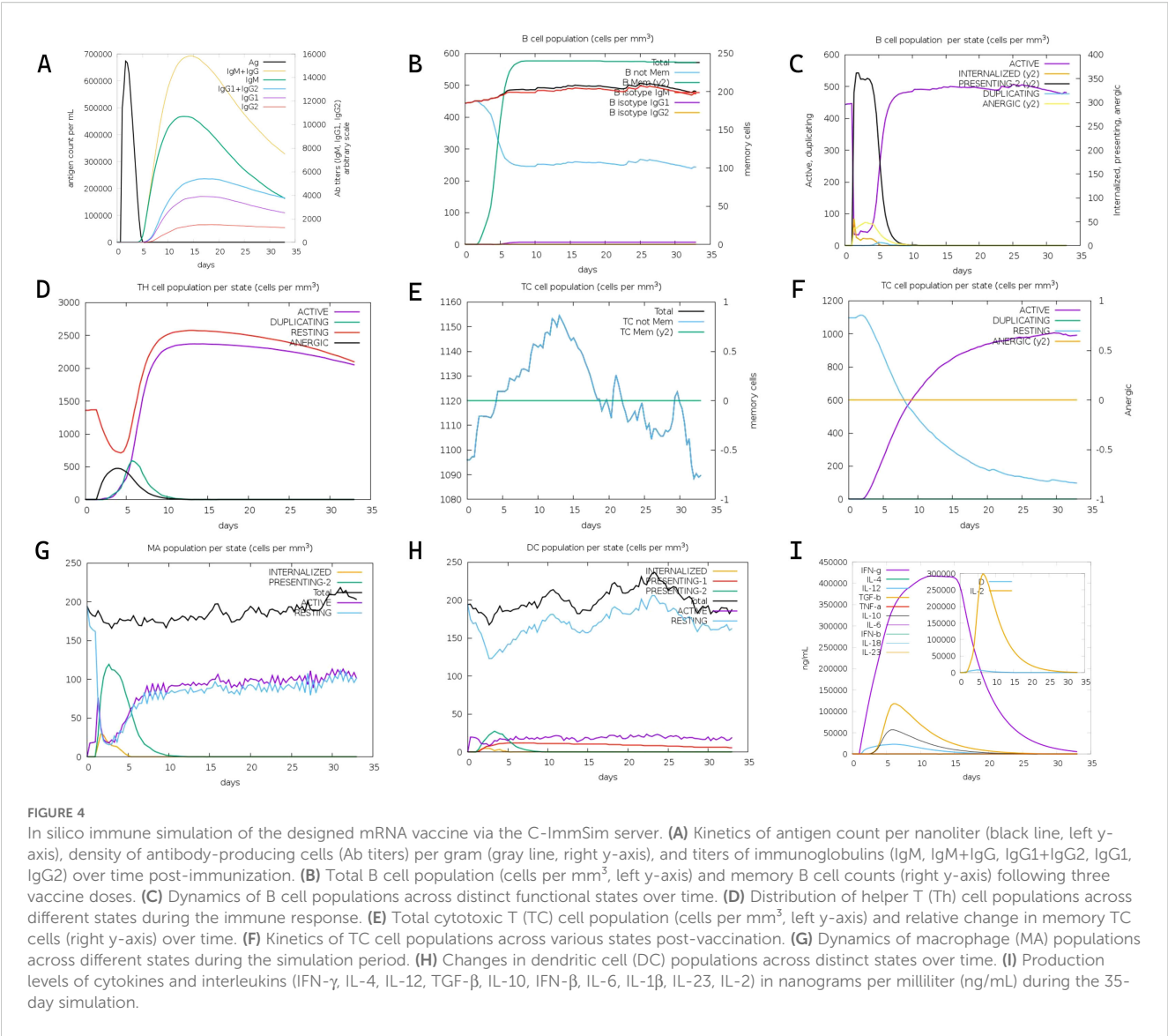
3.12 Molecular docking validation of TLR interaction

Protein-peptide docking between the vaccine-encoded epitope structure and human Toll-like receptor 3 (TLR3; PDB ID 1ZIW) and Toll-like receptor 4 (TLR4; PDB ID 3FXI) was performed using

ClusPro 2.0 (<https://cluspro.org/>), the peptide sequences produced after the translation of the mRNA vaccine are predicted by AlphaFold2 (40). with the highest-ranked TLR3-vaccine complex demonstrating a binding energy of -1378.8 kcal/mol and the highest-ranked TLR4-vaccine complex’s binding energy of -1581.2 kcal/mol (Supplementary Table S4, Figure 6). These results validate the vaccine construct’s capacity to engage TLR3 and TLR4-mediated endosomal sensing pathways, initiating robust dendritic cell maturation and subsequent Th1-polarized adaptive immunity—critical features for eliciting durable antitumor responses in colorectal cancer.

4 Discussion

Colorectal cancer (CRC), defined as cancer of the colon or rectum, has emerged as a persistent global health burden over the past four decades, ranking as the third most prevalent neoplastic disease worldwide (41). According to GLOBOCAN 2020 estimates by the World Health Organization’s International Agency for Research on Cancer (IARC), CRC accounts for approximately 1.93 million new cases and 935,000 deaths annually, with disproportionate incidence rates in industrialized nations (42). The clinical success of Sipuleucel-T, an autologous dendritic cell vaccine approved by the FDA in 2010 for metastatic castration-resistant prostate cancer, demonstrated the therapeutic potential of cancer vaccines and catalyzed paradigm-shifting advancements in tumor immunology (43). Cancer vaccines predominantly utilize tumor-associated antigens (TAAs) and tumor-specific antigens (TSAs) to activate antitumor immunity, enabling targeted elimination of malignant cells through immunological surveillance mechanisms (44). In this study, we engineered a multi-



epitope mRNA vaccine targeting CRC by integrating six TSAs with validated oncogenic roles.

Cytotoxic CD8+ T cells play a pivotal role in tumor eradication, with clinical studies demonstrating that tumor-infiltrating CD8+ T

cell density correlates with improved survival outcomes (45). Our vaccine strategy therefore prioritizes sustained activation of tumor-specific CD8+ T cells through dual mechanisms: a) Circumvention of T cell exhaustion via epitope optimization, and b) Synergistic engagement of B cell-mediated antitumor responses. Emerging evidence confirms B cells' multifaceted roles in cancer immunity, including cytokine-mediated enhancement of CD8+ T cell cytotoxicity, granzyme B secretion for direct tumor lysis, and antibody-dependent cellular phagocytosis (46). Following epitope selection, molecular docking simulations validated robust binding affinities between prioritized epitopes and their cognate HLA alleles (Supplementary Table S2). The endogenous antigen processing pathway ensures cytosolic peptides complex with HLA class I molecules for CD8+ T cell recognition, a mechanism critical for eliminating malignant cells (47). Our docking results demonstrated exceptional interface stability ($I_{sc} < -100$ kcal/mol), with key epitopes like MSDTEEQY achieving complete accommodation within HLA-A*01:01's binding groove (Figure 1F).

TABLE 4 Comparison of five codon optimization methods for mRNA vaccine sequences: Evaluation via CAI (Codon Adaptation Index, reflects translation efficiency), GC content (impacts mRNA stability and immunogenicity), and mRNA Secondary Structure MFE(minimum free energy, indicates mRNA secondary structure stability).

Optimization method	CAI	GC	MFE
Jcat+RNAFold	0.96	67.43%	-651.20 kcal/mol
Gensmart+RNAFold	0.92	57.68%	-596.90 kcal/mol
LinearDesign	0.75	57%	-1193.70 kcal/mol
ExpOptimizer+RNAFold	0.81	52.25%	-564.30 kcal/mol
Vectorbuilder+RNAFold	0.92	56.88%	-617.10 kcal/mol

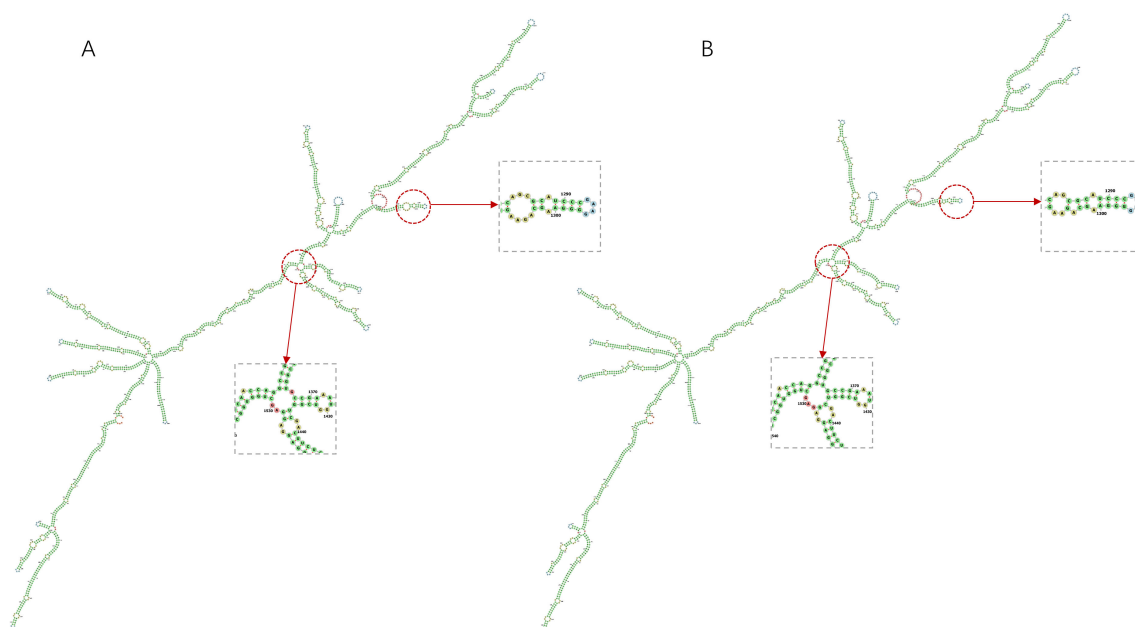


FIGURE 5

Prediction of the secondary structures of mRNA vaccine construct by two methods and their subtle differences in structural composition. **(A)** The secondary structure of the mRNA vaccine construct predicted by LinearDesign. **(B)** The secondary structure of the mRNA vaccine construct predicted by RNAfold.

Population coverage analysis revealed geographical immunogenic efficacy disparities, with peak coverage in Europe (89.13%) and North America (84.69%) versus reduced efficacy in Central America (15.29%) (Table 1). This heterogeneity reflects regional HLA allele distribution patterns, necessitating future iterations incorporating population-specific HLA haplotypes for global applicability.

Leveraging the degeneracy of the genetic code, we performed a systematic comparative analysis of five codon optimization algorithms to maximize translational efficiency and structural stability. Optimization prioritized the minimum free energy (MFE) of mRNA secondary structures as the critical stability

determinant, with LinearDesign achieving superior performance (MFE = -1193.70 kcal/mol) through constraint-based programming (Table 4). Post-assembly bioinformatic characterization confirmed the vaccine construct's antigenicity (VaxiJen score: 0.7284), non-allergenicity, and thermostability (instability index: 47.94), while C-IMMSIM immune simulations predicted robust lymphocyte activation (Figure 3). Molecular docking with Toll-like receptor 3 demonstrated exceptional binding affinity (-1378.8 kcal/mol, Supplementary Table S4).

Although this study has completed the design of the CRC vaccine, two points still need further research and discussion: 1) Will the antigen ordering within the vaccine construct affect the

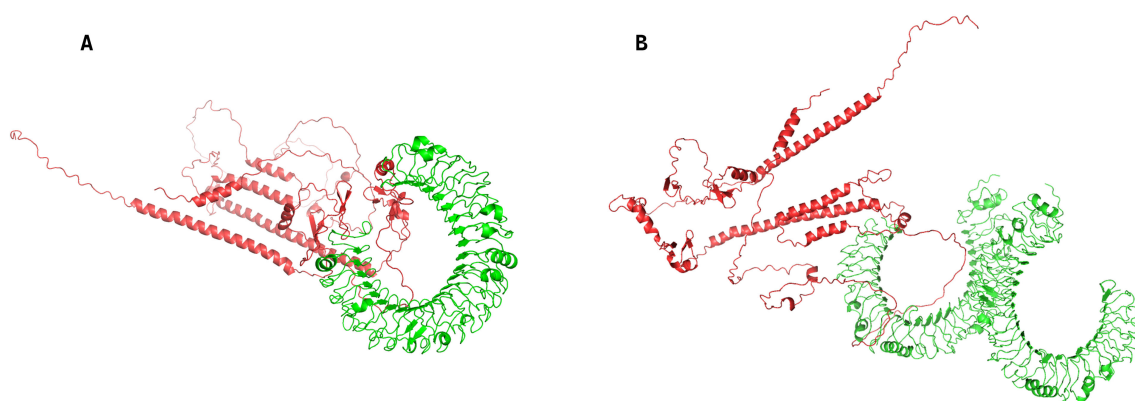


FIGURE 6

Protein-peptide docking complexes of the vaccine-encoded epitope with human Toll-like receptors, generated via ClusPro 2.0. **(A)** Docked complex between the vaccine-encoded peptide and Toll-like receptor 3 (TLR3; PDB ID 1ZIW). **(B)** Docked complex between the vaccine-encoded peptide and Toll-like receptor 4 (TLR4; PDB ID 3FXI).

vaccine's efficacy? 2) Can immune simulations outcomes accurately predict *in vivo* immune responses?

Among these questions, the potential impact of antigen ordering on vaccine efficacy warrants in-depth exploration, especially, for multi-antigen mRNA vaccines against CRC, where the sequence arrangement of individual antigens within the mRNA construct not only affects core antigen properties like expression and folding but also directly shapes immunogenicity and overall vaccine performance. Firstly, antigen order may influence its immunogenicity. Elena et al. (48) designed a combined influenza/COVID-19 mRNA vaccine. When the SARS-CoV-2 RBD antigen was linked to the N-terminus of the trivalent influenza antigens, the immune responses against influenza B and H3N2 antigens decreased significantly. In contrast, when the SARS-CoV-2 RBD antigen was positioned at the C-terminus, the immune responses against all influenza antigens were comparable to those induced by the trivalent influenza antigens alone. This finding indicates that antigen order may affect protein folding or antigen accessibility. Secondly, antigen order may impact its expression efficiency and cellular localization. Fang et al. (49) developed a monkeypox mRNA vaccine. The results showed significant differences in the cell surface display rates of different antigens. This suggests that the position of an antigen in the tandem structure, together with its inherent structural characteristics, collectively determines its final localization. In turn, this may affect the intensity of B cell receptor recognition and antibody responses. Interestingly, Fang et al. proposed a modular vaccine platform (MVP) in another study (50) to enhance the immunogenicity of antigens in mRNA vaccines. In that study, they pointed out that the immune responses to individual antigens in multi-antigen mixed vaccines were relatively independent, with no obvious inter-antigen interference observed. Nevertheless, the inherent cell surface trafficking capacity of different antigens varied significantly. Therefore, the order of antigens in the vaccine is less important than whether the antigens are appropriately modified. Currently, we are developing the second version of the mRNA vaccine design workflow, and based on the above conclusions, we believe that it is highly necessary to conduct further analysis and discussion on the sequence arrangement of antigen epitopes in current vaccines.

Regarding the second question, numerous previous studies have explored the reliability of simulated immune outcomes. For instance, Pappalardo et al. (51) simulated the immunoprophylactic effect of the Triplex vaccine against breast tumors in HER-2/neu transgenic mice. The model successfully reproduced the tumor-free survival curves of mice under four vaccination regimens (early, late, very late, and chronic). Additionally, the simulation results indicated that antibody (Ab) responses played a dominant role in controlling tumor growth, particularly in the long-term phase post-vaccination—this aligns with the conclusion from *in vivo* experiments that Th1-type antibody responses (e.g., IgG2a) are critical for long-term protection. Another study by Bonin et al. (52) simulated and validated the human immune response to the yellow fever vaccine. The model was highly consistent with clinical data in terms of antibody responses and viremia. Simulation results revealed that viral load peaked on day 5 post-vaccination, then decreased rapidly, and dropped to

undetectable levels after 10 days—consistent with clinical observations. Similarly, in a study by Fan et al. (53), immune simulation was performed using C-IMMSIM platform as us. The results showed that the changes in B-cell and T-cell counts were highly consistent with the splenocyte subset proportions and antibody levels observed in *in vivo* experiments, demonstrating that immune simulation can effectively predict the activation trends of actual immune cells. The predicted increases in interferon-gamma (IFN- γ) and interleukin-4 (IL-4) were validated both *in vitro* and *in vivo*.

Although immune simulation exhibits a strong correlation with *in vivo* immune responses, immune simulation still faces numerous challenges in the translation to *in vivo* experiments. Firstly, immune simulation typically relies on existing knowledge, assumptions, and incorporate empirical datasets; however, these datasets may be biased or incomplete. Secondly, the human immune system is an extremely complex and multi-scale network, involving interactions at the molecular, cellular, tissue, organ, and population levels (54). These complex mechanisms are difficult to be fully captured and reproduced in a single simulation (55). Finally, there may be inter-individual differences in the immunogenicity (56).

To verify the reliability of our designed vaccine, we propose that subsequent experimental validations should include the following: conducting cellular uptake assays to evaluate the transfection efficiency of mRNA; performing protein expression assays to assess whether mRNA can be successfully translated into target antigens within cells; detecting the activation markers of CD4+ and CD8+ T cells via flow cytometry to evaluate the vaccine's immunogenicity; and finally, evaluating whether the vaccine can inhibit tumor growth in mouse models.

5 Conclusion

In conclusion, the design of a novel multi-epitope mRNA vaccine for CRC in this study offers a promising framework for advancing research in CRC therapy.

However, it is crucial to note that further *in vitro* and *in vivo* studies are essential to confirm the findings of this study. These additional studies will be necessary to evaluate the vaccine's safety, efficacy, and potential limitations in real-world scenarios. The high predicted antigenicity, the ability to interact with immune receptors, and the stable structure of the proposed vaccine suggest that it may be a promising approach to combat CRC. Overall, this study highlights the potential of *in silico* approaches for vaccine design and provides valuable insights into the development of effective vaccines against CRC.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author/s.

Author contributions

LW: Conceptualization, Data curation, Formal Analysis, Investigation, Writing – original draft, Writing – review & editing. XZ: Supervision, Writing – review & editing. YW: Supervision, Writing – review & editing. JL: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This study was supported by Tianjin Science and Technology Plan 24ZXZSS00020.

Conflict of interest

Authors XZ and YW were employed by Tianjin BioAI-global Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Davidson KW, Barry MJ, Mangione CM, Cabana M, Caughey AB, Davis EM, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *Jama*. (2021) 325:1965–77. doi: 10.1001/jama.2021.6238
- Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag C, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. (2023) 72:338–44. doi: 10.1136/gutjnl-2022-327736
- Spaander MC, Zauber AG, Syngal S, Blaser MJ, Sung JJ, You YN, et al. Young-onset colorectal cancer. *Nat Rev Dis primers*. (2023) 9:21. doi: 10.1038/s41572-023-00432-7
- Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodríguez Yoldi MJ. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *Int J Mol Sci*. (2017) 18:197. doi: 10.3390/ijms18010197
- Moertel CG. Chemotherapy for colorectal cancer. *New Engl J Med*. (1994) 330:1136–42. doi: 10.1056/NEJM199404213301608
- Häfner MF, Debus J. Radiotherapy for colorectal cancer: current standards and future perspectives. *Visceral Med*. (2016) 32:172–7. doi: 10.1159/000446486
- Tol J, Koopman M, Cats A, Rodenburg CJ, Creemers GJ, Schrama JG, et al. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. *New Engl J Med*. (2009) 360:563–72. doi: 10.1056/NEJMoa0808268
- Sebag-Montefiore D, Stephens RJ, Steele R, Monson J, Grieve R, Khanna S, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. *Lancet*. (2009) 373:811–20. doi: 10.1016/S0140-6736(09)60484-0
- Schuster M, Nechansky A, Kircheis R. Cancer immunotherapy. *Biotechnol Journal: Healthcare Nutr Technology*. (2006) 1:138–47. doi: 10.1002/biot.200500044
- Saxena M, van der Burg SH, Melief CJ, Bhardwaj N. Therapeutic cancer vaccines. *Nat Rev Cancer*. (2021) 21:360–78. doi: 10.1038/s41568-021-00346-0
- Slingluff CL Jr. The present and future of peptide vaccines for cancer: single or multiple, long or short, alone or in combination? *Cancer J*. (2011) 17:343–50. doi: 10.1097/PPO.0b013e318233e5b2
- Lee KP, Raez LE, Podack ER. Heat shock protein-based cancer vaccines. *Hematology/Oncology Clinics*. (2006) 20:637–59. doi: 10.1016/j.hoc.2006.02.007

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1649091/full#supplementary-material>

- De Gruijl TD, Van Den Eertwegh AJ, Pinedo HM, Scheper RJ. Whole-cell cancer vaccination: from autologous to allogeneic tumor-and dendritic cell-based vaccines. *Cancer Immunology Immunother*. (2008) 57:1569–77. doi: 10.1007/s00262-008-0536-z
- Melief CJ. Cancer immunotherapy by dendritic cells. *Immunity*. (2008) 29:372–83. doi: 10.1016/j.immuni.2008.08.004
- Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines—a new era in vaccinology. *Nat Rev Drug discovery*. (2018) 17:261–79. doi: 10.1038/nrd.2017.243
- Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. (2017) 547:222–6. doi: 10.1038/nature23003
- Sahin U, Karikó K, Türeci Ö. mRNA-based therapeutics—developing a new class of drugs. *Nat Rev Drug discovery*. (2014) 13:759–80. doi: 10.1038/nrd4278
- Liu C, Papukashvili D, Dong Y, Wang X, Hu X, Yang N, et al. Identification of tumor antigens and design of mRNA vaccine for colorectal cancer based on the immune subtype. *Front Cell Dev Biol*. (2022) 9:783527. doi: 10.3389/fcell.2021.783527
- Fei W, Chen L, Chen J, Shi Q, Zhang L, Liu S. RBP4 and THBS2 are serum biomarkers for diagnosis of colorectal cancer. *Oncotarget*. (2017) 8:92254–64. doi: 10.18632/oncotarget.21173
- Karagiannis GS, Berk A, Dimitromanolakis A, Diamandis EP. Enrichment map profiling of the cancer invasion front suggests regulation of colorectal cancer progression by the bone morphogenetic protein antagonist, gremlin-1. *Mol Oncol*. (2013) 7:826–39. doi: 10.1016/j.molonc.2013.04.002
- Hao Y-H, Yu S-Y, Tu R-S, Cai Y-Q. TNNT1, a prognostic indicator in colon adenocarcinoma, regulates cell behaviors and mediates EMT process. *Bioscience Biotechnology Biochem*. (2020) 84:111–7. doi: 10.1080/09168451.2019.1664891
- Ma Y-S, Huang T, Zhong X-M, Zhang H-W, Cong X-L, Xu H, et al. Proteogenomic characterization and comprehensive integrative genomic analysis of human colorectal cancer liver metastasis. *Mol cancer*. (2018) 17:139. doi: 10.1186/s12943-018-0890-1
- Xing X, Gu X, Ma T. Knockdown of biglycan expression by RNA interference inhibits the proliferation and invasion of, and induces apoptosis in, the HCT116 colon cancer cell line. *Mol Med Rep*. (2015) 12:7538–44. doi: 10.3892/mmr.2015.4383

24. Yang X-M, You H-Y, Li Q, Ma H, Wang Y-H, Zhang Y-L, et al. *CTHRC1* promotes human colorectal cancer cell proliferation and invasiveness by activating Wnt/PCP signaling. *Int J Clin Exp pathology*. (2015) 8:12793.
25. Lin X-L, Yang L, Fu S-W, Lin W-F, Gao Y-J, Chen H-Y, et al. Overexpression of *NOX4* predicts poor prognosis and promotes tumor progression in human colorectal cancer. *Oncotarget*. (2017) 8:33586. doi: 10.18632/oncotarget.16829
26. Naveed M, Jabeen K, Naz R, Mughal MS, Rabaan AA, Bakhrebah MA, et al. Regulation of host immune response against *Enterobacter cloacae* proteins via computational mRNA vaccine design through transcriptional modification. *Microorganisms*. (2022) 10:1621. doi: 10.3390/microorganisms10081621
27. Naveed M, J-u H, Ahmad M, Naeem N, MS M, AA R, et al. Designing mRNA-and peptide-based vaccine construct against emerging multidrug-resistant *Citrobacter freundii*: a computational-based subtractive proteomics approach. *Medicina*. (2022) 58:1356. doi: 10.3390/medicina58101356
28. DeLano WL. Pymol: An open-source molecular graphics tool. *CCP4 News Protein Crystallogr*. (2002) 40:82–92.
29. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. *Methods Enzymology*. (2004) 383:66–93. doi: 10.1016/S0076-6879(04)83004-0
30. Leppke K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol*. (2018) 19:158–74. doi: 10.1038/nrm.2017.103
31. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science*. (2016) 352:1413–6. doi: 10.1126/science.aad9868
32. Chen J, Hu Z, Sun S, Tan Q, Wang Y, Yu Q, et al. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. *arXiv preprint arXiv*. (2022) 2204:00300. doi: 10.1101/2022.08.06.503062
33. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proc 2019 Conf North Am chapter Assoc Comput linguistics: Hum Lang Technol*. (2019) 1:4171–86. doi: 10.18653/v1/N19-1423
34. Chu Y, Yu D, Li Y, Huang K, Shen Y, Cong L, et al. A 5' UTR language model for decoding untranslated regions of mRNA and function predictions. *Nat Mach Intelligence*. (2024) 6:449–60. doi: 10.1038/s42256-024-00823-9
35. Li T, Liu G, Bu G, Xu Y, He C, Zhao G. Optimizing mRNA translation efficiency through rational 5' UTR and 3' UTR combinatorial design. *Gene*. (2025) 942:149254. doi: 10.1016/j.gene.2025.149254
36. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New Engl J Med*. (2020) 383:2603–15. doi: 10.1056/NEJMoa2034577
37. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. (1986) 44:283–92. doi: 10.1016/0092-8674(86)90762-2
38. Zhang H, Zhang L, Lin A, Xu C, Li Z, Liu K, et al. Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature*. (2023) 621:396–403. doi: 10.1038/s41586-023-06127-z
39. McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers: original Res biomolecules*. (1990) 29:1105–19. doi: 10.1002/bip.360290621
40. Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun*. (2022) 13:1265. doi: 10.1038/s41467-022-28865-w
41. Lewandowska A, Rudzki G, Lewandowski T, Strykowska-Góra A, Rudzki S. Risk factors for the diagnosis of colorectal cancer. *Cancer control*. (2022) 29:10732748211056692. doi: 10.1177/10732748211056692
42. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660
43. Gardner T, Elzey B, Hahn NM. Sipuleucel-T (Provenge) autologous vaccine approved for treatment of men with asymptomatic or minimally symptomatic castrate-resistant metastatic prostate cancer. *Hum Vaccines immunotherapeutics*. (2012) 8:534–9. doi: 10.4161/hv.19795
44. Miao L, Zhang Y, Huang L. mRNA vaccine for cancer immunotherapy. *Mol cancer*. (2021) 20:41. doi: 10.1186/s12943-021-01335-5
45. Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. *Br J cancer*. (2021) 124:359–67. doi: 10.1038/s41416-020-01048-4
46. Tsou P, Katayama H, Ostrin EJ, Hanash SM. The emerging role of B cells in tumor immunity. *Cancer Res*. (2016) 76:5597–601. doi: 10.1158/0008-5472.CAN-16-0431
47. Rock KL, Reits E, Neefjes J. Present yourself! By MHC class I and MHC class II molecules. *Trends Immunol*. (2016) 37:724–37. doi: 10.1016/j.it.2016.08.010
48. Thornhill-Wadolowski E, Ruter DL, Yan F, Gajera M, Kurt E, Samanta L, et al. Development of an influenza/COVID-19 combination mRNA vaccine containing a novel multivalent antigen design that enhances immunogenicity of influenza virus B hemagglutinins. *Vaccines*. (2025) 13:628. doi: 10.3390/vaccines13060628
49. Fang Z, Monteiro VS, Renauer PA, Shang X, Suzuki K, Ling X, et al. Polyvalent mRNA vaccination elicited potent immune response to monkeypox virus surface antigens. *Cell Res*. (2023) 33:407–10. doi: 10.1038/s41422-023-00792-5
50. Fang Z, Monteiro VS, Oh C, Janabi KA, Romero L, Ahsan N, et al. A modular vaccine platform for optimized lipid nanoparticle mRNA immunogenicity. *Nat Biomed Engineering*. (2025), 1–16. doi: 10.1038/s41551-025-01478-6
51. Pappalardo F, Lollini P-L, Castiglione F, Motta S. Modeling and simulation of cancer immunoprevention vaccine. *Bioinformatics*. (2005) 21:2891–7. doi: 10.1093/bioinformatics/bti426
52. Bonin CR, Fernandes GC, Dos Santos RW, Lobosco M. A qualitatively validated mathematical-computational model of the immune response to the yellow fever vaccine. *BMC Immunol*. (2018) 19:15. doi: 10.1186/s12865-018-0252-1
53. Fan Y, He Y, Li Y, Yin Z, Shi J, Tian T, et al. Design of a novel EmTSP-3 and EmTIP based multi-epitope vaccine against *Echinococcus multilocularis* infection. *Front Immunol*. (2024) 15:1425603. doi: 10.3389/fimmu.2024.1425603
54. Kwiatkowska M, Urban J, Dagasso G. Context specification in the computational modelling of human immune system response to viral infections. *Proc Comput Science*. (2021) 185:330–7. doi: 10.1016/j.procs.2021.05.016
55. Li J, Lee Y, Li Y, Jiang Y, Lu H, Zhang W, et al. Co-inhibitory molecule B7 superfamily member 1 expressed by tumor-infiltrating myeloid cells induces dysfunction of anti-tumor CD8+ T cells. *Immunity*. (2018) 48:773–86. e5. doi: 10.1016/j.immuni.2018.03.018
56. Wei H-H, Zheng L, Wang Z. mRNA therapeutics: new vaccination and beyond. *Fundam Res*. (2023) 3:749–59. doi: 10.1016/j.fmre.2023.02.022