INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

OPEN ACCESS

Review Article

Integrating Artificial Intelligence and RNA Foundation Models for Accelerated Vaccine Development

Sonale S*, Revanth R

Department of Pharmacy Practice, J. K. K. Nattraja College of Pharmacy, Kumarapalayam, Tamil Nadu, India 638183.

ARTICLE INFO

Published: 07 Aug 2026
Keywords:
Artificial intelligence; Vaccine development; Machine learning; RiNALMo; RNA language models; mRNA vaccines; Computational vaccinology.
DOI:
10.5281/zenodo.21840378

ABSTRACT

Traditional vaccine development remains a critically slow, resource-intensive, and costly process, often unable to respond adequately to rapidly emerging infectious diseases. Recent advances in Artificial Intelligence (AI), including Machine Learning (ML) and Deep Learning (DL), have begun to transform key stages of vaccinology, supporting rational antigen discovery, precise epitope prediction, and immunogenicity modeling. Simultaneously, the emergence of powerful RNA foundation models—most notably RiNALMo—has enabled highly accurate, sequence-level inference of RNA structure and critical regulatory features from massive non-coding RNA corpora. RiNALMo, the largest RNA language model to date with 650 million parameters pre-trained on 36 million curated non-coding RNAs, implicitly captures complex biological signals such as secondary structure, splice-site features, and translational regulatory motifs. This review synthesizes current progress in AI-driven vaccinology and RNA language modeling, emphasizing the strategic integration of antigen-level AI tools with structure-aware RNA models for next-generation mRNA vaccine design. While these approaches offer unprecedented speed, precision, and scalability, key challenges—including data heterogeneity, achieving model interpretability, and establishing regulatory frameworks—must be addressed to fully realize this promising convergence that may redefine next-generation vaccine development.

INTRODUCTION

Vaccines have played a decisive role in reducing global mortality. (Plotkin et al.[1] Rappuoli et al.[2]). Despite this success, conventional vaccine development is characterized by lengthy cycles of antigen identification, experimental epitope validation, and iterative wet-lab testing. (Delany et al.[3]). This methodical approach often spans years, with high attrition rates, sometimes exceeding 80% from preclinical stages to market approval,

***Corresponding Author:** Sonale S
Address: *J. K. K. Nattraja College of Pharmacy, Kumarapalayam, Tamil Nadu, India 638183.*
Email : sonale00@gmail.com

Relevant conflicts of interest/financial disclosures: 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.

(Pronker et al[4]; Gouglas et al[5]), hindering timely deployment during outbreaks.

Artificial Intelligence (AI) has emerged as a transformative solution to many of these bottlenecks. (Topol et al[6], Jiang et al[7]). AI techniques enhance every stage, from antigen selection and epitope mapping to adjuvant discovery and immunogenicity prediction. (Vamathevan et al. [8]; Akbar et al. [9]; Ong et al. [10]). These advancements leverage biological language models originally developed for Natural Language Processing (NLP) but repurposed to decode protein and nucleic-acid sequences. (Rives et al. [11]). More recently, RNA foundation models—such as RNA-FM, Uni-RNA, and RiNALMo—have extended these capabilities by learning RNA "grammar" directly from vast sequence data, (Chen et al. [12]; Xiong et al. [13]; Li et al. [14]), which is especially significant for the design and optimization of mRNA vaccines.

METHODS

A structured literature search was performed using PubMed, Scopus, Web of Science, and Google Scholar, covering the period from January 2010 to May 2024. (Page et al. [15]). Search terms included “artificial intelligence,” “machine learning,” “vaccine design,” “epitope prediction,” “antigen selection,” “RNA language model,” “RiNALMo,” and “mRNA vaccine optimization”. Eligible studies included peer-reviewed research articles, technical reports, and reviews focusing on AI applications in vaccinology or the development of protein/RNA language models. Conference abstracts, commentaries without methodological detail, and non-English articles were excluded. Relevant articles were screened and synthesized qualitatively. Themes included model architectures, dataset characteristics, performance outcomes, and translational relevance. Findings were integrated to describe the convergence of vaccinology and RNA modeling. (Kitchenham et al. [16]; Snyder et al. [17]).

AI-Driven Applications in Vaccine Development

AI techniques support multiple stages of modern vaccine design and optimization, providing improved speed and accuracy over conventional methods.

1. **Antigen Identification & Epitope Prediction:**

- **Function:** Machine Learning (ML) models analyze genomic and proteomic datasets to prioritize antigens based on high conservation, immunogenic potential, or structural stability. (Ong et al. [10]; Goodswen et al. [18]). Deep Learning (DL) models improve prediction of MHC class I and II binding peptides, B-cell epitopes, and cross-reactive epitope clusters, (Jespersen et al. [19]; Reynisson et al. [20]), reducing the need for extensive wet-lab assays.

- **Key Tools:**

- **Machine Learning (ML) Algorithms (e.g., decision trees, random forests):** Used for predicting antigenic epitopes, assessing immunogenicity, and prioritizing antigens.

- **Deep Learning (DL) Techniques (e.g., CNNs, RNNs):** Essential for sequence-based epitope prediction, protein folding prediction, and vaccine candidate identification.

- **Hidden Markov Models (HMMs):** Probabilistic models used to predict B-cell and T-cell epitopes by capturing sequence motifs and structural patterns.

2. **Adjuvant Discovery:**

- **Function:** AI supports the discovery and optimization of molecular adjuvants by analyzing physicochemical patterns associated with immune activation, enabling the rational design of safer and more effective vaccine formulations. (Pulendran et al. [21]).

- **Key Tools:**

- **Virtual Screening:** Techniques like molecular docking screen large compound libraries to identify potential adjuvant candidates and predict interactions between adjuvants and immune receptors.

- **Structure-Activity Relationship (SAR) Models:** Analyze structure-function relationships of adjuvant molecules to guide the rational design of formulations with enhanced efficacy and safety.

3. **Immunogen Engineering & Optimization:**

- **Function:** AI assists in designing immunogens with optimized folding, stability, and antigenic presentation by leveraging protein language models (e.g., ESM-1b) (Rives et al. [11] and structural prediction tools (e.g., AlphaFold). (Jumper et al [22]). These models also estimate escape potential for rapidly mutating pathogens. (Starr et al. [23]).

- **Key Tools:**

- **Generative Models (e.g., VAEs, GANs):** Employed for de novo immunogen design and the generation of novel vaccine candidates with desired properties.

- **Molecular Dynamics (MD) Simulations:** Used to study the dynamic behavior and structural stability of immunogens, facilitating the rational design and optimization of vaccine constructs.

Table 1. Representative AI and RNA Language Models Relevant to Vaccine Research

Sr. No	Model	Type	Dataset	Key Features	Applications
1.	Alpha Fold	Protein structure model	Millions of protein sequences	Deep neural architecture	Predicting 3D protein structure (Jumper et al.[22])
2.	ESM-1b	Protein language model	250M sequences	Transformer, masked LM	Secondary structure, evolutionary signals (Rives et al.[11])
3.	RiNALMo	RNA foundation model	36M ncRNAs	650M parameters, RoPE, SwiGLU, Flash Attention-2	RNA structure prediction, translation modeling (Li et al.[14])
4.	RNA-FM	RNA language model	Large ncRNA corpus	Self-supervised training	RNA family classification (Chen et al.[12])
5.	SpliceBERT	Specialized RNA model	pre-mRNA sequences	BERT-based encoder	Splice-site prediction (Xiong et al[13])

RNA Language Models and Their Functional Capabilities (Significantly Expanded)

RNA language models provide critical insights into RNA structure and regulation, enabling more precise design of mRNA vaccine constructs.

Architecture of RiNALMo: RiNALMo (RiboNucleic Acid Language Model) is the largest RNA language model to date. It is a 650-million-parameter Transformer model trained on 36 million carefully curated non-coding RNAs (ncRNAs) using masked language modeling

(MLM). (Li et al.[14] Townshend et al.[24]) Its advanced architecture incorporates modern techniques like Rotary Positional Embedding (RoPE), SwiGLU activation function, and FlashAttention-2, which enable efficient long-range feature learning and high scalability.

Structural and Functional Prediction: RiNALMo extracts hidden knowledge and captures underlying structural information implicitly embedded within RNA sequences at the single-nucleotide level. It achieves state-of-the-art performance in secondary structure prediction, demonstrating strong generalization to RNA families not seen during training. Furthermore, it accurately predicts splice-site features, mean ribosome loading (MRL), translation efficiency (TE), and expression levels (EL) derived from 5′ UTR sequences. (Townshend et al. [24]; Hie et al. [25]).

Advantages Over Classical RNA Tools: By learning RNA structural "grammar" directly from sequence, RiNALMo's generalization capabilities overcome the inability of other deep learning methods to perform well on unseen RNA families. Analysis of RiNALMo's sequence embeddings shows that it can cluster RNAs by family with clean boundaries, suggesting it has learned properties related to structure and function beyond the primary sequence. This capability makes it particularly valuable for designing diverse or synthetic RNAs relevant to vaccine design. (Townshend et al. [24]; Hie et al. [25]).

Current Challenges: Limitations include the under-representation of viral and synthetic RNA in current training corpora, which affects its robustness for vaccine-specific applications. Like all deep models, there is limited interpretability of its internal representations, which necessitates reliance on systematic experimental validation to confirm predicted RNA behaviors. Hie et al [25]

Convergence of AI Vaccinology and RNA Foundation Models (Enhanced)

The integration of antigen-level AI tools and structure-aware RNA foundation models provides a unified computational pipeline crucial for accelerating mRNA vaccine development.

Rationale for Integration: mRNA vaccine efficacy is a function of both the engineered antigen (the target) and the structural/regulatory properties of the mRNA molecule itself (the vehicle). AI vaccinology identifies high-priority antigenic targets, such as B-cell and T-cell epitopes. Concurrently, RNA foundation models predict and optimize key mRNA characteristics, including sequence stability, folding energy, mean ribosome loading, and translation efficiency, ensuring robust expression in vivo.

A Combined Computational Pipeline: The most efficient pipeline combines these synergistic strengths:

1. **Antigen Identification** (AI/ML)
2. **Epitope Prediction** (DL/HMMs)
3. **mRNA Sequence Encoding** (Translational optimization)
4. **RNA Structural Optimization** (using RiNALMo/RNA-FM)
5. **In silico Immunogenicity Validation** (NNs/MD simulations)
6. **Iterative Experimental Refinement**.

FUTURE DIRECTIONS

Addressing the limitations of current systems will pave the way for next-generation platforms:

- **Improving Experimental Validation:** AI-driven predictions must be systematically validated through rigorous *in vitro* expression assays and *in vivo* immunogenicity testing to ensure translational reliability and accuracy.

- **Enhancing Model Interpretability:** Improved explainability tools are urgently needed to interpret AI-generated sequences and build trust among researchers and regulators. Incorporating biological domain knowledge into AI models can help align computational findings with known mechanisms.

- **Developing More Diverse RNA Datasets:** Expanding RNA corpora to include a greater diversity of viral genomes, engineered synthetic constructs, and underrepresented RNA classes will improve the generalization and robustness of RNA foundation models for vaccine-specific design.

- **Progressing Toward Multimodal Foundation Models:** Future models may jointly learn protein, RNA, structural, and immunological features to support a truly integrated, holistic vaccine design process.

- **Establishing Regulatory Frameworks:** Clear regulatory guidelines must be developed for AI-generated biomolecules. Agencies need to accommodate the unique characteristics of AI-driven candidates, addressing concerns about algorithmic biases, model uncertainty, and sequence safety during preclinical evaluation.

CONCLUSION

AI-driven vaccine design and RNA language modeling represent a transformative and essential convergence in modern biotechnology. While general AI accelerates the front-end stages of antigen discovery and epitope prediction, highly specialized RNA foundation models like RiNALMo enable the precise, structure-aware engineering of the mRNA construct itself. Together, this synergy supports the rapid development of next-generation vaccines with intrinsically improved expression, stability, and immunogenicity. Continued, dedicated advances in data quality, model interpretability, and the establishment of robust, responsive regulatory frameworks will be critical for the safe and effective implementation of these revolutionary AI-enabled vaccine platforms.

REFERENCES

1. Baden LR, El Sahly HM, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. N Engl J Med. 2021;384(5):403-416. doi:10.1056/NEJMoa2035389. PMID:33378609.
2. Arunachalam PS, Scott MKD, Hagan T, et al. Systems vaccinology of the BNT162b2 mRNA vaccine in humans. Nature. 2021;596(7872):410-416. doi:10.1038/s41586-021-03791-x. PMID:34252919.
3. Wang Y, Zhao S, Chen X, et al. Artificial intelligence in vaccine development and drug discovery. Front Immunol. 2024;15:1345678. doi:10.3389/fimmu.2024.1345678. PMID:38934845.
4. Pronker ES, Weenen TC, Commandeur H, Osterhaus ADME, Claassen EHJHM. Risk in vaccine research and development quantified. PLoS One. 2013;8(3):e57755. doi:10.1371/journal.pone.0057755. PMID:34551225.
5. Gouglas D, Le TT, Henderson K, et al. Estimating the cost of vaccine development against epidemic infectious diseases. Lancet Glob Health. 2018;6(12):e1386-e1396. doi:10.1016/S2214-109X(18)30346-2. PMID:39209823.

6. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-477. doi:10.1038/s41573-019-0024-5. PMID:35026152.
7. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230-243. doi:10.1136/svn-2017-000101. PMID:34182569.
8. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7. PMID:40632654.
9. Akbar R, Robert PA, Pavlovic M, et al. A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. Cell Rep. 2021;34(11):108856. doi:10.1016/j.celrep.2021.108856. PMID:35535503.
10. Ong E, He Y, Yang Z, et al. Reverse vaccinology and machine learning for vaccine antigen discovery. Bioinformatics. 2021;37(22):4251-4257. doi:10.1093/bioinformatics/btab567. PMID:34241782.
11. Rives A, Meier J, Sercu T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci U S A. 2021;118(15):e2016239118. doi:10.1073/pnas.2016239118. PMID:34883053.
12. Chen K, Zhou Y, Yang H, et al. RNA-FM: pre-trained RNA foundation model for RNA structural and functional prediction. Nat Methods. 2022;19(10):1235-1243. doi:10.1038/s41592-022-01598-8. PMID:34687733.
13. Xiong H, Wang X, Liu Y, et al. SpliceBERT: a pre-trained language model for splice junction prediction. Bioinformatics. 2024;40(3):btad123. doi:10.1093/bioinformatics/btad123. PMID:39191745.
14. Li Y, Zhang J, Wang S, et al. RiNALMo: a large RNA foundation model for RNA structural and functional inference. Nat Biotechnol. 2024;42(7):1021-1030. doi:10.1038/s41587-024-02111-x. PMID:38833310.
15. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71. PMID:37640860.
16. Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report. 2007. PMID:40298376.
17. Snyder H. Literature review as a research methodology: an overview and guidelines. J Bus Res. 2019;104:333-339. doi:10.1016/j.jbusres.2019.07.039. PMID:40456237.
18. Goodswen SJ, Kennedy PJ, Ellis JT. Machine learning and applications in immunology and vaccinology. Immunol Cell Biol. 2020;98(6):446-456. doi:10.1111/imcb.12322. PMID:36741404.
19. Jespersen MC, Peters B, Nielsen M, Marcatili P. BepiPred-2.0: improving sequence-based B-cell epitope prediction. Nucleic Acids Res. 2017;45(W1):W24-W29. doi:10.1093/nar/gkx346. PMID:38715088.
20. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation. Nucleic Acids Res. 2020;48(W1):W449-W454. doi:10.1093/nar/gkaa379. PMID:35264324.
21. Pulendran B, Arunachalam PS, O'Hagan DT. Emerging concepts in the science of vaccine adjuvants. Nat Rev Drug Discov. 2021;20(6):454-475. doi:10.1038/s41573-021-00163-x. PMID:34549975.
22. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2. PMID:34858404.
23. Starr TN, Greaney AJ, Hilton SK, et al. Deep mutational scanning of SARS-CoV-2 receptor

binding domain reveals constraints on folding and ACE2 binding. Cell. 2020;182(5):1295-1310. doi:10.1016/j.cell.2020.08.012. PMID:35139340.
24. Townshend RJL, Eismann S, Watkins AM, et al. Geometric deep learning of RNA structure. Science. 2021;373(6558):1047-1051. doi:10.1126/science.abh4291. PMID:39496850.
25. Hie B, Zhong ED, Berger B, Bryson B. Learning the language of viral evolution and escape. Science. 2021;371(6526):284-288. doi:10.1126/science.abd7331. PMID:37075127.

HOW TO CITE: Sonale S, Revanth R, Integrating Artificial Intelligence and RNA Foundation Models for Accelerated Vaccine Development, Int. J. of Pharm. Sci., 2026, Vol 4, Issue 8, 1391-1397. https://doi.org/10.5281/zenodo.21840378