

Application of Artificial Intelligence in the Design of Tumor mRNA Vaccines

Xiaohe He¹,
Pengyu Huang²,
Zubaire Shalai^{3,*},
Ruiwen Xie⁴

¹ International Department,
Guangdong County Garden School,
Guangzhou, Guangdong, China

² Department of Pharmaceutical
Engineering, School of Chemical
Engineering, Chongqing University
of Science and Technology,
Chongqing, China.

³ Department of Biotechnology,
School of Food and Biological
Engineering, Jiangsu University,
Zhenjiang, Jiangsu, China.

⁴ International Department, Yali
Senior High School, Yuhua District,
Changsha, Hunan, China.

*Corresponding author:
3220904009@stmail.ujs.edu.cn

Abstract:

For many years, scientists have been focused on developing mRNA-based vaccines, an area that has transitioned from theoretical concepts to practical treatments in clinical settings. This remarkable progress highlights the speed at which scientific advancements can occur when collaboration and innovation intersect. mRNA vaccines offer several inherent advantages, such as robust immunogenicity, a lack of gene integration risks, and the potential for cost-effective production. However, challenges related to mRNA stability and degradation continue to pose significant hurdles. To address these challenges and enhance vaccine development, traditional trial-and-error methods are increasingly being complemented and, in some cases, replaced by rational design strategies that leverage artificial intelligence (AI). In this review, we explore the core technologies underpinning mRNA vaccines and discuss how AI-driven design methods are optimizing their development. We also present a case study of trRosettaRNA, a deep learning-based approach for automating the prediction of RNA three-dimensional structures. This method successfully predicts the 3D structure of target RNA, demonstrating the transformative potential of AI in optimizing vaccine design and further advancing mRNA-based therapeutic strategies.

Keywords: mRNA vaccine; artificial intelligence; AI-driven design.

1. Introduction

In recent years, mRNA tumor vaccines have rapidly become one of the most promising directions in the field of tumor individualized immunotherapy due to its multiple advantages, such as short preparation cycle, high safety, and simultaneous activation of humoral and cellular immunity. However, from the

precise screening of tumor new antigen, rational optimization of mRNA sequence, to the innovative design of drug delivery system, every step in the whole technology chain is full of challenges. Firstly, the high heterogeneity of tumors makes the number of potential antigens huge and persistent mutations, which is difficult to be fully covered by traditional experimental methods. Secondly, the mRNA mol-

ecule itself is easy to degrade. Its coding region (CDS) sequence space is exponentially expanded. If the experimental strategy of one-by-one verification is adopted, the time and economic cost are almost unbearable.

The vigorous development of artificial intelligence provides new ideas and tools to solve the above bottlenecks. The AI platform, represented by the lineardesign algorithm, is dedicated to end-to-end modeling of the structure function relationship of complex CDS regions. Deep learning networks can instantly lock highly conservative and immunogenic epitopes in massive omics data to achieve precise enrichment. The generative model can evaluate the stability, translation efficiency, and immunogenicity of millions of CDS variants in parallel at one time. This completely eliminates the inefficient cycle of “trial and error verification”. The reinforcement learning framework can also iterate and optimize the composition ratio, surface modification, and targeted ligands of lipid nanoparticles (LNP) in a virtual environment, significantly reducing the experimental cycle and the cost of trial and error.

This study focuses on “the whole process design of AI-driven tumor gene vaccine”. Firstly, the three pain points of new antigen prediction, mRNA structure optimi-

zation, and LNP precise delivery were sorted out through literature research. Then, trosetarna was used as the core tool to analyze its optimization strategy in RNA secondary structure prediction, stability evaluation, and mutation tolerance analysis. Finally, the far-reaching significance of the AI paradigm in improving the efficiency of individualized treatment, reducing the cost of research and development, and rapidly responding to viral mutations was discussed, providing replicable paths and practical experience for the development of mRNA tumor vaccines. We hope to take this case to provide a feasible path and practical experience for the application of artificial intelligence in the research and development of tumor mRNA vaccines.

2. mRNA Vaccine

2.1 Mechanism of mRNA Vaccines

mRNA vaccines are administered via intramuscular injection, and their antigens are ultimately presented to immune cells.

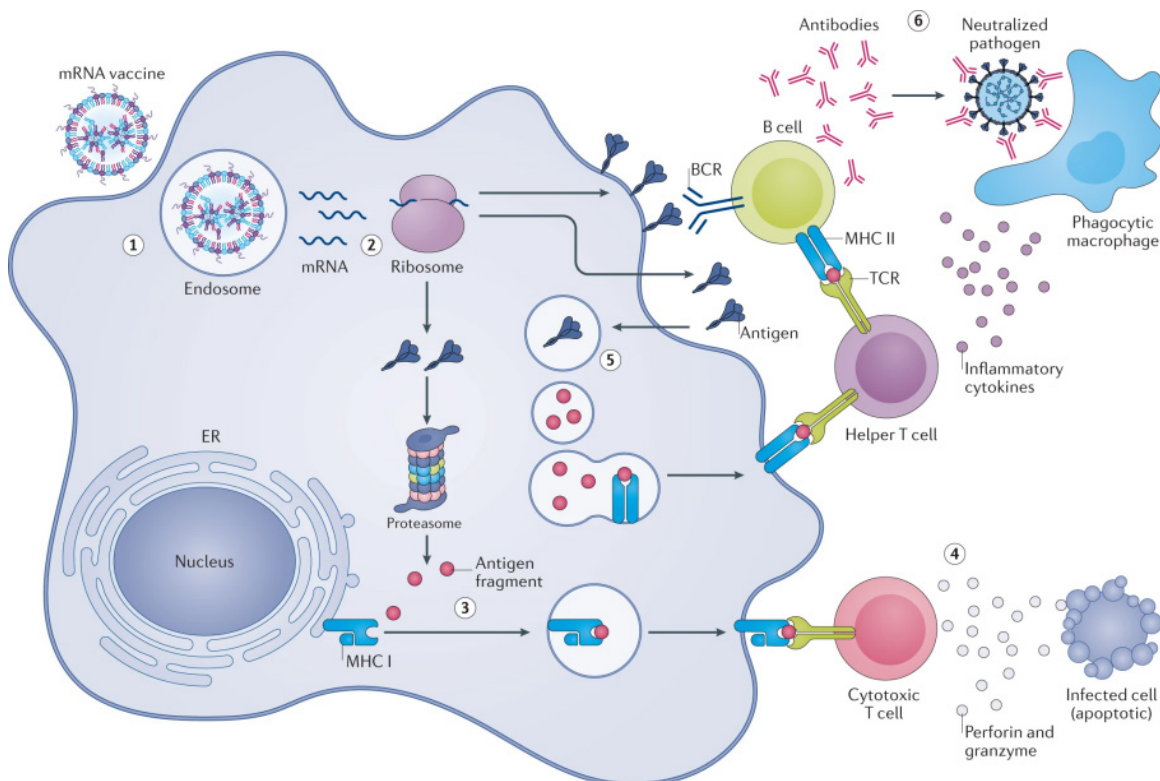


Fig.1 mRNA Vaccine of Immune Activation Pathway [1]

The specific mechanism can be seen in Fig.1. Firstly, the body's antigen-presenting cells absorb the mRNA through a process called endocytosis. Secondly, once the mRNA

escapes the endosome and reaches the cytoplasm, ribosomes synthesize it into the target protein. This newly produced antigenic protein can then trigger immune respons-

es through multiple pathways. Thirdly, the proteasome processes intracellular antigens into smaller peptides, which are then presented on the cell surface to cytotoxic T cells via major histocompatibility complex (MHC) class I molecules. Moreover, the release of toxic substances such as perforin and granzyme by activated cytotoxic T cells results in the elimination of infected cells. Additionally, secreted antigens may be internalized by other cells, degraded within endosomes, and displayed on the cell surface to helper T cells through MHC class II molecules. Finally, helper T cells aid in clearing pathogens by promoting antibody production in B cells and activating phagocytic cells, such as macrophages, via inflammatory cytokines.

2.2 Advantages of mRNA Vaccines

For one thing, rapid development and production, no need to culture pathogens or proteins- only the target antigen's mRNA sequence must be synthesized. Production is simple (taking weeks), making it ideal for emerging diseases or personalized cancer treatments. For another, high safety. mRNA does not integrate into the host genome, avoiding potential insertional mutagenesis risks associated with DNA vaccines. It degrades and clears naturally via metabolism, leaving no residue. Nucleotide modifications (e.g., pseudouridine) can reduce immunogenicity and minimize inflammatory responses. In addition, strong immune response. Simultaneously activates humoral immunity (antibody production) and cellular immunity (T-cell killing), making it more effective against cancer or viral infections. Immune effects can be enhanced by adjusting delivery systems or adjuvants. mRNA vaccines uniquely stimulate robust CD8⁺T cell responses alongside high antibody titers, overcoming a major limitation of subunit vaccines which predominantly elicit humoral immunity [2].

Besides, flexibility and Versatility, the same platform can

encode different antigens-simply changing the mRNA sequence allows for new vaccine development.

Supports multivalent vaccines (e.g., encoding multiple tumor antigens simultaneously). Furthermore, personalization potential, custom mRNA vaccines can be designed based on a patient's tumor mutations, addressing tumor heterogeneity. Personalized mRNA neoantigen vaccines can be designed within 6 weeks by sequencing patient-specific tumor mutations, enabling truly individualized cancer immunotherapy [3] Cost-Effectiveness as well, industrial-scale mRNA synthesis is relatively low-cost, and complex purification steps (e.g., recombinant protein vaccines) are unnecessary. Finally, comprehensive Advantages, mRNA vaccines combine the best attributes of live-attenuated (strong immunity) and subunit (safety) vaccines while avoiding their respective limitations -a 'game-changer' for vaccinology [4].

2.3 Design of mRNA Vaccines

The immunogenicity and expression efficiency of mRNA vaccines first depend on the modular design of their sequences. The mRNA vaccine consists of five domains. These domains are shown in Figure 2: the 5' cap structure helps mRNA stabilize, promotes translation initiation, and avoids degradation. The untranslated region at the 5' end does not encode protein, but determines the protein expression level by regulating translation efficiency and mRNA stability. Open Reading Frame (ORF), from the start codon to the stop codon, contains the region encoding the target protein (CDS) and can determine the protein function. Similar to 5' UTR, 3' UTR functions to regulate protein expression level and maintain mRNA stability. Some 3' UTR sequences can guide mRNA localization to specific cellular regions. Poly-A tail helps mRNA stability and translation efficiency, while the length of poly (a) tail indirectly regulates mRNA translation and half-life.

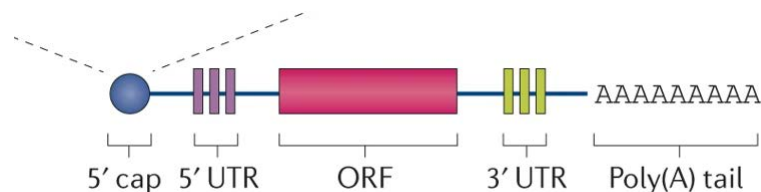


Fig.2 mRNA five domains [1]

The preparation of mRNA vaccine can be simply divided into six steps that showed in Fig.3. In the first step---sequence design. The sequence of the target antigen will be designed and inserted into the plasmid DNA structure after the pathogen genome is sequenced. Then in vitro, the plasmid DNA is transcribed into mRNA by phage polymerase, and the mRNA transcript is purified by high-per-

formance liquid chromatography in the third step of purification to remove pollutants and reactants. The fourth step is to mix the purified mRNA with lipids in a microfluidic mixer to form lipid nanoparticles. The rapid mixing makes the lipid wrap mRNA instantly and precipitate into self-assembled nanoparticles. The fifth step is to remove the non-aqueous solvent in the nanoparticle solution and

the mRNA that has not been encapsulated by dialysis or filtration. Finally, the filtered mRNA vaccine solution was

put into a sterile vial for storage. The above is the general process of mRNA vaccine preparation.

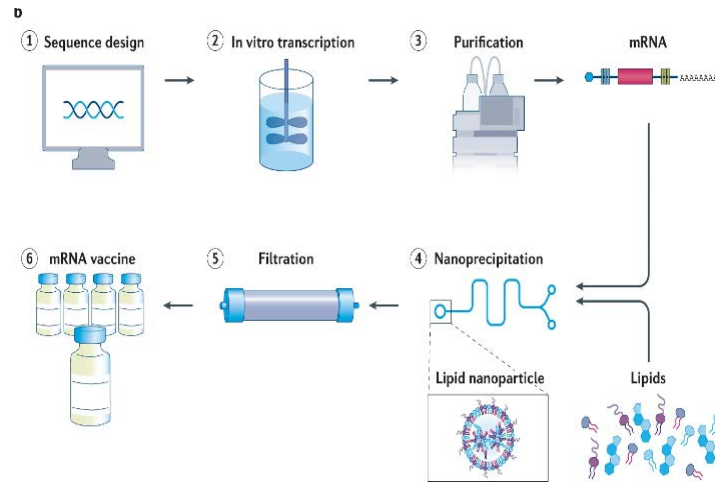


Fig.3 AI-Assisted mRNA Vaccine Design [1]

3. AI-Assisted mRNA Vaccine Design

3.1 Challenges in mRNA Vaccine Development and AI Solutions

3.1.1 challenges in CDS region design

Due to the presence of multiple possible sequences for the mRNA coding region targeting the same protein, the design of the CDS region becomes highly complex. The

Linear Design algorithm, developed by Zhang, H. et al., is an AI tool specifically designed for this purpose [6]. The research team drew inspiration from classical grid analysis in computational linguistics to identify the most stable mRNA in the DFA. LinearDesign substantially improves mRNA half-life. It also improves protein expression. Furthermore, it profoundly increases antibody titre. It can increase antibody titre by up to 128 times in mice [6].

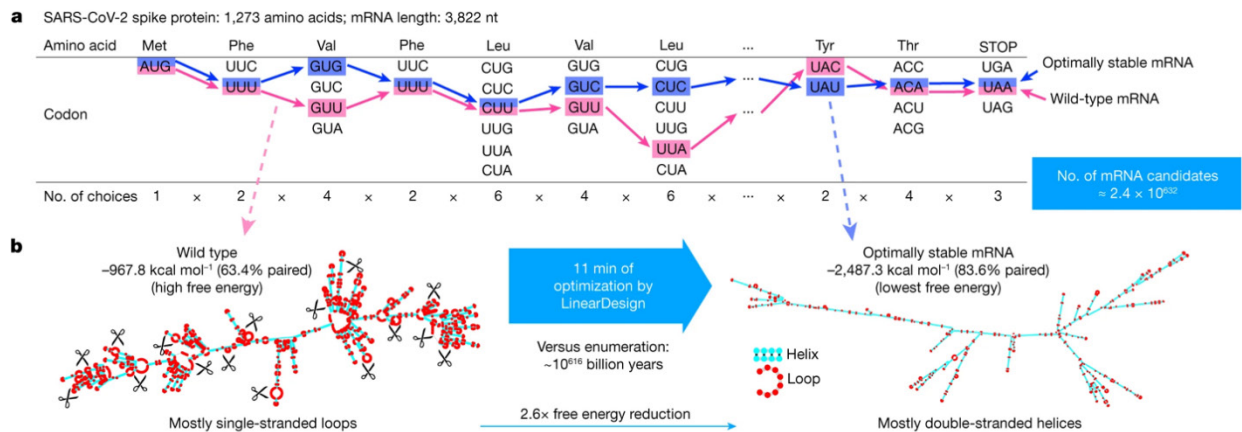


Fig. 4 Challenges in CDS Region Design (a) Wild-type (pink path) and optimally stable (blue path) sequences (b) The secondary structure of wild-type(left) and optimally stable(right) spike mRNAs [6]

As shown in Fig. 4a, due to codon redundancy and the exponential growth of combinations, approximately 2.4×10^{632} different mRNA sequences can encode the spike protein. Evaluating all possible sequences would require around 10^{616} billion years. The pink and blue pathways represent the wild-type sequence and the most stable

sequence (with the lowest free energy), respectively. nt refers to nucleotides. In Fig. 4b, the secondary structures of the wild-type (left) and optimally stable (right) spike mRNAs are depicted. LinearDesign's optimization process takes around 11 minutes [6].

3.1.2 Insufficient targeting of delivery systems

To address the issue of insufficient targeted delivery efficiency, lipid nanoparticles (LNPs) have been shown to effectively deliver mRNA. However, the impact of various LNP components on its functionality presents

challenges for precise design. Wang, W. et al. used AI and virtual screening techniques to predict two key properties of LNPs: apparent pKa values and mRNA delivery efficiency. This has contributed to the development of more effective and rational LNP design strategies [7].

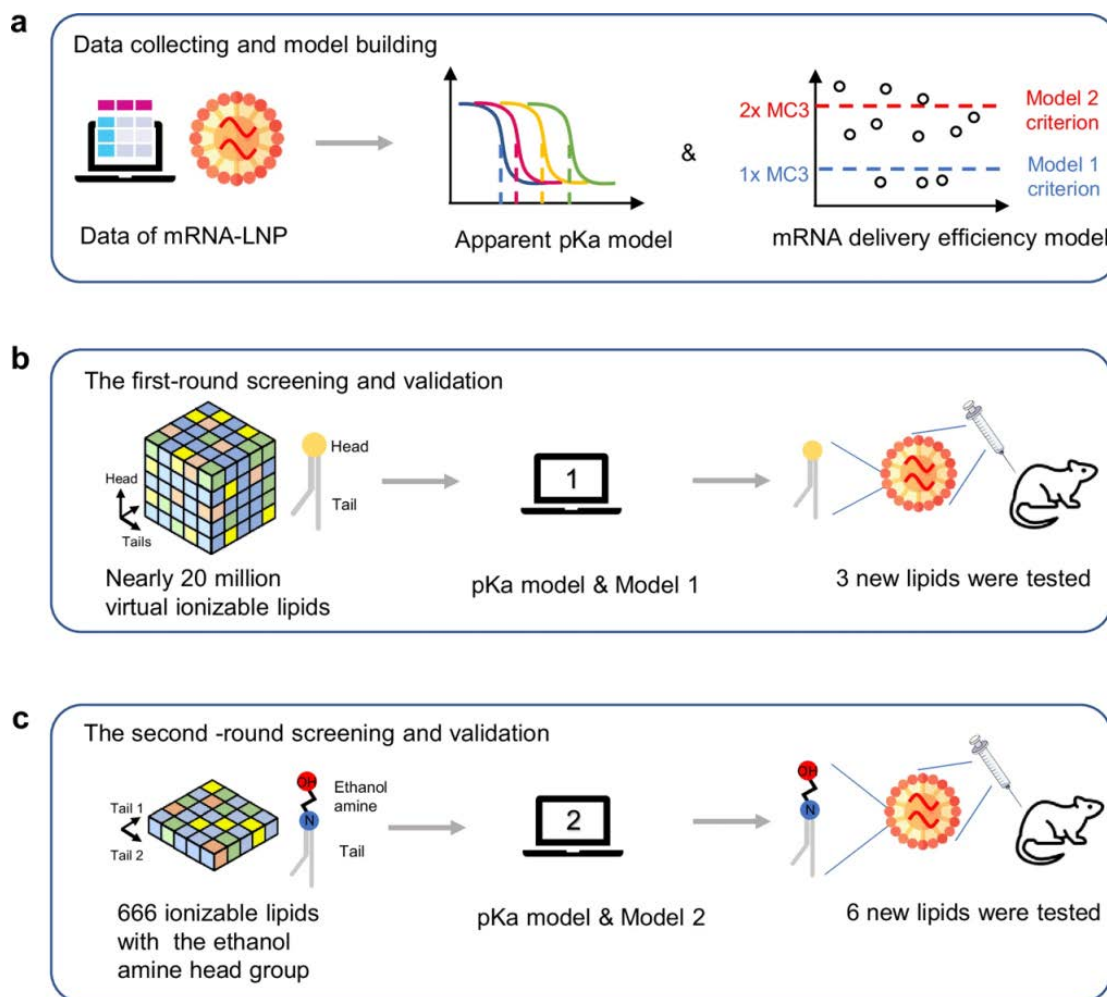


Fig.5 AI-guided Rational Design of Ionizable Lipids of mRNA LNPs (a)Data-driven models predicting pKa and mRNA delivery efficiency (b) Initial virtual screening and validation of ionizable lipids with pKa model and Model 1 (c) Second-round virtual screening and validation of ionizable lipids with pKa model and Model 2, MC3,DLin-MC3-DMA [7]

Figure 5 shows the AI-driven rational design process for ionisable lipids in mRNA LNP formulations. First, an AI model was used to predict the apparent pKa and mRNA delivery efficiency of the LNP. The delivery efficiency standards for the two models (Model 1 and Model 2) are set at one and two times that of the MC3, respectively (Figure 5a). A first round of screening was conducted, ultimately identifying three ionizable lipids. However, due to potential influences from the data source, the performance of the models was suboptimal (Figure 5b). A second round of lipid screening was therefore performed, focusing on lipids containing the ethanolamine head group. This re-

sulted in the selection of six lipids. In this screening, the performance of all six lipids met or exceeded that of MC3, with one comparable to that of SM-102 [7].

3.1.3 Viral mutations

Predicting viral variants in advance is crucial for developing optimal vaccines. To effectively predict viral escape, the research team developed a flexible AI tool, EVEscape. This algorithm can predict the early stages of a pandemic. By spotting potential mutations early on, it helps to develop effective vaccines and treatments [8].

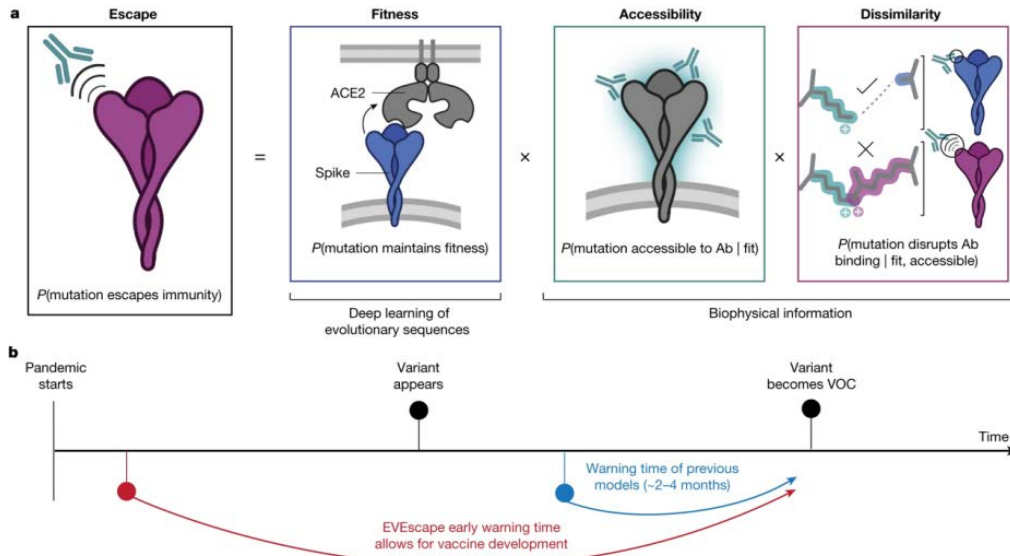


Fig. 6 Early antibody escape prediction using generative models: Structural and Biophysical Constraints (a) Estimating mutation escape likelihood via fitness, accessibility, and dissimilarity (b) Leveraging early-pandemic data for immune escape prediction [8]

As shown in Fig. 6a, The probability of a mutation inducing immune escape is expressed as the product of three probabilities. The first is the likelihood that the mutation will preserve viral fitness, this is known as the ‘fitness’ term. The second is the likelihood that the mutation will occur within an antibody-accessible region, this is known as the ‘accessibility’ term. The third is the likelihood that the mutation will disrupt antibody binding, this is known as the ‘dissimilarity’ term. As illustrated in Fig. 6b, these components can be informed by prepandemic data sources, enabling early warning [8].

3.2 Artificial Intelligence Technology: trRosettaRNA

For mRNA vaccine design, RNA three-dimensional structure prediction has been a long-standing challenge. The development of trRosettaRNA, an automated deep learning method for RNA 3D structure prediction, by the research team is a significant step towards resolving the complex issue of RNA structure prediction using deep learning.

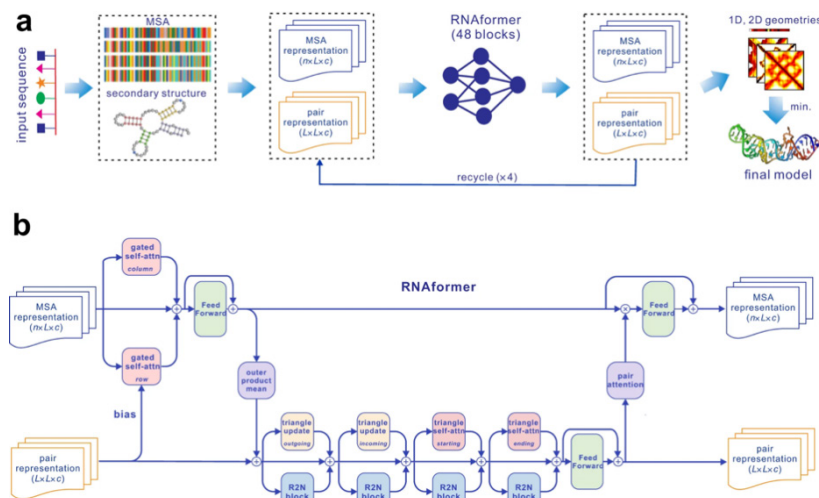


Fig. 7 Overall architecture of trRosettaRNA (a) flowchart of trRosettaRNA (b) structure of each RNA former block [9]

Fig. 7 shows the architecture of trRosettaRNA. Starting from the nucleotide sequence of the RNA of interest,

the programs rMSA31 and SPOT-RNA32 first generate a multiple sequence alignment (MSA) and a secondary

structure, respectively. The process involves converting them into representations in MSA and pair formats. These are then fed into a transformer network, the purpose of which is to predict 1D and 2D geometries.

In a manner similar to that of trRosetta, these geometries are converted into restraints. The purpose of this is to guide the final step of 3D structure folding, which is based on energy minimisation [9].

3.3 Usage of trRosettaRNA

Though importing the current RNA sequence into the website, the predicted RNA structure and LDDT format are obtained [9].

3.3.1 Advantages of the Model for mRNA Vaccine Design

Firstly, though the deep study and de novo folding, this model has a higher confidence and it can offer a simulated RNA model that is more similar to the natural structure on the mRNA vaccine designing part, and make it easier to be recognized by the immune system.

Moreover, this website provides a download tunnel and we can get the structure directly predicted by the end-to-end neural network. End-to-end neural networks can be

used to predict the distribution of RNA secondary structures, and then it provides advantages in the design of mRNA vaccines, such as optimizing stability and half-life, improving translation efficiency, reducing immunogenicity, and facilitating antigen design [6].

Third, it provides the multiple sequence alignment used. Multiple sequence alignment analyzes the conserved regions of RNA sequences related to different pathogens to identify those sequence fragments that have remained relatively stable during evolution. These regions have the potential to become vaccine targets. Designing vaccines against them can make the vaccines more broad-spectrum and effective^[10].

3.3.2 Core Ideas for Tumor mRNA Vaccine Design

The core ideas for designing tumor mRNA vaccines based on the structural characteristics of met-p10-c10 are as follows. First, giving priority to using the high-LDDT region to lock the key functional areas and ensure the core function of the vaccine, and second, sequencing optimization in the low-LDDT region to reduce the “structural risk” of the vaccine. Last, combining the characteristics of the tumor microenvironment, structural features are utilized to enhance the targeting of vaccines [11].

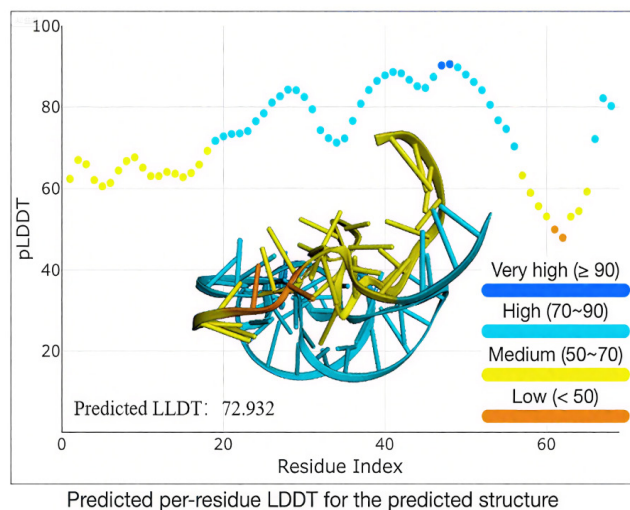


Fig. 8 Predict 3D model and LDDT format of “Met-p10-c10” from trRosettaRNA

To gain a deeper understanding of the practicality of trRosettaRNA, a practical test was ran. As shown in Figure 8, the ‘Met-p10-c10’ RNA sequence (data from NCBI) was entered into the input box. After a brief wait, the three-dimensional structure diagram and LDDT value were obtained. The predicted LDDT is 72.932, this means that it has a moderately high structural reliability, but regional differences need to be taken into consideration.

4. Conclusion

Thanks to their fast development, high safety and strong ability to trigger immune responses, tumour mRNA vaccines have become a promising option for personalised cancer treatment. Their effectiveness depends on smart sequence design—including parts like the 5’ cap, UTRs, ORF, and poly-A tail—and getting them delivered properly, but there are still problems, that is, designing the CDS region is complicated, delivery systems don’t target well

enough, and predicting tumor-related mutations is hard and scientists must spend so much time on finding their sequences.

Artificial intelligence is solving these issues well. Tools like LinearDesign use computer algorithms to make mRNA more stable and better at making proteins, which has led to much higher antibody levels in early tests. AI also helps improve lipid nanoparticles to target delivery more accurately, while EVEscape predicts how tumors might change, helping vaccines stay effective. Importantly, trRosettaRNA, a deep learning tool, handles the long-standing challenge of predicting RNA 3D structures. It uses genetic comparisons and structure data to create reliable models—like one with an LDDT score of 72.932 for related sequences.

These AI tools simplify the design of tumor mRNA vaccines, from optimizing sequences to checking structures, speeding up the creation of personalized and effective cancer treatments.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Chaudhary, N., Weissman, D., Whitehead, K.A. mRNA vaccines for infectious diseases: principles, delivery and clinical translation. *Nat Rev Drug Discov*, 2021, 20: 817-838.
- [2] Pardi N, Hogan M J, Porter F W, et al. mRNA vaccines — a new era in vaccinology. *Nature Reviews Drug Discovery*, 2022, 21(2): 161-183.
- [3] Verbeke R, Lentacker I, De Smedt S C, et al. Three decades of messenger RNA vaccine development. *Nano Today*, 2021, 38: 101175.
- [4] Ott P A, Hu Z, Keskin D B, et al. An immunogenic personal neoantigen vaccine for melanoma. *Nature*, 2017, 547(7662): 217-221.
- [5] Dolgin E. The tangled history of mRNA vaccines. *Nature*, 2021, 597(7876): 318-324.
- [6] Zhang, H., Zhang, L., Lin, A. et al. Algorithm for optimized mRNA design improves stability and immunogenicity. *Nature*, 2023, 621: 396-403.
- [7] Wang, W., Chen, K., Jiang, T. et al. Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery. *Nat Commun*, 2024, 15: 10804.
- [8] Thadani, N.N., Gurev, S., Notin, P. et al. Learning from prepandemic data to forecast viral escape. *Nature*, 2023, 622, 818-825.
- [9] Wang, W., Feng, C., Han, R. et al. trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nat Commun*, 2023, 14: 7266.
- [10] Zhao, Y., Ni, W., Liang, S. et al. Vaccination with Span, an antigen guided by SARS-CoV-2 S protein evolution, protects against challenge with viral variants in mice. *Sci Transl Med*, 2023, 15(677): eabo3332.
- [11] Shrestha, B., Adhikari, B. et al. Scoring protein sequence alignments using deep learning. *Bioinformatics*, 2022, 38(11), 2988–2995.