Computational biology and artificial intelligence in mRNA vaccine design for cancer immunotherapy

Saber Imani[1], Xiaoyan Li[1], Keyi Chen[2], Mazaher Maghsoudloo[3], Parham Jabbarzadeh Kaboli[4], Mehrdad Hashemi[5,6], Saloomeh Khoushab[5,6] and Xiaoping Li[2]*

[1]Shulan International Medical College, Zhejiang Shuren University, Hangzhou, Zhejiang, China, [2]Key Laboratory of Artificial Organs and Computational Medicine in Zhejiang Province, Shulan International Medical College, Zhejiang Shuren University, Hangzhou, Zhejiang, China, [3]Key Laboratory of Epigenetics and Oncology, the Research Center for Preclinical Medicine, Southwest Medical University, Luzhou, Sichuan, China, [4]Department of Biochemistry, Faculty of Medicine, Medical University of Warsaw, Warsaw, Poland, [5]Department of Genetics, Faculty of Advanced Science and Technology, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran, [6]Farhikhtegan Medical Convergence sciences Research Center, Farhikhtegan Hospital Tehran Medical sciences, Islamic Azad University, Tehran, Iran

Messenger RNA (mRNA) vaccines offer an adaptable and scalable platform for cancer immunotherapy, requiring optimal design to elicit a robust and targeted immune response. Recent advancements in bioinformatics and artificial intelligence (AI) have significantly enhanced the design, prediction, and optimization of mRNA vaccines. This paper reviews technologies that streamline mRNA vaccine development, from genomic sequencing to lipid nanoparticle (LNP) formulation. We discuss how accurate predictions of neoantigen structures guide the design of mRNA sequences that effectively target immune and cancer cells. Furthermore, we examine AI-driven approaches that optimize mRNA-LNP formulations, enhancing delivery and stability. These technological innovations not only improve vaccine design but also enhance pharmacokinetics and pharmacodynamics, offering promising avenues for personalized cancer immunotherapy.

1 Introduction

Messenger RNA (mRNA) vaccines have emerged as a rapid, flexible, and scalable strategy in cancer immunology. This innovative method elicits a robust and targeted immune response (Lorentzen et al., 2022; Yao et al., 2024). The effectiveness of mRNA vaccines during the COVID-19 pandemic has underscored their potential in addressing

infectious diseases (Chakraborty et al., 2021; Pennisi et al., 2024). However, moving from concept to clinical implementation involves navigating significant scientific and technical challenges, necessitating a comprehensive, interdisciplinary approach (Lorentzen et al., 2022; Sayour et al., 2024). mRNA vaccines in oncology are considered personalized, representing a key advance in precision medicine by targeting the unique genetic mutations in an individual's tumor cells (Lorentzen et al., 2022). By crafting a vaccine that targets these specific anomalies, this personalized method seeks to elicit a precise immune response, minimizing off-target effects and significantly enhancing therapeutic outcomes (May, 2024).

Unlike traditional vaccines, which use inactivated or attenuated pathogenic proteins, mRNA vaccines deliver tumor-associated antigens (TAAs) or neoantigens directly to antigen-presenting cells (APCs) like dendritic cells (DCs) or macrophages. After the tumor antigen is presented on the surface of the APCs, a cascade of immune responses is triggered, initiating adaptive immunity (Pardi et al., 2018; Gote et al., 2023). These neoantigens are processed and displayed on the cell surface via major histocompatibility complex (MHC) class I molecules, allowing the immune system to recognize the tumor proteins as foreign and triggering an immune response (Esprit et al., 2020). The primary immune response involves cytotoxic T lymphocytes (CTLs), which recognize and eliminate cancer cells expressing specific tumor antigens (Vishweshwaraiah and Dokholyan, 2022). Additionally, APC activation stimulates CD4+ T helper 1 (TH1) cells, which release cytokines to boost CTL activity and recruit macrophages, creating an immune-reactive tumor microenvironment (TME) (Li et al., 2022b; Ramirez et al., 2023). By enhancing the infiltration of immune cells, such as CTLs and macrophages, and overcoming immune checkpoint inhibition, mRNA vaccines can help shift the balance in favor of anti-tumor immunity. This reprogramming of the TME supports a more effective and sustained immune response against cancer cells, ultimately improving the overall efficacy of cancer immunotherapy (Gote et al., 2023; Ramirez et al., 2023).

In the case of naked mRNA vaccines, the mRNA is delivered directly into the body without any protective carrier. Once administered, the naked mRNA is taken up by cells, including DCs, through endocytosis or direct membrane fusion (Hasan et al., 2023). After entering the cytoplasm, the mRNA is translated into the target tumor antigen, which is processed and presented on MHC class I molecules, stimulating a robust immune response, specifically activating CTLs that target and destroy tumor cells expressing the same antigen. However, naked mRNA has some limitations, particularly in terms of stability and delivery efficiency (Abbasi et al., 2024).

To overcome these challenges, lipid nanoparticles (LNPs)-encapsulated mRNA are commonly used, and they are the only FDA-approved delivery vehicles for mRNA vaccines (Igyártó and Qin, 2024). LNPs are designed to encapsulate the mRNA, protecting it from degradation and improving its stability in the bloodstream. They also facilitate the efficient delivery of mRNA into target cells. Once inside the cell, the mRNA is released from the LNPs and enters the cytoplasm, where translation occurs, leading to the production of tumor antigens. LNPs are especially advantageous for improving cellular uptake. They interact with the cell membrane, facilitating endocytosis and ensuring that the mRNA is delivered into cells in a controlled manner. Once inside, the mRNA is translated into the antigen, processed, and presented by APCs on MHC class I molecules, leading to the activation of CTLs and the initiation of a strong anti-tumor immune response (Imani et al., 2024). By using LNPs, the delivery of mRNA vaccines becomes more efficient, enhancing both the stability of the mRNA and the ability of APCs to initiate a targeted immune response (Alameh et al., 2021; Shuptrine et al., 2024).
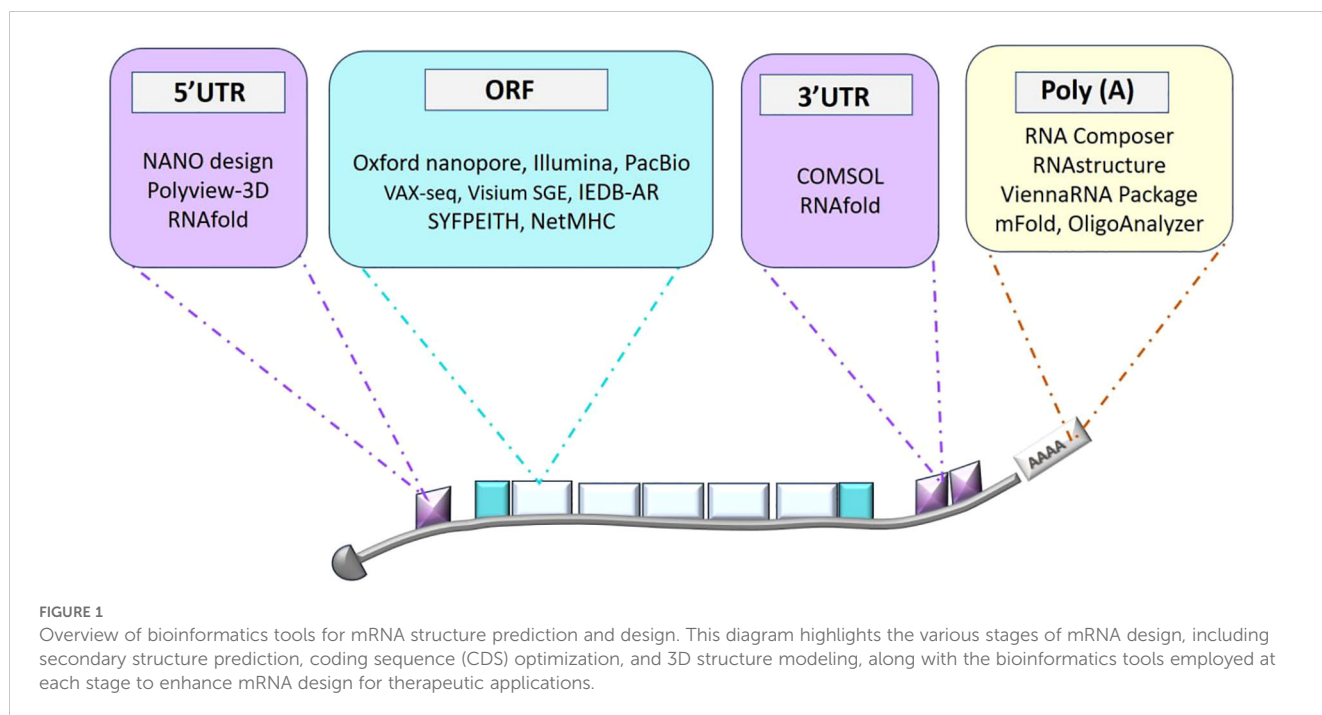
The development of personalized mRNA vaccines involves several crucial steps, each supported by advanced bioinformatics tools. Initially, next-generation sequencing (NGS) is used to analyze the genome of the pathogen or tumor, identifying unique mutations and neoantigens (Alburquerque-González et al., 2022; Al Fayez et al., 2023). Comprehensive genetic data is crucial for designing mRNA vaccines. Tools like NetMHCpan and the Immune Epitope Database (IEDB) identify the most immunogenic HLA-I and MHC class I epitopes to trigger a strong T-cell response (Kim et al., 2012; Cai et al., 2021). To enhance stability and efficiency, RNAfold and mfold predict the mRNA's secondary structure, reducing degradation and improving effectiveness (Chen and Chan, 2023). LNP formulation tools, such as NanoAssembler, optimize delivery by protecting the mRNA and aiding its entry into host cells for effective antigen expression (Wang et al., 2022).

On the other hand, machine learning algorithms further refine these predictions by analyzing extensive immunological data. Incorporating machine learning and AI into this process is vital. Algorithms like Random Forest, Support Vector Machines (SVMs), and Convolutional Neural Networks (CNNs) analyze large datasets to predict vaccine efficacy and potential side effects. These AI-driven insights help optimize vaccine design, enhancing efficacy and safety (Bravi, 2024).

While advancements in bioinformatics and AI are significant, comprehensive comparative studies in this field are lacking, which limits our understanding of their full potential. This paper explores the role of these technologies in developing personalized mRNA vaccines, focusing on genome sequencing, epitope prediction, RNA structure analysis, and LNP formulation. We discuss the challenges, insights, and future directions, highlighting how AI improves vaccine development by analyzing data, identifying patterns, and optimizing design to predict side effects and enhance effectiveness. This paper aims to address current knowledge gaps and encourages further research in oncology and immunology, where personalized mRNA vaccines have the potential to transform cancer treatment.

2 Sequencing and initial data acquisition

Figure 1 presents a schematic overview of bioinformatics tools for mRNA structure prediction and design, covering methods for secondary structure prediction, coding sequences (CDS) optimization, and 3D modeling. Sequencing and initial data acquisition are fundamental steps in developing mRNA vaccines, providing essential genetic information about target viruses, and

FIGURE 1
Overview of bioinformatics tools for mRNA structure prediction and design. This diagram highlights the various stages of mRNA design, including secondary structure prediction, coding sequence (CDS) optimization, and 3D structure modeling, along with the bioinformatics tools employed at each stage to enhance mRNA design for therapeutic applications.

setting the stage for vaccine design and optimization (Gunter et al., 2023). Key sequencing technologies such as Illumina, Oxford Nanopore, and PacBio play crucial roles in this process. Illumina's high-throughput short-read sequencing offers extensive coverage of viral genomes, helping to identify genetic variations that are important for understanding viral diversity and evolution (Lemay et al., 2022). Oxford Nanopore's real-time long-read sequencing provides insights into full-length RNA transcripts and complex genomic regions, which is useful for detecting diverse viral variants and structural features (Stefan et al., 2022). PacBio's high-accuracy long-read sequencing allows for detailed genomic characterization and variant analysis, particularly beneficial for studying RNA viruses. Bioinformatics tools play a crucial role in maintaining the quality, preprocessing, and comprehensive analysis of sequencing data across all three sequencing technologies. FASTQC (Fast Quality Control) assesses key quality metrics like base quality scores and GC content, while Trimmomatic eliminates artifacts and adapter sequences from raw reads (Bolger et al., 2014), thereby improving the accuracy of subsequent analyses. SAMtools manages aligned sequences in Sequence Alignment/Map (SAM) and Binary Alignment/Map (BAM) formats, which is vital for variant calling and in-depth genomic analysis, offering valuable insights for vaccine design. The workflow begins with alignment tools such as the Burrows-Wheeler Aligner (BWA) and Bowtie, which align short-read mRNA sequences to reference genomes or transcriptomes (Rajan-Babu et al., 2021). Also, Visium Spatial Gene Expression (Visium SGE) is an advanced platform that combines spatially resolved transcriptomics with histological imaging to map gene expression within the structural context of tissues, enabling precise insights into cellular activity and tissue architecture (Toyama et al., 2023). These tools are instrumental in identifying conserved regions and potential immunogenic epitopes within the mRNA sequences. Following alignment, assembly algorithms

reconstruct full-length mRNA sequences by integrating sequence overlaps and pairing information, ensuring the integrity and completeness of mRNA constructs for vaccine production.

In practice, during the development of mRNA vaccines for novel viral outbreaks, alignment tools like BWA are used to compare mRNA sequences with known sequences of related viruses. This process helps identify conserved regions critical for vaccine design, ensuring effective targeting of the virus and the induction of protective immune responses in vaccinated individuals. Table 1 compares the main key sequencing technologies, including Illumina, Oxford Nanopore, and PacBio.

2.1 Illumina

Illumina sequencing is a high-throughput technology known for its precision in generating short DNA or RNA sequence reads, which can produce fragments hundreds of bases in length and is vital for mRNA vaccine development. The process starts with fragmenting DNA samples of approximately 300-500 bp or RNA samples of about 200 bp, followed by the attachment of adapters. These fragments are then amplified on a flow cell through bridge amplification, forming clusters of identical sequences. During sequencing, fluorescently labeled nucleotides are incorporated into the growing DNA strands. Techniques like RELIC are used to correct dye bias in Illumina data, ensuring accurate sequencing results (Xu et al., 2017). Illumina can also help length-sequencing platforms such as ONT get high-quality genomes more efficiently (Lerminiaux et al., 2024). Each nucleotide emits a unique color when excited by a laser, and high-resolution cameras capture these colors to determine the nucleotide sequence. This technology is crucial for identifying genetic variations and viral genome features, aiding in the development of effective mRNA vaccines. For

TABLE 1 Comparison of sequencing technologies for mRNA vaccine development.

Feature	Illumina	Oxford Nanopore	PacBio	VAX-seq	Visium SGE
Read length	Short-read	Long-read	Long-read	Short-read to long-read	Short-read to long-read
Tech.	Seq. by Fluor. nucleotides	Nanopore-based detect.	SMRT tech.	Sequencing of mRNA sequences	Spatial transcriptomics
Data	High-throughput	Real-time seq.	High-accuracy	High-throughput	High-throughput
Accuracy	Very high	Moderate	Very high	High	Moderate to high
Tools	BWA, FASTQC	Minimap2, GraphMap	HGAP, SMRT	Custom tools (specific to VAX-seq)	Space Ranger, Seurat
Complexity	Moderate	Lower	High	Moderate	Moderate
Seq. depth	Very high	Moderate	High	High	High
Turnaround	Moderate	Fast	Moderate	Fast	Moderate
Cost	Mod. to high	Moderate	High	Moderate	High
Capabilities	No	Yes	No	Yes	Yes
Suitability	High	High	Moderate	High	High
Advantages	Extensive coverage, high acc.	Real-time seq., long reads, min. lib. prep	High acc., long reads, detailed analysis	High-resolution immune profiling	Spatial resolution, gene expression mapping
Disadvantages	Short reads, complex lib. prep	Lower acc.	Higher cost, complex lib.	Limited to vaccine studies	Limited to tissue samples

Seq., Sequencing, Fluor., Fluorescent, Tech., Technology, Detect., Detection, SMRT, Single Molecule Real-Time, VAX-seq, Vaccination Sequencing, Acc., Accuracy, Min., Minimal, Lib., Library, Prep., Preparation, Genom., Genomic, Mod., Moderate, BWA, Burrows-Wheeler Aligner, BWT, Burrows-Wheeler Transform, FM-index, Full-text Minute-space Index, HGAP, Hierarchical Genome Assembly Process, FASTQC, FAST Quality Control, Trimmomatic, Trimming tool for quality control, Minimap2, Alignment tool for long reads, GraphMap, Alignment tool for long reads; SMRT Link; Software suite for managing SMRT sequencing data; Space Ranger, Software for analyzing Visium spatial gene expression data; Seurat, Tool for analyzing single-cell RNA-seq and spatial data.

processing sequencing data, algorithms like BWA (Guo and Huo, 2024) and Bowtie are essential. BWA uses the Burrows-Wheeler Transform (BWT) (Keel and Snelling, 2018) for efficient sequence alignment, while Bowtie utilizes the FM-index for indexing and searching data (Table 1). Brittney N. Keel's comparison shows that BWA is more robust, whereas HISAT2 is faster and uses less memory than both BWA and Bowtie2 (Keel and Snelling, 2018). Data quality is ensured with preprocessing tools such as FASTQC, which evaluates base quality scores, and Trimmomatic, which removes low-quality bases and adapter sequences to enhance alignment accuracy and variant detection. Detailed mathematical formulations and specific operational details of these methods are available in Supplementary file S1.

2.2 Oxford nanopore

Oxford Nanopore sequencing is a state-of-the-art long-read technology that plays a crucial role in mRNA vaccine development. This method uses nanopore sensors to detect changes in ionic current as nucleic acids pass through a protein nanopore (Xue et al., 2020), such that a negatively charged single-stranded DNA or RNA molecule is driven from the negatively charged "cis" side through the nanopore to the positively charged "trans" side, which is recorded and analyzed to infer the base sequence (Su et al., 2023). Currently, there are eight versions of the system, with R9 achieving an impressive translocation rate of 250 bases per second and R9.4 achieving a translocation rate of 450 bases per second, which is a

significant improvement over R7's 70 bases per second. The other different systems have their advantages (Wang et al., 2021).

For mRNA vaccine development, Oxford Nanopore sequencing has been modified to sequence them directly without reverse transcription. Although the accuracy of direct sequencing of RNA is lower than that of DNA sequencing, about 83% to 86%. Similarly, Oxford Nanopore sequencing also provides direct sequencing of complementary DNA (cDNA) without the need for polymerase chain reaction (PCR) amplification (Wang et al., 2021). This capability is essential for understanding the complete structure and function of RNA, including secondary structures and complex genomic regions vital for designing effective vaccines.

Key bioinformatics tools for Oxford Nanopore sequencing are MinION Knowledge Base (MinKNOW) and Guppy. MinKNOW manages the sequencing device and collects raw data (Oeck et al., 2023), while Guppy performs base-calling to convert the raw signal data into nucleotide sequences (Wick et al., 2019). After base-calling, alignment tools such as Minimap2 are used to map these long reads to reference genomes. During mRNA vaccine development, this technology allows real-time sequencing of viral genomes, aiding in the identification of conserved regions and potential epitopes crucial for effective vaccine design. Although the average accuracy of ONT sequencing is improving, certain subsets of reads or read fragments have very low accuracy, and the error-rate reads of 1D reads and 2D/1D reads are still much higher than the short reads produced by NGS technologies (Wang et al., 2021). Oxford Nanopore sequencing also excels at detecting viral variants by analyzing complete sequences and complex genomic

regions, with the characteristics of short turnaround time and low cost (Xu et al., 2022). This ability to identify mutations and variations is essential for designing mRNA vaccines that elicit strong immune responses against diverse viral strains.

2.3 PacBio

PacBio sequencing, utilizing Single Molecule Real-Time (SMRT) targeting technology that does not require pausing between read steps, so kinetic changes interpreted from light-pulse movies can be analyzed to detect base modifications, such as methylation, and accurate detection and discovery of all variant types, even in hard-to-reach regions of the genome (Rhoads and Au, 2015), has the potential to revolutionize physical health, reproduction, cancer research, as well as microbial and viral genetic testing (Ardui et al., 2018), is crucial for mRNA vaccine development due to its capability to produce highly accurate long-read sequences. This technique involves DNA polymerase synthesizing complementary DNA strands with fluorescently labeled nucleotides. The emitted light from these nucleotides is detected in real-time, enabling immediate base calling. For RNA molecules, PacBio finds novel genes, transcripts, and alternative splicing through a complete view of transcript isoform diversity to sequence them (Rhoads and Au, 2015).

The long-read capability of PacBio sequencing, which can extend up to 60 kb, provides significant advantages in identifying and quantifying subtypes, including novel ones (Rhoads and Au, 2015). According to *Jia H.* et al. findings, this technology allows for low-input library preparation, requiring only 100 ng of DNA for the Sequel system and 400 ng for the Sequel II system (Jia et al., 2024). This is particularly useful for comprehensive viral genome sequencing, including the identification of new variations and genetic mutations in viruses like SARS-CoV-2 (Nicot et al., 2023).

The SMRT Link software suite manages data collection and processing, including base calling and error correction. Algorithms such as the Hierarchical Genome Assembly Process (HGAP) (Chin et al., 2013) and Canu et al. (Prjibelski et al., 2023) address the challenges of assembling long reads by correcting errors and constructing complete genome sequences. HGAP builds consensus sequences from long reads. PacBio sequencing is crucial for identifying conserved regions and potential immunogenic epitopes within viral genomes, which helps in designing effective mRNA vaccines. However, the technology has limitations, including lower throughput with fewer sub-reads or CCS reads and a higher error rate of about 11-15% for CLR reads (Rhoads and Au, 2015).

2.4 VAX-seq

VAX-seq, a novel sequencing technology, plays a pivotal role in advancing the field of mRNA vaccine development. This high-throughput sequencing method is specifically tailored for the identification and quantification of vaccine-induced immune responses (Gunter et al., 2023b). VAX-seq is a specialized

technology focused on sequencing mRNA in the context of immune profiling. Its ability to detect modified nucleosides is limited and primarily inferred through indirect analyses or complementary assays (Gunter et al., 2023). By providing a more detailed understanding of the interactions between mRNA vaccines and the immune system, VAX-seq enables the identification of specific mRNA sequences that contribute to optimal immune activation. This technology allows researchers to profile the genetic composition of mRNA vaccines and their translation products with greater accuracy, improving both the design and efficacy of these vaccines (Gote et al., 2023).

One of the key advantages of VAX-seq over traditional sequencing methods, such as Illumina and Oxford Nanopore, lies in its ability to offer higher-resolution insights into the transcriptome (Gunter et al., 2023). This enables a more comprehensive analysis of vaccine-induced responses, allowing for the detection of rare or subtle immune reactions that might be missed with other methods. The technique enhances the ability to tailor mRNA vaccine sequences to better stimulate desired immune responses, which is crucial for optimizing vaccine formulations for various pathogens, including those that require more precise immune targeting. Incorporating VAX-seq into mRNA vaccine development holds significant potential for both enhancing vaccine design and guiding clinical decision-making. By combining its high sensitivity with the ability to sequence and quantify complex mRNA sequences, VAX-seq aids in the identification of critical sequence motifs and epitopes (Jeeva et al., 2021). This level of detail is essential for the development of more effective mRNA vaccines, capable of eliciting stronger, more targeted immune response, and ultimately providing better protection against infectious diseases (Gunter et al., 2023).

2.5 Visium SGE

Visium SGE by 10x Genomics has emerged as a transformative analytical tool, integrating spatially resolved transcriptomic data with high-resolution tissue histology (Ståhl et al., 2016). This platform allows researchers to map gene expression patterns directly onto histological sections, providing unparalleled insights into the spatial context of mRNA translation and immune cell dynamics within tissues (Toyama et al., 2023). Visium SGE combines spatial transcriptomics with high-throughput short-read sequencing. While it offers spatial resolution and gene expression mapping, its capability to detect modified nucleosides is restricted to indirect bioinformatic inferences (Williams et al., 2022). By combining transcriptomics with histopathological features, Visium enables the identification of specific cell populations and their molecular activities about their precise tissue location. For example, in the context of mRNA vaccine development, Visium can localize mRNA-encoded antigen expression to immune-competent regions, such as lymphoid aggregates, while simultaneously identifying structural changes in surrounding tissue architecture. This dual-layer information is invaluable for validating predictive models like AlphaFold, ensuring that computationally predicted antigens are accurately expressed and situated in biologically

relevant microenvironments (Smith et al., 2024). In mRNA vaccine development, Visium has proven instrumental in refining antigen design and delivery strategies. For instance, studies leveraging Visium have demonstrated its capability to map DCs activity in lymphoid tissues following mRNA-LNP administration, directly linking antigen presentation to CTLs recruitment. In one example, Visium analysis identified specific tissue regions where mRNA vaccines encoding TAAs were translated most efficiently, allowing researchers to pinpoint the spatial co-localization of antigen-expressing cells and CD8+ T-cell activation zones. This spatial information guided the optimization of LNP formulations to ensure antigen delivery to DCs located in lymphoid-rich areas, thereby enhancing CTL priming and overall vaccine efficacy (Melo Ferreira et al., 2021; Hudson and Sudmeier, 2022).

Moreover, Visium facilitates the identification of off-target effects and unintended mRNA expression in non-target tissues, a critical consideration in vaccine safety profiling. For example, spatial transcriptomic analysis using Visium uncovered ectopic expression of mRNA constructs in hepatocytes during preclinical studies, revealing suboptimal LNP biodistribution. Based on these findings, LNP formulations were redesigned to incorporate specific targeting ligands that preferentially deliver mRNA to DCs while minimizing liver uptake (Ståhl et al., 2016). This iterative approach underscores the power of Visium in bridging computational predictions with experimental outcomes, ensuring the spatial fidelity of mRNA expression, and advancing the rational design of mRNA vaccines for cancer immunotherapy (Toyama et al., 2023).

3 Antigen and epitope prediction

Antigen prediction uses bioinformatics to analyze pathogen genomes or proteomes, identifying specific epitopes that trigger immune responses through various MHC classes or DC and macrophages (Capelli et al., 2023). In mRNA vaccine development, choosing the right antigen targets is essential for effective expression and a strong immune response. Neo-antigen prediction technologies enhance vaccine safety and effectiveness by finding highly immunogenic epitopes, which can shorten development timelines and reduce costs (Soria-Guerra et al., 2015). The following section will detail the specific tools used in this process. Selecting optimal epitopes is crucial for robust immune stimulation, and epitope prediction tools are key in developing effective mRNA vaccines.

3.1 NetMHC

NetMHC is a user-friendly bioinformatics tool that utilizes information from both data types for training on binding affinity and eluting ligand data, thus being used to predict peptide-MHC interactions, addressing the challenge of identifying peptides that effectively bind to MHC molecules. NetMHC has undergone several transformative updates since its inception in the early 2000s, embracing the latest computational advancements and

significantly enhancing its database of interactions between peptides and MHC. These updates have incorporated sophisticated scoring matrices, intricate hidden Markov models, and cutting-edge artificial neural networks (ANNs) (Zhou et al., 2023), collectively enhancing the tool's predictive capabilities and broadening its application scope within the field, for example, NetMHCpan-4.0 achieves better performance, and ligands in all cases are predicted with very strong eluting ligand likelihood values (Jurtz et al., 2017). It has become an essential resource in immunoinformatics, crucial for understanding how peptide fragments derived from pathogens can activate CD8+ T cells and trigger immune responses, particularly neoantigens in cancer immunology (Wu et al., 2023).

Cytotoxic T cells play a central role in the pathogenesis and immunomodulation of malignancies, and the binding of peptides to MHC molecules is the most selective single step in the antigen presentation pathway. It has recently been shown that over 90% of naturally occurring MHC ligands are identified with 98% specificity (Nielsen and Andreatta, 2016). In vaccine development, NetMHC evaluates the binding affinity between peptides and MHC molecules, aiding researchers in selecting optimal peptides for vaccine inclusion to induce robust CD8+ T-cell responses. This capability enhances vaccine specificity and efficacy by focusing on peptides with the strongest interactions. During the intricate process of vaccine development, the versatile NetMHC tool harmoniously integrates with existing peptide-MHC data, leveraging computational simulations to accurately predict potential antigen epitopes – a pivotal step in vaccine design. This training approach integrates larger data content and can directly learn the length of each MHC molecule from the experimental binding data to present the optimal peptide (Andreatta and Nielsen, 2016). NetMHC provides highly accurate predictions due to its use of extensive training datasets and advanced modeling techniques like neural networks and position-specific scoring matrices (PSSMs). While it focuses on MHC class I molecules, its performance depends on the quality and breadth of peptide-MHC interaction data and may require substantial computational resources. Although NetMHC is excellent at predicting MHC interactions, it does not cover all aspects of antigen processing and presentation, such as class II MHC interactions. Despite these limitations, integrating NetMHC into the vaccine development process greatly improves the design of specific and effective vaccines.

3.2 IEDB Analysis Resource

The IEDB-AR (Immune Epitope Database Analysis Resource) is a crucial tool for designing mRNA vaccines against variable antigens, especially for virus-based vaccines such as those targeting influenza viruses, SARS-CoV-2, and HIV.I EDB-AR has T cell epitope prediction tools, B cell epitope prediction tools, and tools for the analysis of known epitope sequences or sequence groups. The IEDB-AR platform stands as an ideal choice for addressing diseases characterized by substantial antigenic variation or requiring a robust, multifaceted immune response. Its

applicability extends to a wide spectrum of conditions, including those associated with infections, allergies, autoimmune disorders, and transplantations, where its capabilities are particularly well-suited to inform and guide therapeutic strategies (Vita et al., 2018).

Several new tools have been added to IEDB-AR. Among the T cell epitope prediction tools are TepiTool, MHC-NP, Immunogenicity, CD4EpiScore, and Deimmunization. These tools have their different functions, such as TepiTool, It can be used to predict naturally processed MHC class I and II ligands, deimmunization of therapeutic proteins, and prediction of T cell immunogenicity beyond MHC binding affinity (Dhanda et al., 2019). IEDB-AR also adds a new tool called LYRA (Automated Modeling of Lymphocyte Receptors), which allows for the simulation of 3D structures of B and T cell receptors (Klausen et al., 2015), allowing for the prediction of canonical structures per cycle, when necessary.

By utilizing algorithms such as ANNs and SVMs to predict both class I and II peptide-MHC binding affinities, T-cell and B-cell epitopes, and cross-reactive epitopes (Yan et al., 2024), IEDB-AR identifies optimal antigenic targets, such as pHLA-target Ags (Gerber et al., 2020), to stimulate both CD4$^+$ and CD8$^+$ T-cell responses, as well as antibody responses. However, its effectiveness depends on the quality and comprehensiveness of the peptide-MHC interaction data, which can impact prediction accuracy. This variability in data coverage may affect the tool's precision.

3.3 SYFPEITHI

SYFPEITHI, a free bioinformatics tool from the late 1990s, predicts peptide-MHC interactions for MHC class I and II molecules. Its user-friendly interface and high accuracy help identify peptides that bind to specific MHC molecules and predict epitopes (Zhang et al., 2023a). The database includes peptide sequences (approximately 200 peptide motifs and 2000 peptide sequences), anchor position, MHC specificity, source protein, source organism, and publication references. The tool employs PSSMs as its primary algorithm to evaluate the binding affinity of peptides to MHC molecules, which can sequence the MHC-eluting peptides directly. The adopted scoring approach simplifies the identification of promising vaccine candidates by providing detailed binding scores and rankings. *Chao Shen* et al. findings show that this method effectively balances scoring and docking tasks, making the selection process both rigorous and efficient (Shen et al., 2023). But instead of synthesizing and testing dozens or even hundreds of peptides, SYFPEITHi prescreens a set of peptides and enables epitope prediction of the sequence, restriction elements, and their respective motifs of proteins or their genes., which aids in the design of effective vaccines (Rammensee et al., 1999). The accuracy of SYFPEITHI's predictions depends on the quality and completeness of the peptide-MHC interaction data, with gaps potentially affecting reliability. SYFPEITHI does not account for critical aspects of antigen processing and presentation, such as peptide transport into the endoplasmic reticulum via TAP, proteasome trimming, or competition for MHC binding. These

factors are essential for a full understanding of immune responses and peptide presentation (Larsen et al., 2005; Lee et al., 2024a).

4 Codon optimization

Before analyzing mRNA structure, it's essential to focus on the Coding Sequence (CDS) and codon optimization. Codon optimization is crucial for improving CDS expression in a host organism. This process involves modifying codons to match the host's preferred codon profile, which enhances gene expression efficiency and reduces costs (Hanson and Coller, 2018). Codon optimization takes into account factors such as codon usage bias, tRNA abundance, GC content, and RNA secondary structure. By carefully selecting codon combinations, researchers can improve protein expression, reduce mRNA degradation, and enhance stability. This also impacts protein folding, post-translational modifications, and immunogenicity (Zhang et al., 2023b). Software tools like GeneOptimizer and JCAT (Java Codon Adaptation Tool) help in this process by choosing the most efficient codons based on the host's tRNA abundance and codon usage patterns. Here is a summary of their advantages and disadvantages.

4.1 GeneOptimizer

GeneOptimizer is a powerful tool for optimizing DNA sequences. It uses a sliding window method to adjust codon usage, GC content, and other factors to improve translation efficiency (Fu et al., 2020). It handles large gene sequences and manages key processes such as transcription, splicing, translation, and mRNA degradation. GeneOptimizer can complete gene optimization in minutes. Synthetic genes were designed by uploading sequences, selecting expression systems, specifying cloning vectors, and sequence details. At the same time, based on the data related to a given organism and the user's sequence requirements, the DNA sequence that is most suitable for the user's research requirements is generated. Researchers can use this tool to select optimal codon combinations for specific organisms, enhancing gene expression efficiency and scaling up protein production to meet experimental needs. Despite being a premium tool, GeneOptimizer empowers users with the autonomy to meticulously craft gene sequences, circumventing the necessity for DNA templates. It achieves this through the implementation of sophisticated codon optimization and sequence alignment algorithms, exemplified by its utilization of sliding windows for refining multiparameter DNA sequences and FOGSAA for executing swift, global sequence alignments (Chakraborty and Bandyopadhyay, 2013). Importantly, GeneOptimizer enhances mRNA stability and prolongs its half-life within cells through codon optimization (Schwanhäusser et al., 2011; Luo et al., 2023). Optimized mRNA sequences, with more favorable codons, related studies have shown that using GeneOptimizer at the same dose can significantly increase protein expression and produce more antigen

proteins, leading to stronger immune responses and improved disease prevention. However, altering mRNA sequences with GeneOptimizer may have some unknown risks, such as potential interactions with other RNA and proteins within cells, which could lead to adverse reactions or reduced vaccine efficacy.

4.2 JCAT

The JCAT uses advanced algorithms, such as the Codon Adaptation Index and the Relative Codon Adaptation model, to enhance the production of heterologous proteins and there is no need to manually define highly expressed genes. Significantly, JCAT not only enhances gene sequence design but also safeguards against undesirable outcomes such as the emergence of restriction enzyme cleavage sites and Rho-independent transcription terminators. *Grote* et al.*'s* study underscores this capability, demonstrating how JCAT successfully adapted the codon usage of the *P. aeruginosa* exbD gene to that of *E. coli* while simultaneously evading the formation of identical restrictive sites, ensuring the stability of CDS. On the output, JCAT can be either a graph or a CAI (Codon Adaptation Index) value given by the pasted sequence and the newly adapted sequence. In addition, users can calculate CAI values by uploading gene sequences in FASTA format (Grote et al., 2005), which can help researchers quickly understand key biological information during mRNA vaccine development. JCAT is usually a codon optimization of a single gene in the laboratory. Therefore, experiments are comparing the original *Pseudomonas aeruginosa* DNA sequence with the DNA sequence optimized for Escherichia coli to demonstrate the degree of optimization. JCAT is user-friendly, offering high automation and precision, which allows researchers to efficiently analyze and adjust codon combinations. This optimization improves mRNA vaccine expression levels in host cells and avoids Rho-independent transcription terminators in codon-optimized DNA sequences (Postle and Good, 1985; Ermolaeva et al., 2000).

JCAT is built on biological insights into translational optimization, particularly the significance of codon adaptation in heterologous protein production. By leveraging algorithms like the Codon Adaptation Index (CAI), it aligns codon usage with host-specific tRNA pools, improving translation efficiency and reducing translational errors (Sample et al., 2019). Studies have shown that codon optimization not only enhances protein yield but also stabilizes mRNA expression by avoiding undesired sequence features, such as Rho-independent transcription terminators, which can destabilize transcripts (Leppek et al., 2022). Biologically, JCAT addresses critical factors in mRNA vaccine development, such as ensuring optimal ribosome loading to maximize protein translation while avoiding ribosome clustering that could lead to mRNA degradation. The tool's ability to safeguard against restriction enzyme cleavage sites and transcriptional terminators highlights its utility in designing sequences for experimental and therapeutic applications. These features align with the broader understanding of how codon adaptation influences mRNA stability and protein expression, making JCAT an invaluable resource for precise, biologically informed sequence optimization (Grote et al., 2005).

5 Secondary structure prediction

Predicting the secondary structures of mRNA, including elements like α-helices and β-sheets, is essential for understanding its tertiary structure and function (Jiang et al., 2023). This prediction helps identify regions prone to degradation, allowing researchers to optimize gene sequences for greater mRNA stability. By analyzing the secondary structure, scientists can design mRNA sequences that are more efficient for translation, thereby improving vaccine expression in the host. Additionally, understanding the mRNA structure aids in selecting the most effective delivery systems, ensuring that mRNA efficiently enters cells and translates into target proteins, which enhances vaccine efficacy. Notably, CRISPR-Cas gene editing technology exemplifies its immense potential in addressing disease-causing mutations stemming from various cellular origins, highlighting the transformative impact of such evaluations on biomedical research and therapeutics (Cheng et al., 2020), and guiding the selection of those that can elicit stronger immune responses. In this section, we compare key tools for predicting mRNA secondary structure: RNAfold, mFold, and Inverse Prediction of RNA Knot (IPKnot).

5.1 RNAfold

RNAfold, part of the Vienna RNA Package, uses a thermodynamic model, such as the nearest neighbor thermodynamic model (Calonaci et al., 2020) to predict RNA secondary structures by computing the minimum free energy (MFE) and the thermodynamic regularized RNAfold can be used to calculate folding fractions that are highly correlated with the true free energy (Sato et al., 2021). RNAfold predicts RNA secondary structures by analyzing sequence inputs along with folding constraints, algorithms, and energy parameters. Users can select options for dangling ends, modified bases, and SHAPE reactivity data. The output can be customized to include interactive RNA secondary structure maps, reliability annotations, or mountain plots.

While RNAfold does not engage in direct codon optimization, its profound capability in predicting mRNA structures lays a solid foundation for subsequent codon optimization endeavors. Notably, key regions within mRNA, such as the 5' UTR, 3' UTR, and Poly(A) tail, play pivotal roles in facilitating vaccine translation, where the application of advanced techniques like sparsification can further enhance their efficacy (Gray et al., 2024), and RNAfold can assist in optimizing these regions to enhance vaccine expression levels. RNAfold also has several servers, such as RNAalifold, which can predict a set of common structures of aligned DNA or RNA sequences (Hofacker et al., 2002), which can calculate the hybridization energy and base pairing pattern of two RNA sequences (Bernhart et al., 2006). However, accurate RNA sequence data is essential for RNAfold's predictions, and due to the complex diversity of RNA sequences, there may be some margin of error. Analyzing longer mRNA sequences also requires more computational time, which can significantly extend the development cycle.

5.2 Mfold

Mfold is a bioinformatics tool for predicting the secondary structure of RNA and DNA, similar to RNAfold. Similarly, Mfold contains several separate applications that can be used to predict nucleic acid folding, hybridization, and melting temperatures (Zuker, 2003). Mfold predicts the most likely secondary structure of a nucleic acid sequence by computing the most thermodynamically stable configuration (Zuker and Stiegler, 1981). This process involves calculating the free energy of various possible structures to determine which one is the most stable, which is essential for ensuring the stability of mRNA vaccines. Mfold uses dynamic programming algorithms to provide an optimal secondary structure based on the sequence and environmental conditions, such as Pknots-RE, NUPACK, gfold, and Knotty (Marchand et al., 2023) and the user can also change the rotation Angle to get the desired molecular folding orientation (Zuker, 2003). Unlike RNAfold, Mfold can identify regions in mRNA that might be prone to instability (Binet et al., 2023), such as regions with a high likelihood of forming secondary structures that may lead to degradation or poor folding. This sophisticated functionality empowers researchers to refine mRNA sequences with heightened precision for vaccine design, ensuring optimal performance. Furthermore, Mfold's unique capability to anticipate the intricate interplay between mRNA molecules and delivery systems, and to visualize these interactions in various graphic formats including PostScript, PNG, or JPG, further augments its value in the realm of vaccine development (Zuker, 2003). Figure 2 presents a detailed comparison between RNAFold and Mfold predictions for mRNA secondary structure. As shown in Figure 2A, RNAFold provides a comprehensive visualization, with a color-coded structure based on base-pairing probabilities. Warmer colors highlight highly stable regions, particularly within the UTRs and near the poly-A tail—key areas for mRNA stability and translational efficiency. This allows for an in-depth understanding of structural stability across the mRNA sequence. In contrast, Mfold offers a simpler structural model without color-coding or probabilistic information. While it generates quicker results, Mfold's predictions lack the depth required for a thorough stability analysis.

RNAFold further enhances its predictions with a detailed entropy analysis, using overlapping curves for MFE, probable folding pathways, and Centroid structures to illustrate structural variability at each nucleotide (Figure 2B). This analysis confirms low entropy in regions like the UTRs and poly-A tail, indicating stability in these essential areas. Although Mfold provides an energy-based prediction, it does not offer the same clarity in entropy distribution. As a result, RNAFold's combination of structural and stability data makes it better suited for precise applications, while Mfold remains useful for rapid, less detailed evaluations.

5.3 IPKnot

IPKnot is a specialized computational tool used to predict the secondary structure of mRNA molecules, providing critical insights into the folding process based on dynamic programming and thermodynamic principles (Kato et al., 2012). By simulating base-pairing interactions, IPKnot predicts structures such as hairpins, loops, and stems, which play a significant role in mRNA stability, translation initiation, and susceptibility to degradation by ribonucleases (Sato et al., 2011).

In the context of mRNA-LNP delivery, IPKnot's folding predictions are essential for optimizing the interaction between mRNA and LNPs. The predicted mRNA secondary structure influences the mRNA's ability to be encapsulated into LNPs, as well as the subsequent release and translation inside the target cell. IPKnot aids in designing mRNA sequences with secondary structures that are compatible with LNP formulations, enhancing encapsulation efficiency and promoting stable, controlled release into the cytoplasm. This stability is vital for maintaining the functional integrity of mRNA once inside the cell, ensuring that it can be efficiently translated to produce the encoded protein (Jabbari and Condon, 2014).

In cancer immunotherapy, specifically mRNA-based cancer vaccines, IPKnot plays a pivotal role in optimizing the mRNA sequence and its secondary structure for enhanced immune system activation. The folding pattern of the mRNA influences the conformation of the encoded antigen in MHC (Solheim et al., 1995). Efficient MHC class I and class II presentation is critical for triggering both CD8+ cytotoxic T cell responses and B cell-mediated antibody production against tumor-associated antigens. By fine-tuning the mRNA sequence to achieve an optimal secondary structure, IPKnot contributes to more efficient antigen presentation, thereby improving the activation of both the innate and adaptive immune systems. This leads to stronger and more sustained immune responses, which are essential for targeting and eradicating tumor cells (Bell et al., 2017). Moreover, IPKnot's role in optimizing mRNA folding extends to improving the translational efficiency of mRNA in clinical applications, including gene therapies and personalized vaccines. The tool is integral in ensuring that mRNA molecules remain stable during synthesis, storage, and delivery, providing a foundation for the development of mRNA-based therapies with high efficacy and minimal degradation (Lee et al., 2024b). This capability is particularly crucial in the design of mRNA vaccines, where the accurate prediction of secondary structures ensures that the mRNA sequences are robust and capable of eliciting the desired immune response (Bon et al., 2008).

6 Protein structure prediction

Protein structure prediction is crucial for understanding how mRNA vaccines generate their target antigens and interact within the host. Unlike costly proteomics techniques like gas chromatography-mass spectrometry (GC-MS), which analyze chemical compounds but don't directly reveal protein structures, protein structure prediction provides theoretical insights crucial for refining antigen design before empirical testing. Accurate predictions ensure proteins fold correctly and function as intended, enhancing immune response. Key methods in this field include AlphaFold and Rosetta (Genc and McGuffin, 2025), which help identify potential folding and stability issues early, guiding experimental strategies and reducing extensive
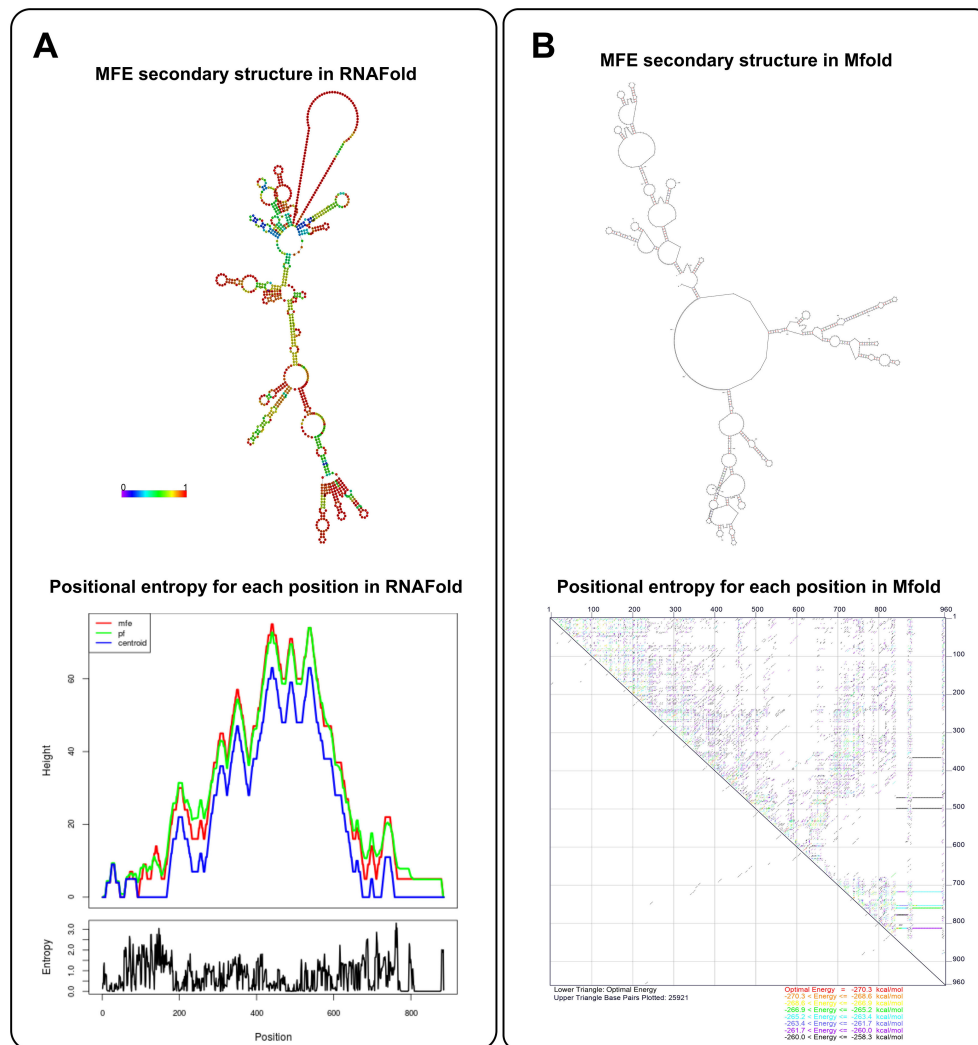
FIGURE 2

mRNA secondary structure prediction by RNAfold and Mfold. **(A)** The MFE structures for the template mRNA predicted by RNAFold and Mfold are shown. RNAFold (left) provides a detailed structural prediction with base-pairing probabilities, emphasizing stable regions-particularly in the UTRs and near the poly-A tail. This helps assess folding stability, essential for efficient translation. Mfold (right), while simpler, offers faster predictions, making it useful for quick structural overviews. **(B)** RNAFold's positional entropy analysis shows low entropy in the UTRs and poly-A tail, confirming greater structural stability in these regions. Mfold provides a broader structural view but lacks detailed entropy data, making RNAFold more precise for stability assessment, while Mfold excels in speed and simplicity. The template used is a 962 bp mRNA encoding 12 neoantigens, with key regulatory elements like the HGH 5' UTR, AES- mtRNA1-3' UTR, and a 121-base pair long poly-A tail, designed for stability and efficient translation. The predictions were generated using RNAFold (ViennaRNA Package 2.4.18) and Mfold (version 3.6) (Zuker and Stiegler, 1981), with both tools sourced from their respective official repositories.
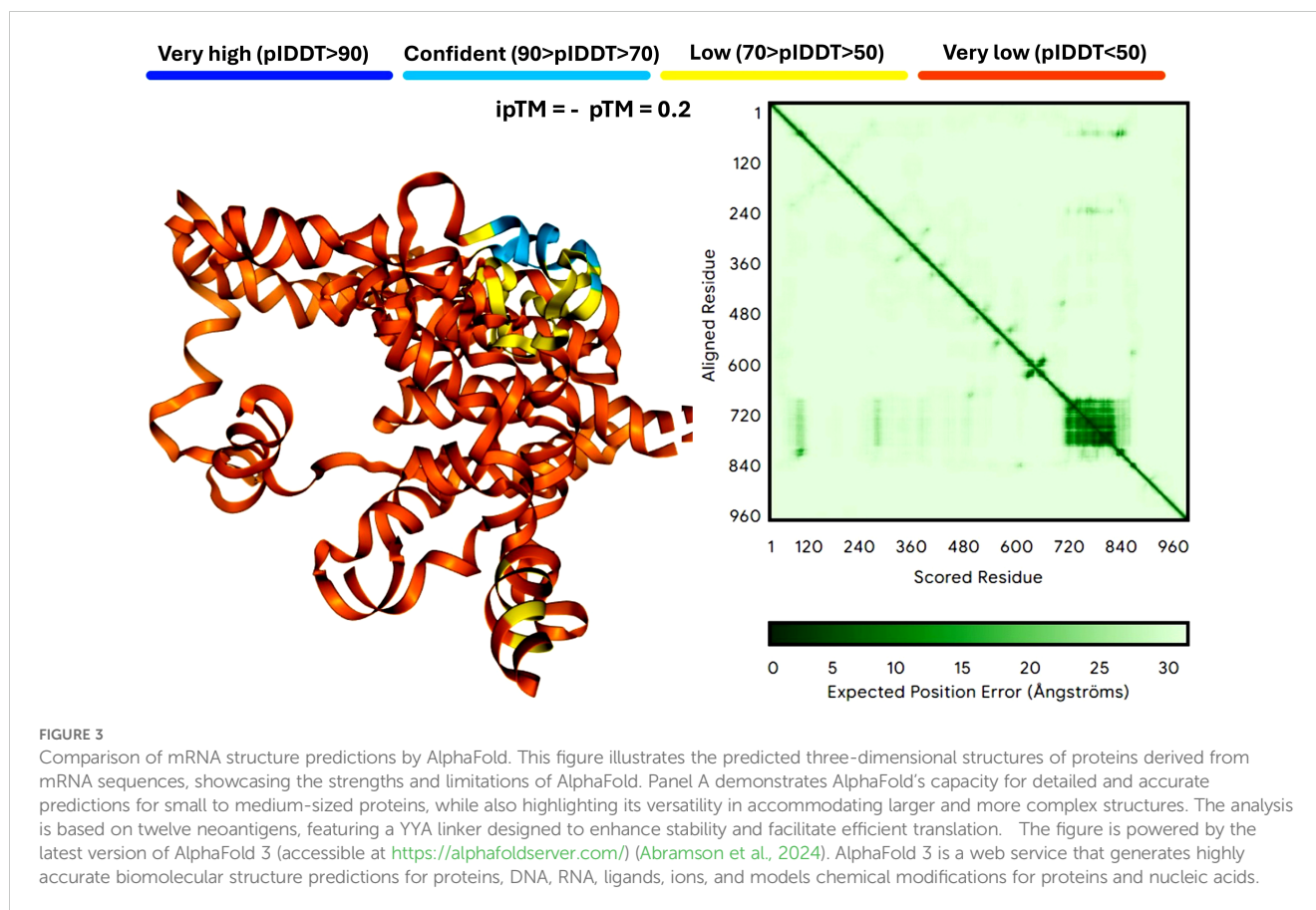
laboratory testing. Figure 3 compares mRNA structure predictions from AlphaFold and Rosetta.

Protein structure prediction is essential for understanding how mRNA vaccines produce their intended antigens and how these proteins interact within the host. Distinct from intricate and expensive proteomics approaches, such as GC-MS, which delve into chemical compounds yet fall short in directly illuminating protein structures, protein structure prediction stands as a theoretical cornerstone for refining antigen design before empirical validation. Its precision is paramount, as it ensures that synthesized proteins adopt their correct folds and execute their intended functions, thereby fostering a potent immune response. Key methods, such as AlphaFold and Rosetta (Genc and McGuffin, 2025), are commonly used in this field. These approaches help

identify potential issues in protein folding and stability early in the development process, guiding more effective experimental strategies and reducing the need for extensive laboratory testing.

6.1 AlphaFold

AlphaFold leverages deep learning techniques, such as the Attention Mechanism and Evolutionary Coupling algorithms, to predict the intricate 3D architectures of proteins from their amino acid sequences with atomic-level precision, even in the absence of prior structural knowledge (Jumper et al., 2021). Figure 3 shows the results of mRNA structure predictions by AlphaFold. In particular, AlphaFold can handle the missing physical environment and

FIGURE 3
Comparison of mRNA structure predictions by AlphaFold. This figure illustrates the predicted three-dimensional structures of proteins derived from mRNA sequences, showcasing the strengths and limitations of AlphaFold. Panel A demonstrates AlphaFold's capacity for detailed and accurate predictions for small to medium-sized proteins, while also highlighting its versatility in accommodating larger and more complex structures. The analysis is based on twelve neoantigens, featuring a YYA linker designed to enhance stability and facilitate efficient translation. The figure is powered by the latest version of AlphaFold 3 (accessible at https://alphafoldserver.com/) (Abramson et al., 2024). AlphaFold 3 is a web service that generates highly accurate biomolecular structure predictions for proteins, DNA, RNA, ligands, ions, and models chemical modifications for proteins and nucleic acids.

generate accurate models in challenging situations, such as intertwined homologs or proteins that fold only in the presence of an unknown heme group (Abramson et al., 2024). At the same time, AlphaFold has greatly improved the accuracy of structure prediction by combining a novel neural network architecture and training program based on evolutionary, physical, and geometric constraints of protein structure (Jumper et al., 2021). It combines a novel neural network architecture and training program rooted in evolutionary, physical, and geometric constraints to achieve unparalleled accuracy. With a database exceeding 214 million predicted protein structures, AlphaFold has transformed structural biology and set new benchmarks in protein structure prediction (Varadi et al., 2022; Abramson et al., 2024; Varadi et al., 2024), which are crucial for designing vaccines. However, it demands significant computational power and is dependent on the quality of input data. It works well for small to medium-sized proteins but may struggle with very large or complex proteins due to these resource limitations. DeepMind's AlphaFold has set a new benchmark in protein structure prediction, employing advanced deep learning frameworks such as CNNs and attention mechanisms to achieve unparalleled precision. With a database containing over 214 million predicted protein structures, AlphaFold has profoundly influenced structural biology, providing a median backbone accuracy of 0.96 Å r.m.s.d.95 and an all-atom accuracy of 1.5 Å r.m.s.d.95 (Jumper et al., 2021).

In mRNA vaccine development, AlphaFold's detailed structural insights are pivotal for optimizing antigen design, ensuring their stability and immunogenicity. Its ability to model viral proteins encoded by mRNA is crucial for assessing antigenicity, which is essential for effective vaccine formulations (Asediya et al., 2024). Furthermore, AlphaFold plays a significant role in enhancing the development of mRNA-LNPs delivery systems (2020; Asediya et al., 2024). LNPs encapsulate and protect mRNA during systemic circulation, facilitating targeted delivery and efficient intracellular release. AlphaFold's structural predictions can guide the design of mRNA sequences with stable secondary structures, such as stem-loops or pseudoknots, to improve binding affinity and stability within LNPs. This optimization enhances protection against nuclease degradation and ensures efficient delivery to target cells (2020).

Additionally, AlphaFold can refine mRNA designs to optimize release kinetics from LNPs within target cells. By predicting how mRNA structures interact with LNP components in response to intracellular conditions, such as pH or enzymatic activity, AlphaFold aids in developing formulations that promote efficient unpacking and robust antigen translation upon endosomal escape. These advances enhance DCs activation and antigen presentation to T cells, ensuring a potent adaptive immune response (Jumper et al., 2021; Olawade et al., 2024). AlphaFold also supports the design of mRNA elements encoding immune-stimulatory adjuvants, amplifying immunogenicity when combined with LNP formulations. By optimizing antigen stability and presentation by MHC molecules, AlphaFold contributes to tailored mRNA-LNP formulations that elicit durable and specific immune responses. This capability is particularly valuable in cancer immunotherapy, enabling more precise and effective vaccine designs (Oladipo et al., 2024).

6.2 Rosetta

Rosetta has become one of the leading computational tools for biomolecular structure prediction and design, using energy-based models, including Monte Carlo simulations and the Rosetta Energy Function (Rohl et al., 2004; Varadi et al., 2024), to predict 3D protein structures from mRNA sequences (Koehler Leman and Künze, 2023). It is useful for analyzing protein structures and interactions, and can also model antibodies and antigens (Schoeder et al., 2021), Rosetta can read most glycans in PDB files and automatically detect and score them, helping in the design of mRNA sequences to elicit strong immune responses. Rosetta provides detailed and adaptable predictions for proteins of various sizes, from small peptides to larger proteins, between 10-1000 residues (Du et al., 2021; Schmitz et al., 2021). However, it requires careful setup and parameter tuning, and its accuracy can decrease for very large or complex proteins.

7 Molecular dynamics simulations

MD simulations are crucial for understanding the intricate movements and interactions of atoms in mRNA molecules, especially how they interact with proteins and other cellular components (Hollingsworth and Dror, 2018). These simulations rely on several key parameters: Force fields are used to define atomic interactions, while temperature and pressure controls mimic physiological conditions. The permeability of lipid membranes is also considered to model interactions accurately (Venable et al., 2019). Time steps in simulations balance accuracy with computational efficiency. Cutoff distances manage non-bonded interactions, and periodic boundary conditions help minimize edge effects, enhancing the realism of the model. Solvent models simulate the surrounding aqueous environment, providing a more comprehensive view of mRNA behavior. Electrostatic treatments handle long-range interactions, ensuring that all aspects of the mRNA's behavior are captured accurately. Simulation length and advanced techniques, such as replica exchange (Bock et al., 2023), offer deeper insights into mRNA dynamics. By optimizing these parameters, researchers can ensure that mRNA vaccines maintain their structure and function, thereby improving their effectiveness and

immunogenicity. However, there are MD simulation tools for mRNA vaccine development (Table 2).

7.1 GROMACS

GROMACS is a leading MD simulation tool known for its efficiency and precision. GROMACS can use MD, stochastic dynamics, or path integration methods to simulate any molecule in a solution or crystal, minimize molecular energy, analyze conformation, etc. Its simulation package includes GROMACS force fields (proteins, nucleotides, sugars, etc.) and can range from glass and liquid crystals to polymers, crystals, and biomolecular solutions. It effectively models the movements and interactions of atoms and molecules using advanced force fields (Rawat et al., 2021). GROMACS stands out as an exceptional tool for capturing the intricate dynamics and binding mechanisms of complex macromolecules, particularly mRNA-protein complexes. Its prowess lies in the employment of advanced algorithms, such as the Particle-Mesh Ewald (PME) method, which necessitates seamless all-to-all communication between the computational nodes (Kohnke et al., 2020). This sophisticated approach ensures that GROMACS can accurately model and analyze even the most challenging molecular interactions, providing unparalleled insights into their behavior, for accurate long-range electrostatics and supporting parallel processing for large-scale simulations. This makes it ideal for assessing the stability and behavior of biomolecular structures, crucial for optimizing vaccine designs. While GROMACS provides a powerful toolkit, beginners might need some time to learn how to use it, especially if integrating with R for data analysis and visualization, such as YAMACS, which can show the results in real time (Sarkar et al., 2022).

7.2 AMBER

Assisted Model Building with Energy Refinement (AMBER) is a well-known MD simulation tool used to predict and refine the 3D structures of mRNA. AMBER excels in offering intricate energy calculations and structural analyses of mRNA vaccines, thanks to its harnessing of efficient parallel computing and cutting-edge

TABLE 2 MD simulation tools for mRNA vaccine development.

Feature/Tool	GROMACS	AMBER	NAMD	Rosetta	CHARMM
Primary Use	Large-scale sims	Struct. pred. & refine	Large-scale sims	Prot. model & design	Prot. model & design
Key Algorithms	PME, Parallel Proc.	GB model, PME	PME, MTS integr.	MC, Energy funcs	C36, Leapfrog integr.
Force Fields	Adv. for macro.	ff14SB, NA	CHARMM	Prot.-focused	C36, General
Strengths	Efficient, large sys.	Accurate, energy anal.	Large-scale, dyn. studies	Prot. design, interact.	mRNA-prot. anal.
Learning Curve	Mod.	Steep	Mod.	Mod.	Steep
Computational	High	High	Very high	Mod.	High
Advantages	Fast, ext. data integ.	Energy & struct. anal.	Large-scale, precise elect.	Prot. fold. & design	Free energy & struct. anal.

Sims, Simulations; Struct. pred. & refine, Structure prediction and refinement; Prot., Protein; Model, Modeling; MC, Monte Carlo; Funcs, Functions; Proc., Processing; Adv., Advanced; Macro., Macromolecules; NA, Nucleic Acids; Integr., Integrator; Dyn., Dynamic; Sys., Systems; Anal., Analysis; Interact., Interactions; Mod., Moderate; Elect., Electrostatics; Ext., External; Integ., Integration; Fold., Folding; C36, CHARMM36 (a force field).

algorithms. The integration of the Generalized Born model and the Particle-Mesh Ewald method, among others, ensures that the simulations capture every nuance of the molecular interactions. AMBER has a variety of force fields suitable for different biomolecules, such as AMBER force field, CHARMM force field, etc., which can accurately describe the physical and chemical properties of various biomolecules. Furthermore, AMBER's specialized force fields, notably the ff14SB and nucleic acid force fields, contribute to the highly accurate modeling of nucleic acids and proteins, providing unparalleled insights into the behavior of mRNA vaccines (Mikhailovskii et al., 2022). The package includes advanced techniques for energy minimization and refinement, such as the conjugate gradient and steepest descent methods, which ensure precise structural optimizations. Additionally, AMBER's detailed analysis tools, including the Markov State Models and Principal Component Analysis, deliver deeper insights into tremolo-MD interactions of the mRNA vaccine. Its capabilities for implicit solvation and advanced free energy calculations not only study protein folding but also enhance the understanding of biomolecular stability and interactions (Shao and Zhu, 2018; Mikhailovskii et al., 2024).

7.3 NAMD

Nanoscale molecular dynamics (NAMD)'s advanced parallel computing techniques offer significant benefits for the development of neo-antigen mRNA vaccines. Its capability to perform large-scale simulations (Acun et al., 2018), involving millions of atoms, enables detailed modeling of neo-antigen mRNA vaccines. The use of sophisticated force fields, such as CHARMM, ensures accurate modeling of interactions between neo-antigen mRNA and proteins, which is crucial for predicting how neo-antigens are presented to immune cells and how they might stimulate an immune response. NAMD's PME method stands as a testament to its precision in modeling long-range electrostatic interactions. This innovative approach enables the calculation of complete, non-truncated electrostatic interactions at a minimal computational cost, ensuring that the simulations are both accurate and efficient. With NAMD, researchers can gain unparalleled insights into the intricate electrostatic behavior of their molecular systems (Phillips et al., 2005, 2020). For neo-antigen mRNA vaccines, this means accurately simulating the electrostatic interactions between mRNA and protein targets, which is essential for understanding binding affinities and stability. Additionally, the Multiple Time-Step (MTS) integrator allows NAMD to handle different time scales efficiently (Phillips et al., 2005; Pechlaner et al., 2021), which is particularly useful for studying the dynamic behavior of neo-antigen mRNA and its interactions over time, offering insights into how these interactions evolve and affect the vaccine's efficacy.

7.4 Rosetta

Rosetta is a versatile molecular modeling tool used primarily for protein structure prediction, protein-protein, protein-peptide complexes, and protein-ligand docking, and the design of biomolecules, can also model the RNA molecules in 3D structure (Koehler Leman and Künze, 2023). Unlike traditional MD simulation tools, Rosetta employs energy functions and Monte Carlo sampling methods to explore molecular interactions, and uses standard off-the-shelf computational hardware and all-atomic force fields to model the Large-scale conformational changes in proteins (Alford et al., 2017; Heilmann et al., 2020). Only a handful of structural biomolecule modeling frameworks have similar capabilities to Rosetta, covering applications of structural prediction and experimental data modeling, as well as protein design and small molecule drug discovery (Koehler Leman and Künze, 2023). It excels at predicting protein folding and designing new protein structures, making it valuable for integrating mRNA sequences with protein components to improve vaccine design. Rosetta's ability to model protein interactions and design novel biomolecules complements MD simulations by providing additional insights into the structural and functional aspects of mRNA vaccines, and Some protein sampling limitations were overcome by the combination of MD simulation and Rosetta (Lindert et al., 2013).

7.5 CHARMM

CHARMM (Chemistry at Harvard Macromolecular Mechanics) is a sophisticated MD simulation package known for its in-depth analysis of biomolecular systems. It employs advanced force fields and simulation algorithms to model the movements and interactions of molecules over time (Brooks et al., 2009). CHARMM excels in the study of nucleic acids and proteins, thanks to its highly detailed force fields, such as CHARMM36 and the CHARMM General Force Field. These force fields allow for the precise modeling of mRNA structures and their interactions with proteins, which is essential for understanding the stability and behavior of mRNA vaccines. The software also offers robust integration methods, including the Verlet algorithm and the Leapfrog integrator, which significantly enhance the accuracy and efficiency of simulations. Moreover, CHARMM's advanced energy minimization techniques, like the conjugate gradient and steepest descent methods, ensure thorough structural optimization of biomolecules (Jo et al., 2017). Additionally, CHARMM supports various analyses, such as free energy calculations, principal component analysis, and MD trajectory analysis, providing comprehensive insights into the dynamics and stability of mRNA and its interactions. In vaccine development, CHARMM's capacity to simulate the intricate interactions between mRNA and protein components provides crucial insights into how these interactions impact vaccine efficacy. The package's capabilities for modeling complex biomolecular systems and its extensive set of tools for analysis make it a powerful choice for researchers focused on optimizing mRNA vaccines and other biomolecular studies.

8 mRNA-LNPs formulation

LNPs are the only FDA-approved carriers for mRNA vaccines, ranging from 70 to 200 nm in size. They are crucial for encapsulating and stabilizing mRNA molecules, facilitating their

effective delivery into target cells (Li et al., 2022a). The structural composition of LNPs, typically including lipids, cholesterol, and polyethylene glycol (PEG) (Hald Albertsen et al., 2022), directly influences their efficiency and efficacy. Key structural features, including lipid headgroup interactions and the arrangement of hydrophobic tails, play a critical role in the ability of LNPs to fuse with cell membranes and effectively deliver mRNA payloads. Designing and optimizing LNPs for mRNA cancer vaccines demands advanced computational tools capable of modeling and visualizing these complex structures and interactions. Tools such as NANOdesign, POLYVIEW-3D, and PyMOL are indispensable in this process. Figure 4 depicts the various sections of the mRNA-LNP complex that must be designed, optimized, and characterized using these bioinformatics tools. This figure highlights how these tools contribute to achieving stable, functional, and highly efficient mRNA-LNP formulations, addressing aspects from pharmacology to pharmaceutical applications. By utilizing these resources, researchers can fine-tune parameters such as lipid composition, particle size, and surface properties to improve the performance and stability of LNPs in mRNA cancer vaccines.

8.1 mRNA-LNPs design tools

8.1.1 NANOdesign

NANOdesign is a specialized tool for designing and optimizing LNPs for mRNA vaccines. Nanostructures can be formed by artificial design or by natural self-assembly mechanisms, which rely on intermolecular forces to automatically arrange into ordered

structures. It provides detailed capabilities for modifying lipid types and ratios, essential for achieving optimal mRNA encapsulation and stability.

NANOdesign allows for comprehensive adjustments to lipid composition, which directly impacts the efficiency of mRNA encapsulation (Li et al., 2023). Researchers can explore different lipid types, such as phospholipids, ionizable lipids, and cholesterol, to determine the optimal combination for creating stable and effective nanoparticles. By adjusting these components, they can influence the fluidity and permeability of the lipid bilayer, directly impacting the retention and protection of mRNA within the LNP. NANOdesign also controls particle size, which is crucial for effective cellular uptake. It simulates how different formulation parameters impact the size and uniformity of the nanoparticles, ensuring they fall within the ideal range for delivery. Additionally, NANOdesign allows for modifications to surface properties, such as charge and hydrophilicity, which can alter cell membrane permeability. These adjustments can enhance the vaccine's efficacy and improve its overall stability (Salatin et al., 2015). The tool models the release kinetics, encapsulation efficiency, and stability of mRNA within the LNPs, allowing researchers to optimize the release profile for controlled delivery. It also assesses the stability of LNPs under various conditions, including temperature and pH changes, to ensure the mRNA remains intact throughout storage and administration.

8.1.2 POLYVIEW-3D

POLYVIEW-3D is an advanced tool designed to visualize and analyze the 3D structures of mRNA-LNPs, playing a crucial role in
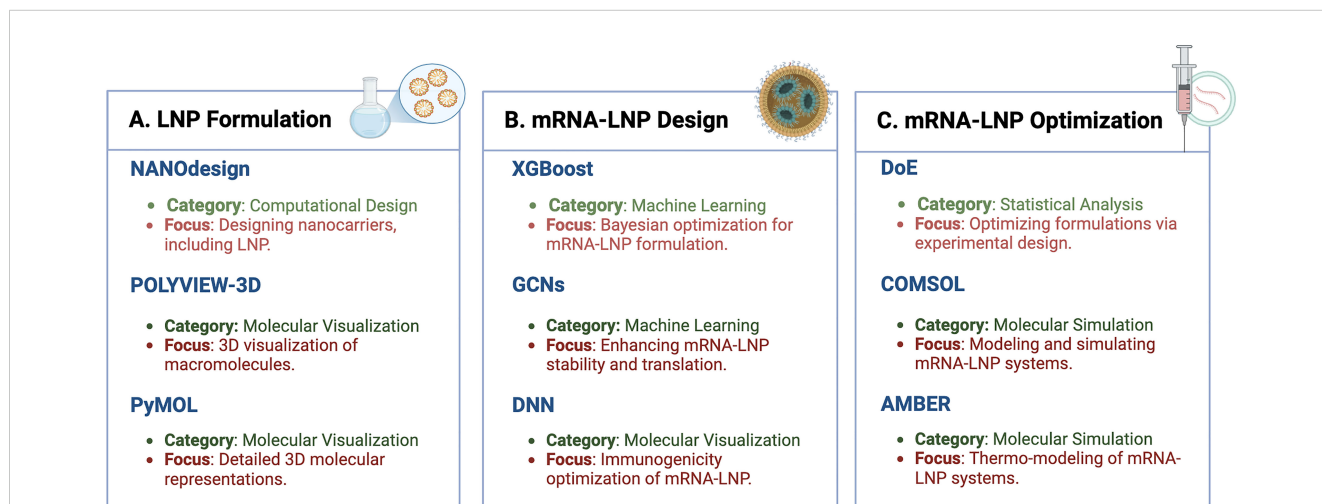

FIGURE 4
Integrated *in silico* framework for mRNA-LNP design, formulation, and optimization. This figure provides a comprehensive overview of the design and optimization process for mRNA-LNPs. **(A)** LNP formulation. This section highlights the tools used for nanoparticle formulation. NANOdesign enables computational modeling of nanocarriers like LNPs. POLYVIEW-3D and PyMOL offer molecular visualization for analyzing macromolecules and creating detailed 3D models of LNP structures **(B)** mRNA-LNP design. Advanced machine learning approaches, such as XGBoost are used for the optimization of mRNA-LNP formulations. Graph convolutional networks (GCNs) enhance mRNA stability and translation through better lipid and matrix design, while deep neural networks (DNNs) assist in optimizing immunogenicity. **(C)** mRNA-LNP optimization. DoE (Design of Experiments) employs statistical methods to streamline and improve formulations. COMSOL and AMBER simulate molecular and thermodynamic properties of LNPs to predict their behavior under various conditions, advancing their delivery efficiency. This figure was created using BioRender to incorporate high-quality symbols and illustrations for clarity.

the design and optimization of these nanoparticles for mRNA delivery (Porollo and Meller, 2007). This software enables researchers to create highly detailed 3D models of LNPs, allowing for a deeper understanding of how these particles interact with mRNA and cellular components (Porollo et al., 2004). By providing clear, high-resolution representations of the nanoparticles, POLYVIEW-3D helps scientists examine the precise arrangement of lipids and other key components within LNPs, which is vital for optimizing their structure for effective mRNA encapsulation, stability, and delivery (Porollo and Meller, 2010).

One of the primary advantages of POLYVIEW-3D is its ability to model lipid bilayer formation and nanoparticle morphology. The software allows researchers to simulate how lipids are organized within the nanoparticle, which is essential for determining the overall stability and functionality of LNPs. Lipid composition plays a crucial role in the efficiency of mRNA encapsulation, as well as the nanoparticle's ability to protect and deliver mRNA to target cells (Abumanhal-Masarweh et al., 2019). POLYVIEW-3D enables users to explore how variations in lipid composition—such as the inclusion of ionizable lipids, phospholipids, and cholesterol—affect the nanoparticle structure, influencing factors such as encapsulation efficiency, particle size, and surface charge. By adjusting these parameters, researchers can fine-tune the LNP design for optimal mRNA delivery (Watson et al., 2005; Arno et al., 2020).

The tool is especially valuable for examining LNP interactions with cell membranes. POLYVIEW-3D leverages the fluorescent marker Rhodamine 123 (Rho123) to visualize and analyze the activity of the P-glycoprotein membrane transporter in the MDCKII-MDR1 transgenic cell line (Sklenářová et al., 2021). This analysis is crucial for understanding how LNPs are taken up by cells, including immune cells like DCs, which play a pivotal role in the immune response to mRNA vaccines. By studying these interactions, POLYVIEW-3D aids in ensuring that mRNA-LNPs are efficiently delivered to the appropriate target cells, such as DCs, and interact with key receptors, including TLRs, to trigger the desired immune response (Porollo and Meller, 2007).

POLYVIEW-3D also facilitates the analysis of nanoparticle morphology and shape, which are key factors in the effectiveness of LNPs. Nanoparticle shape influences how LNPs interact with cell membranes during endocytosis and how they release their encapsulated mRNA into the cytoplasm (Mrazek et al., 2014). By visualizing and manipulating the structure of LNPs, researchers can assess how changes in shape and size impact the delivery process. This is particularly important for optimizing the stability and function of LNPs, as irregularities in shape or size may affect their ability to cross cellular membranes or release mRNA efficiently (Byrgazov et al., 2013).

Another important application of POLYVIEW-3D is in the design of nanocomposite hydrogels, which are often used in conjunction with LNPs to improve the stability and delivery of mRNA vaccines (Baigorria et al., 2023). The software allows researchers to model how changes in lipid composition or particle size affect the hydrogel's properties, helping to optimize the gel's characteristics for enhanced mRNA delivery (Ege et al., 2023). Nanocomposite hydrogels can provide additional protection for LNPs during storage and transport, and POLYVIEW-3D helps

ensure that the LNPs remain stable and effective under different conditions (Baigorria et al., 2023).

By offering these capabilities, POLYVIEW-3D plays a pivotal role in the optimization of mRNA-LNP vaccines. Its ability to model complex interactions at the molecular level, visualize the structural components of LNPs, and analyze their interactions with cell membranes makes it an essential tool for researchers working to improve mRNA vaccine formulations. Whether fine-tuning lipid composition, modeling particle morphology, or studying cellular uptake, POLYVIEW-3D enables researchers to optimize LNP designs for maximum efficacy and stability in mRNA vaccine development (Bates et al., 2001; Porollo et al., 2004).

8.1.3 PyMOL

PyMOL is a widely used molecular visualization and analysis tool that plays a critical role in designing and optimizing mRNA-LNP vaccines. This software enables researchers to create high-resolution, 3D representations of molecular structures, providing deep insights into the interactions between lipids, mRNA, and cellular membranes (Rigsby and Parker, 2016). The ability to visually manipulate and analyze the structures at the atomic level makes PyMOL an invaluable tool for optimizing LNP formulations (Mooers, 2020; Martí-Centelles et al., 2024).

PyMOL helps researchers build and visualize the three-dimensional structures of LNPs. By displaying how lipids are arranged within the nanoparticle, PyMOL allows for detailed structural analysis, including the packing of lipids in the bilayer (Wang and Deserno, 2010). This visualization aids in understanding how changes in lipid composition impact the overall stability and encapsulation efficiency of mRNA. By adjusting parameters like lipid chain length, headgroup types, and ionizable lipids, researchers can explore how these factors influence the structural integrity of LNPs and their ability to encapsulate mRNA efficiently.

PyMOL, using the molecular lipophilicity potential (MLP), a well-established method to calculate and visualize lipophilicity in molecules, allows researchers to observe interactions between the hydrophilic and hydrophobic regions of lipids and the charged and polar components of mRNA. By modeling these interactions, PyMOL helps identify optimal lipid compositions that improve mRNA encapsulation and stability, ensuring that the mRNA remains protected during delivery and is efficiently released once the LNP reaches its target cells (Oberhauser et al., 2014).

One of the critical steps in LNP design is the selection of the appropriate lipid mixture. PyMOL provides a platform to examine how varying lipid components—such as phospholipids, cholesterol, and ionizable lipids—affect the structural and functional properties of LNPs Seeliger and de Groot, 2010. By visualizing the changes in nanoparticle morphology and surface charge as lipid composition is modified, researchers can determine the best formulation for maximizing mRNA encapsulation, delivery efficiency, and stability under physiological conditions. PyMOL helps fine-tune these compositions, optimizing the LNP's ability to deliver mRNA effectively while maintaining stability during storage and transport (El Khoury et al., 2023). In addition to lipid composition, the surface properties of LNPs, including charge, hydrophobicity, and hydrophilicity, play a crucial role in their interaction with cellular

membranes. PyMOL allows researchers to model how altering these properties impacts the nanoparticle's ability to be taken up by cells. PyMOL helps in understanding how LNPs interact with cellular membranes during the process of endocytosis. By simulating the insertion of LNPs into the lipid bilayer of a cell membrane, PyMOL enables researchers to visualize how LNPs may fuse with the membrane and release mRNA into the cytoplasm. This insight is vital for designing LNPs that optimize cellular uptake and ensure the efficient delivery of mRNA into cells for translation (Yong, 2015). PyMOL allows researchers to visualize the impact of particle size and morphology on the functionality of LNPs. Particle size is critical for effective cellular uptake, and by using PyMOL, researchers can simulate how varying nanoparticle sizes and shapes affect the overall performance of mRNA delivery. The tool helps visualize how the size and shape of the LNPs influence their stability, encapsulation efficiency, and release kinetics, which are essential factors for improving vaccine efficacy (Cao et al., 2020; Sebastiani et al., 2021).

PyMOL can also be used to simulate the effects of various environmental conditions such as pH, temperature, and ionic strength on the stability and function of LNPs. By visualizing how LNPs change under different conditions, researchers can predict the behavior of the vaccine during storage, transport, and after administration, ensuring that the mRNA remains intact and functional throughout the vaccine's lifecycle (Arno et al., 2020; Sebastiani et al., 2021). One prominent example of PyMOL's application in LNP design is the study by Zhang et al. (2023a), which explored LNP formulations for mRNA vaccines. In their work, the researchers used PyMOL to visualize and model the interaction between the lipid components of the LNPs and the encapsulated mRNA (Zhang et al., 2023a). By adjusting lipid compositions and evaluating the resulting structural and functional properties, the team optimized the LNPs to enhance mRNA encapsulation and improve delivery efficiency. The use of PyMOL in this study enabled the team to refine the nanoparticle design, resulting in a more stable and effective LNP for mRNA delivery (Arévalo-Romero et al., 2024).

8.2 mRNA-LNPs optimization

8.2.1 Design of Experiments

Design of Experiments (DoE) is a systematic approach that allows for strategic compromises on information to significantly reduce the time and resources needed to understand and optimize a given process (Rampado and Peer, 2023), The design space is defined by the mathematical relationship between Critical Process Parameters (CPPs) and Material Attributes (CMAs) and Critical Quality Attributes (CQAs) (Politis et al., 2017). This method has been employed to optimize the formulation of mRNA-LNP vaccines by addressing critical parameters such as lipid composition, release kinetics, lipid-to-mRNA ratio, particle size, and surface charge (Tavares Luiz et al., 2021; Gurba-Bryśkiewicz et al., 2023). For cancer immunotherapy, optimizing these parameters is crucial to ensure that LNPs effectively deliver mRNA to tumor cells, improve half-life, bioavailability, and biodistribution, enhance antigen presentation, and elicit a strong

immune response, for example, mRNA vaccines against COVID-19 are designed using DoE iterations to minimize or increase cell activation and to meet CQA characteristics while improving protein expression (Ly et al., 2022; Morris and Kopetz, 2022). This method also aids in fine-tuning the physicochemical properties of LNPs, including hydrodynamic diameter, zeta potential, and lipid bilayer integrity, to achieve the most effective therapeutic outcomes.

8.2.2 COMSOL

COMSOL Multiphysics is a sophisticated simulation tool used to model the physical behavior of mRNA-LNP formulations (Towne et al., 2021; Zhang et al., 2024). It allows for the simulation of various nanoparticle dynamics, such as diffusion rates, aggregation behavior, and interactions with cellular membranes, and through experimental verification, COMSOL can be used to synthesize nanoparticles down to the nm level (Erdem et al., 2023). The tool supports the optimization of key parameters, including nanoparticle size distribution, release kinetics, and LNP stability under different environmental conditions.

At the same time, COMSOL allows equations from different physics domains to be solved simultaneously in the same model, allowing for a more realistic simulation of the interactions between various physics under real-world operating conditions. This approach includes modeling the behavior of LNPs in various biological fluids, their ability to traverse cellular membranes, and their release profiles under different physiological conditions. For instance, Chenguo Yao et al. team developed a dynamic electroporation model of irregular cells using COMSOL to investigate the effects of ns+mμs pulses on these cells (Yao et al., 2020). The findings by Erdem et al. regarding COMSOL simulations for optimizing flow rates and mixing efficiency in micro-reactors can be adapted to mRNA-LNP cancer therapy development. By employing COMSOL, key processes such as nanoparticle formation, encapsulation efficiency, and controlled mixing of lipids and mRNA can be simulated to ensure consistency and precision. This could optimize LNP size (e.g., 50-100 nm), improve payload stability, and refine production conditions, facilitating scalable and efficient mRNA delivery systems for enhanced therapeutic efficacy in cancer immunotherapy (Erdem et al., 2023). In a parallel effort, Li et al. leveraged advanced combinatorial chemistry and Ml to identify ionizable lipids for mRNA delivery. Ml rapidly optimized lipid libraries, improving encapsulation efficiency and targeting. COMSOL's potential lies in enhancing such designs by simulating factors like nanoparticle stability, bioavailability, and interaction dynamics to refine LNP formulations for applications like mRNA-based cancer immunotherapy (Li et al., 2024).

9 AI and machine learning tools

9.1 XGBoost/Bayesian

XGBoost is a highly effective machine-learning algorithm known for its speed and predictive power, especially when working with structured datasets that include mRNA-LNP

formulation parameters. The model employs an ensemble method that builds multiple decision trees sequentially to correct for errors made by previous trees, which allows it to handle a broad range of input variables such as lipid composition, nanoparticle size, and encapsulation efficiency. XGBoost can be optimized through hyperparameter tuning, allowing the model to refine predictions of key formulation characteristics like mRNA stability, LNP encapsulation efficiency, and delivery performance (Maharjan et al., 2024).

When combined with Bayesian optimization, this approach takes advantage of a probabilistic model to efficiently navigate the hyperparameter space. Bayesian optimization uses prior knowledge (based on previous experimental data or expert knowledge) to predict the most likely optimal formulation parameters and iteratively refines the search based on observed outcomes (Hoseini et al., 2023). For example, in mRNA-LNP optimization, Bayesian methods can help fine-tune lipid-to-mRNA ratios, lipid types (such as ionizable lipids), and nanoparticle characteristics (like surface charge or size) to maximize mRNA encapsulation, stability, and cell delivery (Sato et al., 2024). For mRNA vaccine development, XGBoost/Bayesian optimization accelerates the formulation process by systematically evaluating a range of conditions with minimal experimental trials, ensuring faster production of effective vaccine candidates. Recent applications show that combining these techniques can enhance mRNA-LNP performance, such as improving vaccine stability and optimizing lipid compositions that facilitate efficient mRNA delivery into immune cells (Castillo-Hair and Seelig, 2022). These algorithms have been proven to identify LNP formulations with the ability to trigger stronger immune responses by ensuring efficient antigen presentation and reducing the risk of immunogenicity-related adverse effects (Hoseini et al., 2024).

9.2 Graph Convolutional Networks

Graph Convolutional Networks (GCNs) represent a transformative application of Ml in mRNA-LNP vaccine development. These networks effectively model complex relationships inherent in graph-structured data, making them uniquely suited for tasks involving the intricate design and optimization of both mRNA sequences and their delivery systems (Gao et al., 2024). In mRNA vaccine design, GCNs encode secondary and tertiary structures of mRNA as graph networks, where nodes represent nucleotides, and edges depict structural interactions, such as hydrogen bonds or stacking interactions. This representation enables the identification of key features, like stem-loops or pseudoknots, that contribute to stability and translation efficiency, facilitating the optimization of mRNA constructs for robust antigen expression (Dorsey et al., 2024).

GCNs are also instrumental in designing LNP formulations and modeling the interactions between mRNA and lipid components, including ionizable lipids, phospholipids, and cholesterol. By treating these formulations as molecular graphs, where nodes are individual molecules and edges signify their interactions, GCNs predict encapsulation efficiency, stability during systemic circulation, and

intracellular release dynamics. These predictions guide the development of formulations that enhance mRNA protection against degradation and promote efficient delivery to target cells (Wang et al., 2022). Furthermore, GCNs analyze how LNPs interact with mRNA under varying cellular conditions, such as pH fluctuations or enzymatic activity, ensuring that delivery systems are optimized for endosomal escape and subsequent mRNA release. This facilitates the effective translation of mRNA into antigens, which is crucial for activating immune cells and orchestrating adaptive immune responses. GCNs also evaluate structural variations in mRNA sequences to design antigens and immune-stimulatory adjuvants with enhanced stability and immunogenicity (Kejani et al., 2020). Through their ability to integrate complex structural, chemical, and biological data, GCNs provide a comprehensive framework for addressing challenges in mRNA-LNP vaccine development. These advancements accelerate the creation of vaccines with precise delivery mechanisms and robust immune activation, contributing to innovations in cancer immunotherapy and infectious disease prevention (Wang et al., 2024).

9.3 Deep Neural Networks

Deep Neural Networks (DNN) are a critical tool for mRNA-LNP immunogenicity optimization because they can model the complex, nonlinear relationships between different variables involved in vaccine development (Chen et al., 2020; Mekki-Berrada et al., 2021). DNNs consist of multiple layers of neurons that learn increasingly abstract representations of input data, which makes them highly effective in identifying intricate patterns within high-dimensional datasets. For mRNA-LNP formulation optimization, DNNs can be trained to predict how specific lipid compositions, mRNA modifications, and nanoparticle characteristics influence immune responses, such as the activation of DCs, T-cells, and B-cells (Mekki-Berrada et al., 2021; Konstantopoulos et al., 2022).

To enhance vaccine efficacy, DNNs work by analyzing large datasets that include both formulation details (lipid ratios, LNP size, and composition) and immunological endpoints (antibody production, cytokine levels, T-cell activation). The neural network identifies the relationships between formulation characteristics and immune activation, providing researchers with precise recommendations for optimizing LNP formulations that yield the strongest immune response. This can include optimizing the lipid mixture for better cellular uptake, adjusting particle size to improve lymph node targeting, or modifying mRNA constructs to enhance antigen presentation (Zeng et al., 2024).

Recent advancements in DNNs, particularly with frameworks like TensorFlow and PyTorch, have allowed the integration of mRNA-LNP optimization with high-dimensional data sources, including cellular response profiles and *in vivo* animal data (Taylor and Kriegeskorte, 2023). These networks can predict how formulation changes will affect immune responses without the need for extensive trial-and-error testing, which significantly speeds up vaccine development. In immuno-oncology, DNNs can also predict LNP formulations that target cancer cells more efficiently by analyzing tumor microenvironment data, ensuring that the
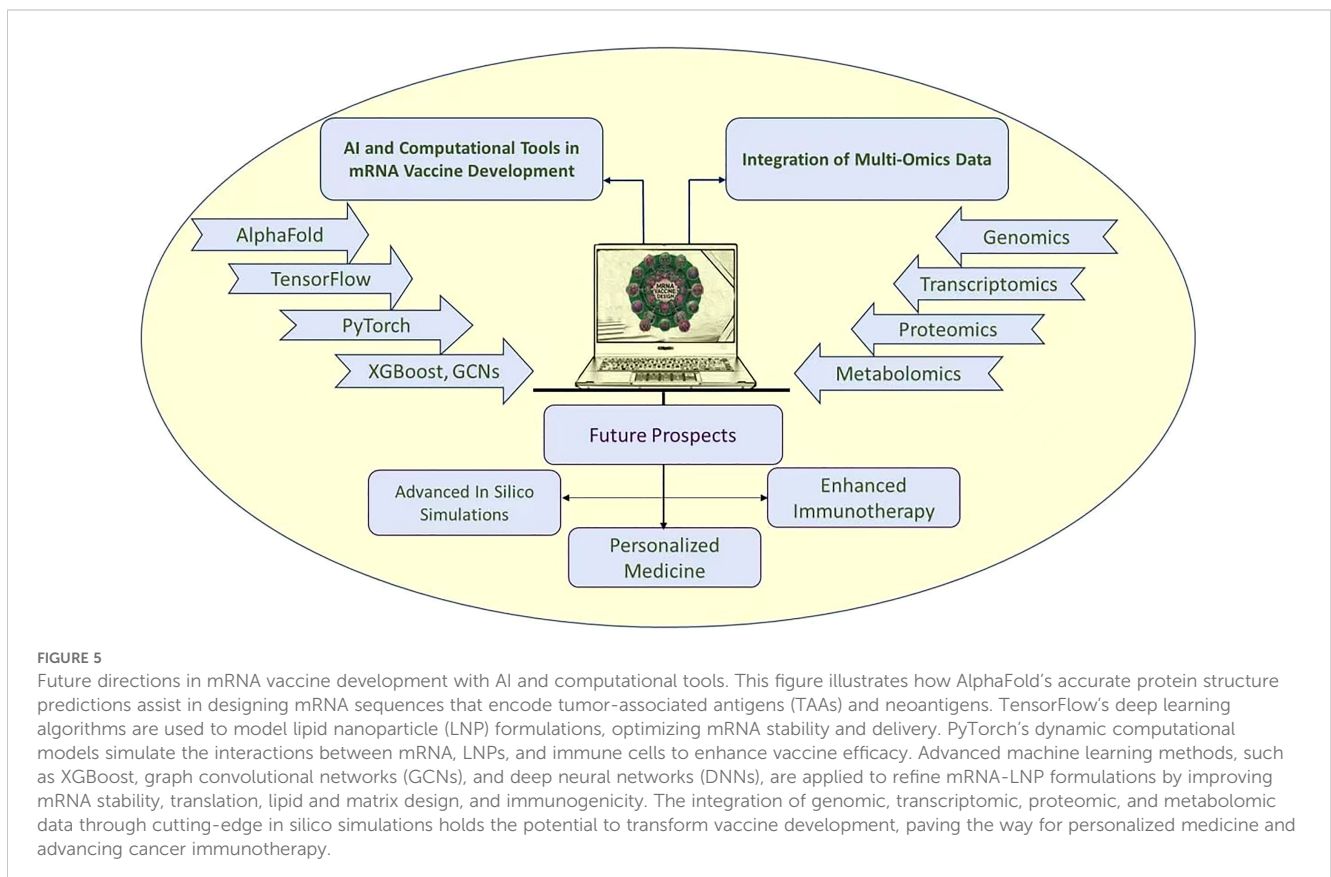
mRNA vaccines not only deliver genetic material but also elicit strong, targeted immune responses against tumor cells (Chen et al., 2020). For example, DNNs can optimize formulations for enhancing the presentation of cancer antigens to immune cells in the tumor microenvironment, significantly improving the therapeutic efficacy of mRNA-based cancer vaccines (Raza et al., 2023). Additionally, DNNs can fine-tune LNP formulations to induce specific immune pathways, such as the activation of Th1 responses, which are crucial for effective anti-cancer immunity. This predictive capability is essential for designing personalized mRNA vaccines, where DNNs can assist in tailoring the vaccine to an individual's unique immune profile, making these tools indispensable for both infectious disease and cancer immunotherapy (Jozwik et al., 2017; Castillo-Hair and Seelig, 2022).

9.4 Prospects

Undoubtedly, the future of mRNA vaccine development is advancing rapidly with the aid of advanced bioinformatics and AI tools. AlphaFold, developed by DeepMind, marks a significant advancement in predicting protein structures with high precision (Service, 2023). Figure 5 highlights key advancements in mRNA vaccine development through the integration of bioinformatics and AI tools. The figure underscores the transformative potential of these technologies in enhancing vaccine efficacy in TME and advancing personalized immunotherapy. By utilizing deep learning techniques such as CNNs and attention mechanisms, AlphaFold provides crucial insights into how proteins fold and assemble. This capability is vital for designing mRNA vaccines that encode TAAs and neoantigens, as it allows researchers to anticipate how these proteins will behave and interact within the body. TensorFlow and PyTorch play pivotal roles in advancing mRNA-LNP vaccine technology (Xu et al., 2023). TensorFlow utilizes advanced algorithms, such as CNNs and RNNs, to model the impact of different lipid formulations on mRNA stability and delivery. This capability is crucial for optimizing LNP designs, and ensuring effective mRNA delivery to DCs, macrophages, and CTLs. PyTorch, known for its dynamic computational graph, facilitates the creation of sophisticated models to simulate interactions between mRNA, LNPs, and immune cells (Zhou et al., 2020). This flexibility is instrumental in optimizing vaccine efficacy and enhancing immune responses.

Looking ahead, the fusion of in silico simulations with multi-omics data is poised to revolutionize vaccine development. AI-driven analysis of genomic (Kamimoto et al., 2023), transcriptomic, proteomic, and metabolomic data will enable the identification of key biomarkers and pathways, leading to more targeted and effective vaccine strategies. These advancements will address current challenges in optimizing LNP formulations and mRNA stability, improving both pharmacokinetics and pharmacodynamics. Ultimately, the integration of these cutting-edge technologies holds the promise of transforming vaccine development and expanding the potential of mRNA-based therapies in personalized medicine and beyond.

FIGURE 5
Future directions in mRNA vaccine development with AI and computational tools. This figure illustrates how AlphaFold's accurate protein structure predictions assist in designing mRNA sequences that encode tumor-associated antigens (TAAs) and neoantigens. TensorFlow's deep learning algorithms are used to model lipid nanoparticle (LNP) formulations, optimizing mRNA stability and delivery. PyTorch's dynamic computational models simulate the interactions between mRNA, LNPs, and immune cells to enhance vaccine efficacy. Advanced machine learning methods, such as XGBoost, graph convolutional networks (GCNs), and deep neural networks (DNNs), are applied to refine mRNA-LNP formulations by improving mRNA stability, translation, lipid and matrix design, and immunogenicity. The integration of genomic, transcriptomic, proteomic, and metabolomic data through cutting-edge in silico simulations holds the potential to transform vaccine development, paving the way for personalized medicine and advancing cancer immunotherapy.

10 Conclusion

The integration of computational biology, bioinformatics, and artificial intelligence is transforming mRNA vaccine development, enhancing their precision and effectiveness. By combining these advanced tools with machine learning, we gain deeper insights into protein structures and optimize LNP formulations and their interactions with immune cells. Additionally, incorporating MD simulations further improves our understanding of mRNA-LNPs' structure and their interactions with cellular machinery, providing critical insights into optimizing stability and translation efficiency. These *in silico* technologies are driving progress in personalized cancer immunotherapy and opening new avenues for addressing global health challenges with next-generation vaccines.

Author contributions

SI: Writing – original draft, Conceptualization, Formal analysis, Investigation, Methodology, and Validation. XL: Writing – original draft, Conceptualization, Formal analysis, Investigation, Methodology, and Validation. KC: Writing – original draft, Conceptualization, Formal analysis, Investigation, Methodology, and Validation. MM: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, and Writing – review & editing. MH: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, and Writing – review & editing. SK: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, and Writing – review & editing. PJK: Visualization, and Writing – review & editing. XL: Resources, Project administration, Funding acquisition, Supervision, Writing – review & editing. All authors contributed to the preparation, review, and editing of the manuscript. All authors read and approved the contents of the manuscript.

Conflict of interest

The 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.

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/fcimb.2024.1501010/full#supplementary-material

References

Abbasi, S., Matsui-Masai, M., Yasui, F., Hayashi, A., Tockary, T. A., Mochida, Y., et al. (2024). Carrier-free mRNA vaccine induces robust immunity against SARS-CoV-2 in mice and non-human primates without systemic reactogenicity. *Mol. Ther.* 32, 1266–1283. doi: 10.1016/j.ymthe.2024.03.022

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. doi: 10.1038/s41586-024-07487-w

Abumanhal-Masarweh, H., Da Silva, D., Poley, M., Zinger, A., Goldman, E., Krinsky, N., et al. (2019). Tailoring the lipid composition of nanoparticles modulates their cellular uptake and affects the viability of triple negative breast cancer cells. *J. Control. Release.* 307, 331–341. doi: 10.1016/j.jconrel.2019.06.025

Acun, B., Hardy, D. J., Kale, L. V., Li, K., Phillips, J. C., and Stone, J. E. (2018). Scalable molecular dynamics with NAMD on the summit system. *IBM. J. Res. Dev.* 62, 1–9. doi: 10.1147/JRD.2018.2888986

Alameh, M. G., Tombácz, I., Bettini, E., Lederer, K., Sittplangkoon, C., Wilmore, J. R., et al. (2021). Lipid nanoparticles enhance the efficacy of mRNA and protein subunit vaccines by inducing robust T follicular helper cell and humoral responses. *Immunity* 54, 2877–2892.e2877. doi: 10.1016/j.immuni.2021.11.001

Alburquerque-González, B., López-Abellán, M. D., Luengo-Gil, G., Montoro-García, S., and Conesa-Zamora, P. (2022). Design of personalized neoantigen RNA vaccines against cancer based on next-generation sequencing data. *Methods Mol. Biol.* 2547, 165–185. doi: 10.1007/978-1-0716-2573-6_7

Al Fayez, N., Nassar, M. S., Alshehri, A. A., Alnefaie, M. K., Almughem, F. A., Alshehri, B. Y., et al. (2023). Recent advancement in mRNA vaccine development and applications. *Pharmaceutics.* 15 (7), 1972–1979. doi: 10.3390/pharmaceutics15071972

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'meara, M. J., Dimaio, F. P., Park, H., et al. (2017). The rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* 13, 3031–3048. doi: 10.1021/acs.jctc.7b00125

Andreatta, M., and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32, 511–517. doi: 10.1093/bioinformatics/btv639

Ardui, S., Ameur, A., Vermeesch, J. R., and Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res.* 46, 2159–2168. doi: 10.1093/nar/gky066

Arévalo-Romero, J. A., López-Cantillo, G., Moreno-Jiménez, S., Marcos-Alcalde, Í., Ros-Pardo, D., Camacho, B. A., et al. (2024). In Silico Design of miniACE2 Decoys with *In Vitro* Enhanced Neutralization Activity against SARS-CoV-2, Encompassing Omicron Subvariants. *Int. J. Mol. Sci.* 25, 10802–10812. doi: 10.3390/ijms251910802

Arno, M. C., Inam, M., Weems, A. C., Li, Z., Binch, A. L. A., Platt, C. I., et al. (2020). Exploiting the role of nanoparticle shape in enhancing hydrogel adhesive and mechanical properties. *Nat. Commun.* 11, 1420. doi: 10.1038/s41467-020-15206-y

Asediya, V. S., Anjaria, P. A., Mathakiya, R. A., Koringa, P. G., Nayak, J. B., Bisht, D., et al. (2024). Vaccine development using artificial intelligence and machine learning: A review. *Int. J. Biol. Macromol.* 282, 136643. doi: 10.1016/j.ijbiomac.2024.136643

Baigorria, E., Souza Dos Santos, S., De Moura, M. R., and Fraceto, L. F. (2023). Nanocomposite hydrogels 3D printed for application in water remediation. *Mater. Today Chem.* 30, 101559. doi: 10.1016/j.mtchem.2023.101559

Bates, P. A., Kelley, L. A., Maccallum, R. M., and Sternberg, M. J. (2001). Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins* Suppl 5, 39–46. doi: 10.1002/(ISSN)1097-0134

Bell, D. R., Cheng, S. Y., Salazar, H., and Ren, P. (2017). Capturing RNA folding free energy with coarse-grained molecular dynamics simulations. *Sci. Rep.* 7, 45812. doi: 10.1038/srep45812

Bernhart, S. H., Tafer, H., Mückstein, U., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2006). Partition function and base pairing probabilities of RNA heterodimers. *Algorithms. Mol. Biol.* 1, 3. doi: 10.1186/1748-7188-1-3

Binet, T., Padiolleau-Lefèvre, S., Octave, S., Avalle, B., and Maffucci, I. (2023). Comparative study of single-stranded oligonucleotides secondary structure prediction tools. *BMC Bioinf.* 24, 422. doi: 10.1186/s12859-023-05532-5

Bock, L. V., Gabrielli, S., Kolář, M. H., and Grubmüller, H. (2023). Simulation of complex biomolecular systems: the ribosome challenge. *Annu. Rev. Biophys.* 52, 361–390. doi: 10.1146/annurev-biophys-111622-091147

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bon, M., Vernizzi, G., Orland, H., and Zee, A. (2008). Topological classification of RNA structures. *J. Mol. Biol.* 379, 900–911. doi: 10.1016/j.jmb.2008.04.033

Bravi, B. (2024). Development and use of machine learning algorithms in vaccine target selection. *NPJ Vaccines* 9, 15. doi: 10.1038/s41541-023-00795-8

Brooks, B. R., Brooks, C. L. 3rd, Mackerell, A. D. Jr., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi: 10.1002/jcc.v30:10

Byrgazov, K., Vesper, O., and Moll, I. (2013). Ribosome heterogeneity: another level of complexity in bacterial translation regulation. *Curr. Opin. Microbiol.* 16, 133–139. doi: 10.1016/j.mib.2013.01.009

Cai, X., Li, J. J., Liu, T., Brian, O., and Li, J. (2021). Infectious disease mRNA vaccines and a review on epitope prediction for vaccine design. *Brief Funct. Genomics* 20, 289–303. doi: 10.1093/bfgp/elab027

Calonaci, N., Jones, A., Cuturello, F., Sattler, M., and Bussi, G. (2020). Machine learning a model for RNA structure prediction. *NAR. Genom. Bioinform.* 2, lqaa090. doi: 10.1093/nargab/lqaa090

Cao, S., Liu, X., Li, X., Lin, C., Zhang, W., Tan, C. H., et al. (2020). Shape matters: comprehensive analysis of star-shaped lipid nanoparticles. *Front. Pharmacol.* 11, 539. doi: 10.3389/fphar.2020.00539

Capelli, R., Serapian, S. A., and Colombo, G. (2023). Computational epitope prediction and design for antibody development and detection. *Methods Mol. Biol.* 2552, 255–266. doi: 10.1007/978-1-0716-2609-2_13

Castillo-Hair, S. M., and Seelig, G. (2022). Machine learning for designing next-generation mRNA therapeutics. *Accounts. Chem. Res.* 55, 24–34. doi: 10.1021/acs.accounts.1c00621

Chakraborty, A., and Bandyopadhyay, S. (2013). FOGSAA: fast optimal global sequence alignment algorithm. *Sci. Rep.* 3, 1746. doi: 10.1038/srep01746

Chakraborty, C., Sharma, A. R., Bhattacharya, M., and Lee, S. S. (2021). From COVID-19 to cancer mRNA vaccines: moving from bench to clinic in the vaccine landscape. *Front. Immunol.* 12, 679344. doi: 10.3389/fimmu.2021.679344

Chen, C. C., and Chan, Y. M. (2023). REDfold: accurate RNA secondary structure prediction using residual encoder-decoder network. *BMC Bioinf.* 24, 122. doi: 10.1186/s12859-023-05238-8

Chen, X., Zhou, M., Gong, Z., Xu, W., Liu, X., Huang, T., et al. (2020). DNNBrain: A unifying toolbox for mapping deep neural networks and brains. *Front. Comput. Neurosci.* 14, 580632. doi: 10.3389/fncom.2020.580632

Cheng, Q., Wei, T., Farbiak, L., Johnson, L. T., Dilliard, S. A., and Siegwart, D. J. (2020). Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nat. Nanotechnol.* 15, 313–320. doi: 10.1038/s41565-020-0669-6

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Dhanda, S. K., Mahajan, S., Paul, S., Yan, Z., Kim, H., Jespersen, M. C., et al. (2019). IEDB-AR: immune epitope database-analysis resource in 2019. *Nucleic Acids Res.* 47, W502–w506. doi: 10.1093/nar/gkz452

Dorsey, P. J., Lau, C. L., Chang, T. C., Doerschuk, P. C., and D'addio, S. M. (2024). Review of machine learning for lipid nanoparticle formulation and process development. *J. Pharm. Sci.* 113 (12), 3413–3433. doi: 10.1016/j.xphs.2024.09.015

Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., et al. (2021). The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* 16, 5634–5651. doi: 10.1038/s41596-021-00628-9

Ege, H., Enguven, G., Gunduz, O., and Ege, Z. R. (2023). "18 - 3D printable nanocomposite hydrogels for biomedical applications," in *Functional nanocomposite hydrogels*. Eds. A. Kumar and V. K. Thakur (Dumfries, United Kingdom: Elsevier), 429–451.

El Khoury, G., Azzam, W., and Rebehmed, J. (2023). PyProtif: a PyMol plugin to retrieve and visualize protein motifs for structural studies. *Amino Acids* 55, 1429–1436. doi: 10.1007/s00726-023-03323-z

Erdem, Ö., Eş, I., Saylan, Y., Atabay, M., Gungen, M. A., Ölmez, K., et al. (2023). *In situ* synthesis and dynamic simulation of molecularly imprinted polymeric

nanoparticles on a micro-reactor system. *Nat. Commun.* 14, 4840. doi: 10.1038/s41467-023-40413-8

Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O., and Salzberg, S. L. (2000). Prediction of transcription terminators in bacterial genomes. *J. Mol. Biol.* 301, 27–33. doi: 10.1006/jmbi.2000.3836

Esprit, A., De Mey, W., Bahadur Shahi, R., Thielemans, K., Franceschini, L., and Breckpot, K. (2020). Neo-antigen mRNA vaccines. *Vaccines (Basel).* 8, 776–783. doi: 10.3390/vaccines8040776

Fu, H., Liang, Y., Zhong, X., Pan, Z., Huang, L., Zhang, H., et al. (2020). Codon optimization with deep learning to enhance protein expression. *Sci. Rep.* 10, 17617. doi: 10.1038/s41598-020-74091-z

Gao, Q., Xu, T., Li, X., Gao, W., Shi, H., Zhang, Y., et al. (2024). Interpretable dynamic directed graph convolutional network for multi-relational prediction of missense mutation and drug response. *IEEE J. BioMed. Health Inform.* 18 (2), 120–136. doi: 10.1109/JBHI.2024.3483316

Genc, A. G., and McGuffin, L. J. (2025). Beyond AlphaFold2: The Impact of AI for the Further Improvement of Protein Structure Prediction. *Methods Mol Biol* 2867, 121–139. doi: 10.1007/978-1-0716-4196-5_7

Gerber, H. P., Sibener, L. V., Lee, L. J., and Gee, M. H. (2020). Identification of antigenic targets. *Trends Cancer* 6, 299–318. doi: 10.1016/j.trecan.2020.01.002

Gote, V., Bolla, P. K., Kommineni, N., Butreddy, A., Nukala, P. K., Palakurthi, S. S., et al. (2023). A comprehensive review of mRNA vaccines. *Int. J. Mol. Sci.* 24, 2700–2736. doi: 10.3390/ijms24032700

Gray, M., Will, S., and Jabbari, H. (2024). SparseRNAfolD: optimized sparse RNA pseudoknot-free folding with dangle consideration. *Algorithms. Mol. Biol.* 19, 9. doi: 10.1186/s13015-024-00256-4

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., et al. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33, W526–W531. doi: 10.1093/nar/gki376

Gunter, H. M., Idrisoglu, S., Singh, S., Han, D. J., Ariens, E., Peters, J. R., et al. (2023). mRNA vaccine quality analysis using RNA sequencing. *Nat. Commun.* 14, 5663. doi: 10.1038/s41467-023-41354-y

Guo, L., and Huo, H. (2024). An efficient Burrows-Wheeler transform-based aligner for short read mapping. *Comput. Biol. Chem.* 110, 108050. doi: 10.1016/j.compbiolchem.2024.108050

Gurba-Bryśkiewicz, L., Maruszak, W., Smuga, D. A., Dubiel, K., and Wieczorek, M. (2023). Quality by design (QbD) and design of experiments (DOE) as a strategy for tuning lipid nanoparticle formulations for RNA delivery. *Biomedicines* 11, 2752–2779. doi: 10.3390/biomedicines11102752

Hald Albertsen, C., Kulkarni, J. A., Witzigmann, D., Lind, M., Petersson, K., and Simonsen, J. B. (2022). The role of lipid components in lipid nanoparticles for vaccines and gene therapy. *Adv. Drug Delivery Rev.* 188, 114416. doi: 10.1016/j.addr.2022.114416

Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* 19, 20–30. doi: 10.1038/nrm.2017.91

Hasan, M., Khatun, A., and Kogure, K. (2023). Intradermal Delivery of Naked mRNA Vaccines via Iontophoresis. *Pharmaceutics* 15, 2678–2695. doi: 10.3390/pharmaceutics15122678

Heilmann, N., Wolf, M., Kozlowska, M., Sedghamiz, E., Setzler, J., Brieg, M., et al. (2020). Sampling of the conformational landscape of small proteins with Monte Carlo methods. *Sci. Rep.* 10, 18211. doi: 10.1038/s41598-020-75239-7

Hofacker, I. L., Fekete, M., and Stadler, P. F. (2002). Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 319, 1059–1066. doi: 10.1016/S0022-2836(02)00308-X

Hollingsworth, S. A., and Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron* 99, 1129–1143. doi: 10.1016/j.neuron.2018.08.011

Hoseini, B., Jaafari, M. R., Golabpour, A., Momtazi-Borojeni, A. A., and Eslami, S. (2023). Optimizing nanoliposomal formulations: Assessing factors affecting entrapment efficiency of curcumin-loaded liposomes using machine learning. *Int. J. Pharm.* 646, 123414. doi: 10.1016/j.ijpharm.2023.123414

Hoseini, B., Jaafari, M. R., Golabpour, A., Rahmatinejad, Z., Karimi, M., and Eslami, S. (2024). Machine learning-driven advancements in liposomal formulations for targeted drug delivery: A narrative literature review. *Curr. Drug Deliv.* 27 (3). doi: 10.2174/0115672018302321240620072039

Hudson, W. H., and Sudmeier, L. J. (2022). Localization of T cell clonotypes using the Visium spatial transcriptomics platform. *STAR. Protoc.* 3, 101391. doi: 10.1016/j.xpro.2022.101391

Igyártó, B. Z., and Qin, Z. (2024). The mRNA-LNP vaccines - the good, the bad and the ugly? *Front. Immunol.* 15, 1336906. doi: 10.3389/fimmu.2024.1336906

Imani, S., Tagit, O., and Pichon, C. (2024). Neoantigen vaccine nanoformulations based on Chemically synthesized minimal mRNA (CmRNA): small molecules, big impact. *NPJ Vaccines* 9, 14. doi: 10.1038/s41541-024-00807-1

Jabbari, H., and Condon, A. (2014). A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures. *BMC Bioinf.* 15, 147. doi: 10.1186/1471-2105-15-147

Jeeva, S., Kim, K. H., Shin, C. H., Wang, B. Z., and Kang, S. M. (2021). An update on mRNA-based viral vaccines. *Vaccines (Basel).* 9, 965–982. doi: 10.3390/vaccines9090965

Jia, H., Tan, S., Cai, Y., Guo, Y., Shen, J., Zhang, Y., et al. (2024). Low-input PacBio sequencing generates high-quality individual fly genomes and characterizes mutational processes. *Nat. Commun.* 15, 5644. doi: 10.1038/s41467-024-49992-6

Jiang, Y., Wang, R., Feng, J., Jin, J., Liang, S., Li, Z., et al. (2023). Explainable deep hypergraph learning modeling the peptide secondary structure prediction. *Adv. Sci. (Weinh).* 10, e2206151. doi: 10.1002/advs.202206151

Jo, S., Cheng, X., Lee, J., Kim, S., Park, S. J., Patel, D. S., et al. (2017). CHARMM-GUI 10 years for biomolecular modeling and simulation. *J. Comput. Chem.* 38, 1114–1124. doi: 10.1002/jcc.v38.15

Jozwik, K. M., Kriegeskorte, N., Storrs, K. R., and Mur, M. (2017). Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Front. Psychol.* 8, 1726. doi: 10.3389/fpsyg.2017.01726

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* 199, 3360–3368. doi: 10.4049/jimmunol.1700893

Kamimoto, K., Stringa, B., Hoffmann, C. M., Jindal, K., Solnica-Krezel, L., and Morris, S. A. (2023). Dissecting cell identity via network inference and in silico gene perturbation. *Nature* 614, 742–751. doi: 10.1038/s41586-022-05688-9

Kato, Y., Sato, K., Asai, K., and Akutsu, T. (2012). Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming. *Nucleic Acids Res.* 40, W29–W34. doi: 10.1093/nar/gks412

Keel, B. N., and Snelling, W. M. (2018). Comparison of burrows-wheeler transform-based mapping algorithms used in high-throughput whole-genome sequencing: application to illumina data for livestock genomes. *Front. Genet.* 9, 35. doi: 10.3389/fgene.2018.00035

Kejani, M. T., Dornaika, F., and Talebi, H. (2020). Graph Convolution Networks with manifold regularization for semi-supervised learning. *Neural Netw.* 127, 160–167. doi: 10.1016/j.neunet.2020.04.016

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., et al. (2012). Immune epitope database analysis resource. *Nucleic Acids Res.* 40, W525–W530. doi: 10.1093/nar/gks438

Klausen, M. S., Anderson, M. V., Jespersen, M. C., Nielsen, M., and Marcatili, P. (2015). LYRA, a webserver for lymphocyte receptor structural modeling. *Nucleic Acids Res.* 43, W349–W355. doi: 10.1093/nar/gkv535

Koehler Leman, J., and Künze, G. (2023). Recent advances in NMR protein structure prediction with ROSETTA. *Int. J. Mol. Sci.* 24, 7835–7857. doi: 10.3390/ijms24097835

Kohnke, B., Kutzner, C., and Grubmüller, H. (2020). A GPU-accelerated fast multipole method for GROMACS: performance and accuracy. *J. Chem. Theory Comput.* 16, 6938–6949. doi: 10.1021/acs.jctc.0c00744

Konstantopoulos, G., Koumoulos, E. P., and Charitidis, C. A. (2022). Digital innovation enabled nanomaterial manufacturing; machine learning strategies and green perspectives. *Nanomater. (Basel).* 12, 2646–2703. doi: 10.3390/nano12152646

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Brunak, S., Lund, O., et al. (2005). An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* 35, 2295–2303. doi: 10.1002/(ISSN)1521-4141

Lee, J., Oldham, M. L., Manon, V., and Chen, J. (2024a). Principles of peptide selection by the transporter associated with antigen processing. *Proc. Natl. Acad. Sci. U.S.A.* 121, e2320879121. doi: 10.1073/pnas.2320879121

Lee, S., Yan, S., Dey, A., Laederach, A., and Schlick, T. (2024b). An intricate balancing act: Upstream and downstream frameshift co-regulatory elements. *bioRxiv.* 27 (06), 599960–599989. doi: 10.1101/2024.06.27.599960

Lemay, M. A., Sibbesen, J. A., Torkamaneh, D., Hamel, J., Levesque, R. C., and Belzile, F. (2022). Combined use of Oxford Nanopore and Illumina sequencing yields insights into soybean structural variation biology. *BMC Biol.* 20, 53. doi: 10.1186/s12915-022-01255-w

Leppek, K., Byeon, G. W., Kladwang, W., Wayment-Steele, H. K., Kerr, C. H., Xu, A. F., et al. (2022). Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat. Commun.* 13, 1536. doi: 10.1038/s41467-022-28776-w

Lerminiaux, N., Fakharuddin, K., Mulvey, M. R., and Mataseje, L. (2024). Do we still need Illumina sequencing data? Evaluating Oxford Nanopore Technologies R10.4.1 flow cells and the Rapid v14 library prep kit for Gram negative bacteria whole genome assemblies. *Can. J. Microbiol.* 70, 178–189. doi: 10.1139/cjm-2023-0175

Li, S., Hu, Y., Li, A., Lin, J., Hsieh, K., Schneiderman, Z., et al. (2022a). Payload distribution and capacity of mRNA lipid nanoparticles. *Nat. Commun.* 13, 5561. doi: 10.1038/s41467-022-33157-4

Li, D. F., Liu, Q. S., Yang, M. F., Xu, H. M., Zhu, M. Z., Zhang, Y., et al. (2023). Nanomaterials for mRNA-based therapeutics: Challenges and opportunities. *Bioeng. Transl. Med.* 8, e10492. doi: 10.1002/btm2.10492

Li, X., Ma, S., Gao, T., Mai, Y., Song, Z., and Yang, J. (2022b). The main battlefield of mRNA vaccine - Tumor immune microenvironment. *Int. Immunopharmacol.* 113, 109367. doi: 10.1016/j.intimp.2022.109367

Li, B., Raji, I. O., Gordon, A. G. R., Sun, L., Raimondo, T. M., Oladimeji, F. A., et al. (2024). Accelerating ionizable lipid discovery for mRNA delivery using machine learning and combinatorial chemistry. *Nat. Mater.* 23, 1002–1008. doi: 10.1038/s41563-024-01867-3

Lindert, S., Meiler, J., and Mccammon, J. A. (2013). Iterative molecular dynamics-rosetta protein structure refinement protocol to improve model quality. *J. Chem. Theory Comput.* 9, 3843–3847. doi: 10.1021/ct400260c

Lorentzen, C. L., Haanen, J. B., Met, Ö., and Svane, I. M. (2022). Clinical advances and ongoing trials on mRNA vaccines for cancer treatment. *Lancet Oncol.* 23, e450–e458. doi: 10.1016/S1470-2045(22)00372-2

Luo, Z., Ma, Q., Sun, S., Li, N., Wang, H., Ying, Z., et al. (2023). Exon-intron boundary inhibits m(6)A deposition, enabling m(6)A distribution hallmark, longer mRNA half-life and flexible protein coding. *Nat. Commun.* 14, 4172. doi: 10.1038/s41467-023-39897-1

Ly, H. H., Daniel, S., Soriano, S. K. V., Kis, Z., and Blakney, A. K. (2022). Optimization of lipid nanoparticles for saRNA expression and cellular activation using a design-of-experiment approach. *Mol. Pharm.* 19, 1892–1905. doi: 10.1021/acs.molpharmaceut.2c00032

Maharjan, R., Kim, K. H., Lee, K., Han, H.-K., and Jeong, S. H. (2024). Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches. *J. Pharm. Anal.* 14 (11), 100996. doi: 10.1016/j.jpha.2024.100996

Marchand, B., Will, S., Berkemer, S. J., Ponty, Y., and Bulteau, L. (2023). Automated design of dynamic programming schemes for RNA folding with pseudoknots. *Algorithms. Mol. Biol.* 18, 18. doi: 10.1186/s13015-023-00229-z

Martí-Centelles, V., Piskorz, T. K., and Duarte, F. (2024). CageCavityCalc (C3): A computational tool for calculating and visualizing cavities in molecular cages. *J. Chem. Inf. Model.* 64, 5604–5616. doi: 10.1021/acs.jcim.4c00355

May, M. (2024). How mRNA is powering a personalized vaccine revolution. *Nat. Med.* 30, 2097–2098. doi: 10.1038/d41591-024-00052-y

Mekki-Berrada, F., Ren, Z., Huang, T., Wong, W. K., Zheng, F., Xie, J., et al. (2021). Two-step machine learning enables optimized nanoparticle synthesis. *NPJ Comput. Mater.* 7, 55. doi: 10.1038/s41524-021-00520-w

Melo Ferreira, R., Sabo, A. R., Winfree, S., Collins, K. S., Janosevic, D., Gulbronson, C. J., et al. (2021). Integration of spatial and single-cell transcriptomics localizes epithelial cell-immune cross-talk in kidney injury. *JCI Insight* 6, 147703–147724. doi: 10.1172/jci.insight.147703

Mikhailovskii, O., Izmailov, S. A., Xue, Y., Case, D. A., and Skrynnikov, N. R. (2024). X-ray crystallography module in MD simulation program amber 2023. Refining the models of protein crystals. *J. Chem. Inf. Model.* 64, 18–25. doi: 10.1021/acs.jcim.3c01531

Mikhailovskii, O., Xue, Y., and Skrynnikov, N. R. (2022). Modeling a unit cell: crystallographic refinement procedure using the biomolecular MD simulation platform Amber. *IUCrJ* 9, 114–133. doi: 10.1107/S2052252521011891

Mooers, B. H. M. (2020). Shortcuts for faster image creation in PyMOL. *Protein Sci.* 29, 268–272. doi: 10.1002/pro.v29.1

Morris, V. K., and Kopetz, S. (2022). Don't blame the messenger: lessons learned for cancer mRNA vaccines during the COVID-19 pandemic. *Nat. Rev. Cancer* 22, 317–318. doi: 10.1038/s41568-022-00463-4

Mrazek, J., Toso, D., Ryazantsev, S., Zhang, X., Zhou, Z. H., Fernandez, B. C., et al. (2014). Polyribosomes are molecular 3D nanoprinters that orchestrate the assembly of vault particles. *ACS Nano.* 8, 11552–11559. doi: 10.1021/nn504778h

Nicot, F., Trémeaux, P., Latour, J., Jeanne, N., Ranger, N., Raymond, S., et al. (2023). Whole-genome sequencing of SARS-CoV-2: Comparison of target capture and amplicon single molecule real-time sequencing protocols. *J. Med. Virol.* 95, e28123. doi: 10.1002/jmv.28123

Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* 8, 33. doi: 10.1186/s13073-016-0288-x

Oberhauser, N., Nurisso, A., and Carrupt, P. A. (2014). MLP Tools: a PyMOL plugin for using the molecular lipophilicity potential in computer-aided drug design. *J. Comput. Aided. Mol. Des.* 28, 587–596. doi: 10.1007/s10822-014-9744-0

Oeck, S., Tüns, A. I., and Schramm, A. (2023). Nanopore sequencing techniques: A comparison of the minKNOW and the alignator sequencers. *Methods Mol. Biol.* 2649, 209–221. doi: 10.1007/978-1-0716-3072-3_10

Oladipo, E. K., Adeyemo, S. F., Akinboade, M. W., Akinleye, T. M., Siyanbola, K. F., Adeogun, P. A., et al. (2024). Utilizing Immunoinformatics for mRNA Vaccine Design against Influenza D Virus. *BioMedInformatics* 4, 1572–1588. doi: 10.3390/biomedinformatics4020086

Olawade, D. B., Teke, J., Fapohunda, O., Weerasinghe, K., Usman, S. O., Ige, A. O., et al. (2024). Leveraging artificial intelligence in vaccine development: A narrative review. *J. Microbiol. Methods* 224, 106998. doi: 10.1016/j.mimet.2024.106998

Pardi, N., Hogan, M. J., Porter, F. W., and Weissman, D. (2018). mRNA vaccines - a new era in vaccinology. *Nat. Rev. Drug Discovery* 17, 261–279. doi: 10.1038/nrd.2017.243

Pechlaner, M., Oostenbrink, C., and Van Gunsteren, W. F. (2021). On the use of multiple-time-step algorithms to save computing effort in molecular dynamics simulations of proteins. *J. Comput. Chem.* 42, 1263–1282. doi: 10.1002/jcc.v42.18

Pennisi, F., Genovese, C., and Gianfredi, V. (2024). Lessons from the COVID-19 pandemic: promoting vaccination and public health resilience, a narrative review. *Vaccines (Basel).* 12, 891–904. doi: 10.3390/vaccines12080891

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.v26:16

Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., et al. (2020). Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* 153, 044130. doi: 10.1063/5.0014475

Politis, S. N., Colombo, P., Colombo, G., and Rekkas, D. M. (2017). Design of experiments (DoE) in pharmaceutical development. *Drug Dev. Ind. Pharm.* 43, 889–901. doi: 10.1080/03639045.2017.1291672

Porollo, A. A., Adamczak, R., and Meller, J. (2004). POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins. *Bioinformatics* 20, 2460–2462. doi: 10.1093/bioinformatics/bth248

Porollo, A., and Meller, J. (2007). Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D. *BMC Bioinf.* 8, 316. doi: 10.1186/1471-2105-8-316

Porollo, A., and Meller, J. (2010). POLYVIEW-MM: web-based platform for animation and analysis of molecular simulations. *Nucleic Acids Res.* 38, W662–W666. doi: 10.1093/nar/gkq445

Postle, K., and Good, R. F. (1985). A bidirectional rho-independent transcription terminator between the E. coli tonB gene and an opposing gene. *Cell* 41, 577–585. doi: 10.1016/S0092-8674(85)80030-1

Prjibelski, A. D., Mikheenko, A., Joglekar, A., Smetanin, A., Jarroux, J., Lapidus, A. L., et al. (2023). Accurate isoform discovery with IsoQuant using long reads. *Nat. Biotechnol.* 41, 915–918. doi: 10.1038/s41587-022-01565-y

Rajan-Babu, I. S., Peng, J. J., Chiu, R., Li, C., Mohajeri, A., Dolzhenko, E., et al. (2021). Genome-wide sequencing as a first-tier screening test for short tandem repeat expansions. *Genome Med.* 13, 126. doi: 10.1186/s13073-021-00932-9

Ramirez, F., Zambrano, A., Hennis, R., Holland, N., Lakshmanaswamy, R., and Chacon, J. (2023). Sending a message: use of mRNA vaccines to target the tumor immune microenvironment. *Vaccines (Basel).* 11 (9), 1465–1478. doi: 10.3390/vaccines11091465

Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A., and Stevanović, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50, 213–219. doi: 10.1007/s002510050595

Rampado, R., and Peer, D. (2023). Design of experiments in the optimization of nanoparticle-based drug delivery systems. *J. Control. Release.* 358, 398–419. doi: 10.1016/j.jconrel.2023.05.001

Rawat, R., Kant, K., Kumar, A., Bhati, K., and Verma, S. M. (2021). HeroMDAnalysis: an automagical tool for GROMACS-based molecular dynamics simulation analysis. *Future Med. Chem.* 13, 447–456. doi: 10.4155/fmc-2020-0191

Raza, A., Chohan, T. A., Buabeid, M., Arafa, E. A., Chohan, T. A., Fatima, B., et al. (2023). Deep learning in drug discovery: a futuristic modality to materialize the large datasets for cheminformatics. *J. Biomol. Struct. Dyn.* 41, 9177–9192. doi: 10.1080/07391102.2022.2136244

Rhoads, A., and Au, K. F. (2015). PacBio sequencing and its applications. *Genomics Proteomics Bioinf.* 13, 278–289. doi: 10.1016/j.gpb.2015.08.002

Rigsby, R. E., and Parker, A. B. (2016). Using the PyMOL application to reinforce visual understanding of protein structure. *Biochem. Mol. Biol. Educ.* 44, 433–437. doi: 10.1002/bmb.20966

Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol.* 383, 66–93. doi: 10.1016/S0076-6879(04)83004-0

Salatin, S., Maleki Dizaj, S., and Yari Khosroushahi, A. (2015). Effect of the surface modification, size, and shape on cellular uptake of nanoparticles. *Cell Biol. Int.* 39, 881–890. doi: 10.1002/cbin.10459

Sample, P. J., Wang, B., Reid, D. W., Presnyak, V., Mcfadyen, I. J., Morris, D. R., et al. (2019). Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* 37, 803–809. doi: 10.1038/s41587-019-0164-5

Sarkar, A., Santoro, J., Di Biasi, L., Marrafino, F., and Piotto, S. (2022). YAMACS: a graphical interface for GROMACS. *Bioinformatics* 38, 4645–4646. doi: 10.1093/bioinformatics/btac573

Sato, K., Akiyama, M., and Sakakibara, Y. (2021). RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.* 12, 941. doi: 10.1038/s41467-021-21194-4

Sato, K., Kato, Y., Hamada, M., Akutsu, T., and Asai, K. (2011). IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* 27, i85–i93. doi: 10.1093/bioinformatics/btr215

Sato, S., Sano, S., Muto, H., Kubara, K., Kondo, K., Miyazaki, T., et al. (2024). Understanding the manufacturing process of lipid nanoparticles for mRNA delivery using machine learning. *Chem. Pharm. Bull. (Tokyo).* 72, 529–539. doi: 10.1248/cpb.c24-00089

Sayour, E. J., Boczkowski, D., Mitchell, D. A., and Nair, S. K. (2024). Cancer mRNA vaccines: clinical advances and future opportunities. *Nat. Rev. Clin. Oncol.* 21, 489–500. doi: 10.1038/s41571-024-00902-1

Schmitz, S., Ertelt, M., Merkl, R., and Meiler, J. (2021). Rosetta design with co-evolutionary information retains protein function. *PloS Comput. Biol.* 17, e1008568. doi: 10.1371/journal.pcbi.1008568

Schoeder, C. T., Schmitz, S., Adolf-Bryfogle, J., Sevy, A. M., Finn, J. A., Sauer, M. F., et al. (2021). Modeling immunity with rosetta: methods for antibody and antigen design. *Biochemistry* 60, 825–846. doi: 10.1021/acs.biochem.0c00912

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi: 10.1038/nature10098

Sebastiani, F., Yanez Arteta, M., Lerche, M., Porcar, L., Lang, C., Bragg, R. A., et al. (2021). Apolipoprotein E binding drives structural and compositional rearrangement of mRNA-containing lipid nanoparticles. *ACS Nano.* 15, 6709–6722. doi: 10.1021/acsnano.0c10064

Seeliger, D., and De Groot, B. L. (2010). Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided. Mol. Des.* 24, 417–422. doi: 10.1007/s10822-010-9352-6

Service, R. F. (2023). DeepMind predicts millions of new materials. *Science* 382, 987. doi: 10.1126/science.adn2116

Shao, Q., and Zhu, W. (2018). Assessing AMBER force fields for protein folding in an implicit solvent. *Phys. Chem. Chem. Phys.* 20, 7206–7216. doi: 10.1039/C7CP08010G

Shen, C., Zhang, X., Hsieh, C. Y., Deng, Y., Wang, D., Xu, L., et al. (2023). A generalized protein-ligand scoring framework with balanced scoring, docking, ranking and screening powers. *Chem. Sci.* 14, 8129–8146. doi: 10.1039/D3SC02044D

Shuptrine, C. W., Chen, Y., Miriyala, J., Lenz, K., Moffett, D., Nguyen, T. A., et al. (2024). Lipid-encapsulated mRNAs encoding complex fusion proteins potentiate antitumor immune responses. *Cancer Res.* 84, 1550–1559. doi: 10.1158/0008-5472.CAN-23-2875

Sklenářová, H., Rosecká, M., Horstkotte, B., Pávek, P., Miró, M., and Solich, P. (2021). 3D printed permeation module to monitor interaction of cell membrane transporters with exogenic compounds in real-time. *Anal. Chim. Acta* 1153, 338296. doi: 10.1016/j.aca.2021.338296

Smith, K. D., Prince, D. K., Macdonald, J. W., Bammler, T. K., and Akilesh, S. (2024). Challenges and opportunities for the clinical translation of spatial transcriptomics technologies. *Glomerular. Dis.* 4, 49–63. doi: 10.1159/000538344

Solheim, J. C., Cook, J. R., and Hansen, T. H. (1995). Conformational changes induced in the MHC class I molecule by peptide and beta 2-microglobulin. *Immunol. Res.* 14, 200–217. doi: 10.1007/BF02918217

Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., and Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J. BioMed. Inform* 53, 405–414. doi: 10.1016/j.jbi.2014.11.003

Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82. doi: 10.1126/science.aaf2403

Stefan, C. P., Hall, A. T., Graham, A. S., and Minogue, T. D. (2022). Comparison of illumina and oxford nanopore sequencing technologies for pathogen detection from clinical matrices using molecular inversion probes. *J. Mol. Diagn.* 24, 395–405. doi: 10.1016/j.jmoldx.2021.12.005

Su, J., Lui, W. W., Lee, Y., Zheng, Z., Siu, G. K., Ng, T. T., et al. (2023). Evaluation of Mycobacterium tuberculosis enrichment in metagenomic samples using ONT adaptive sequencing and amplicon sequencing for identification and variant calling. *Sci. Rep.* 13, 5237. doi: 10.1038/s41598-023-32378-x

Tavares Luiz, M., Santos Rosa Viegas, J., Palma Abriata, J., Viegas, F., Testa Moura De Carvalho Vicentini, F., Lopes Badra Bentley, M. V., et al. (2021). Design of experiments (DoE) to develop and to optimize nanoparticles as drug delivery systems. *Eur. J. Pharm. Biopharm.* 165, 127–148. doi: 10.1016/j.ejpb.2021.05.011

Taylor, J., and Kriegeskorte, N. (2023). Extracting and visualizing hidden activations and computational graphs of PyTorch models with TorchLens. *Sci. Rep.* 13, 14375. doi: 10.1038/s41598-023-40807-0

Towne, J., Carter, N., and Neivandt, D. J. (2021). COMSOL Multiphysics® modelling of oxygen diffusion through a cellulose nanofibril conduit employed for peripheral nerve repair. *BioMed. Eng. Online* 20, 60. doi: 10.1186/s12938-021-00897-1

Toyama, S., Honda, T., Iwabuchi, H., Hashimoto, S., Yamaji, K., Tokunaga, Y., et al. (2023). Application of spatial transcriptomics analysis using the Visium system for the mouse nasal cavity after intranasal vaccination. *Front. Immunol.* 14, 1209945. doi: 10.3389/fimmu.2023.1209945

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439–d444. doi: 10.1093/nar/gkab1061

Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., et al. (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* 52, D368–d375. doi: 10.1093/nar/gkad1011

Venable, R. M., Krämer, A., and Pastor, R. W. (2019). Molecular dynamics simulations of membrane permeability. *Chem. Rev.* 119, 5954–5997. doi: 10.1021/acs.chemrev.8b00486

Vishweshwaraiah, Y. L., and Dokholyan, N. V. (2022). mRNA vaccines for cancer immunotherapy. *Front. Immunol.* 13, 1029069. doi: 10.3389/fimmu.2022.1029069

Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., et al. (2018). The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* 47, D339–D343. doi: 10.1093/nar/gky1006

Wang, Z. J., and Deserno, M. (2010). A systematically coarse-grained solvent-free model for quantitative phospholipid bilayer simulations. *J. Phys. Chem. B.* 114, 11207–11220. doi: 10.1021/jp102543j

Wang, Z., Ding, H., Pan, L., Li, J., Gong, Z., and Yu, P. S. (2024). From cluster assumption to graph convolution: graph-based semi-supervised learning revisited. *IEEE Trans. Neural Netw. Learn. Syst.* 7 (2), 1–34. doi: 10.1109/TNNLS.2024.3454710

Wang, W., Feng, S., Ye, Z., Gao, H., Lin, J., and Ouyang, D. (2022). Prediction of lipid nanoparticles for mRNA vaccines by the machine learning algorithm. *Acta Pharm. Sin. B.* 12, 2950–2962. doi: 10.1016/j.apsb.2021.11.021

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., and Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* 39, 1348–1365. doi: 10.1038/s41587-021-01108-x

Watson, J. D., Laskowski, R. A., and Thornton, J. M. (2005). Predicting protein function from sequence and structural data. *Curr. Opin. Struct. Biol.* 15, 275–284. doi: 10.1016/j.sbi.2005.04.003

Wick, R. R., Judd, L. M., and Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20, 129. doi: 10.1186/s13059-019-1727-y

Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R., and Haque, A. (2022). An introduction to spatial transcriptomics for biomedical research. *Genome Med.* 14, 68. doi: 10.1186/s13073-022-01075-1

Wu, X., Li, T., Jiang, R., Yang, X., Guo, H., and Yang, R. (2023). Targeting MHC-I molecules for cancer: function, mechanism, and therapeutic prospects. *Mol. Cancer* 22, 194. doi: 10.1186/s12943-023-01899-4

Xu, L., Hu, Y., and Liu, W. (2022). Tumor microenvironment-mediated immune profiles characterized by distinct survival outcome and immunotherapeutic efficacy in breast cancer. *Front. Genet.* 13, 840348. doi: 10.3389/fgene.2022.840348

Xu, Z., Langie, S. A., De Boever, P., Taylor, J. A., and Niu, L. (2017). RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip. *BMC Genomics* 18, 4. doi: 10.1186/s12864-016-3426-3

Xu, G., Luo, Z., Zhou, R., Wang, Q., and Ma, J. (2023). OPUS-Fold3: a gradient-based protein all-atom folding and docking framework on TensorFlow. *Brief Bioinform.* 24 (6). doi: 10.1093/bib/bbad365

Xue, L., Yamazaki, H., Ren, R., Wanunu, M., Ivanov, A. P., and Edel, J. B. (2020). Solid-state nanopore sensors. *Nat. Rev. Mater.* 5, 931–951. doi: 10.1038/s41578-020-0229-6

Yan, Z., Kim, K., Kim, H., Ha, B., Gambiez, A., Bennett, J., et al. (2024). Next-generation IEDB tools: a platform for epitope prediction and analysis. *Nucleic Acids Res.* 52, W526–w532. doi: 10.1093/nar/gkae407

Yao, C., Ning, J., Liu, H., Lv, Y., Zhao, Y., and Dong, S. (2020). Nanosecond pulses targeting intracellular ablation increase destruction of tumor cells with irregular morphology. *Bioelectrochemistry* 132, 107432. doi: 10.1016/j.bioelechem.2019.107432

Yao, R., Xie, C., and Xia, X. (2024). Recent progress in mRNA cancer vaccines. *Hum. Vaccin. Immunother.* 20, 2307187. doi: 10.1080/21645515.2024.2307187

Yong, C. W. (2015). Study of interactions between polymer nanoparticles and cell membranes at atomistic levels. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, 20140036. doi: 10.1098/rstb.2014.0036

Zeng, X., Wei, Z., Du, Q., Li, J., Xie, Z., and Wang, X. (2024). Unveil cis-acting combinatorial mRNA motifs by interpreting deep neural network. *Bioinformatics* 40, i381–i389. doi: 10.1093/bioinformatics/btae262

Zhang, T., Liu, L., Huang, X., Gao, X., Huan, X., He, C., et al. (2024). The rapid change of shear rate gradient is beneficial to platelet activation. *Platelets* 35, 2288679. doi: 10.1080/09537104.2023.2288679

Zhang, J., Sun, B., Shen, W., Wang, Z., Liu, Y., Sun, Y., et al. (2023b). In silico analyses, experimental verification and application in DNA vaccines of ebolavirus GP-derived pan-MHC-II-restricted epitopes. *Vaccines (Basel).* 11, 1620–1640. doi: 10.3390/vaccines11101620

Zhang, H., Zhang, L., Lin, A., Xu, C., Li, Z., Liu, K., et al. (2023a). Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature* 621, 396–403. doi: 10.1038/s41586-023-06127-z

Zhou, X., Jia, X., Huang, Z., Yang, C., Li, J., Xie, W., et al. (2023). MHC class II regulation of CD8(+) T cell tolerance and implications in autoimmunity and cancer immunotherapy. *Cell Rep.* 42, 113452. doi: 10.1016/j.celrep.2023.113452

Zhou, G., Nebgen, B., Lubbers, N., Malone, W., Niklasson, A. M. N., and Tretiak, S. (2020). Graphics processing unit-accelerated semiempirical born oppenheimer molecular dynamics using pytorch. *J. Chem. Theory Comput.* 16, 4951–4962. doi: 10.1021/acs.jctc.0c00243

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 9 (1), 133-148. doi: 10.1093/nar/9.1.133