Published by SET Publisher

Journal of Pharmacy and Nutrition Sciences

ISSN (online): 1927-5951

From Sequence to Decision: Large Language Models as End-to-End Engines for Therapeutic RNA Design

Alfi Sophian*

The Indonesian Food and Drug Authority, Jl. Percetakan Negara, No. 23, Jakarta Pusat, 10560, Indonesia

Article Info:

Keywords:
RNA therapeutics,
large language models,
RNA foundation models,
mRNA design,
RNA aptamers.

Timeline:
Received: June 11, 2026
Accepted: July 15, 2026
Published: August 03, 2026

Citation: Sophian A. From sequence to decision: Large language models as end-to-end engines for therapeutic RNA design. J Pharm Nutr Sci 2026; 16: 18-28.

DOI: https://doi.org/10.29169/1927-5951.2026.16.03

*Corresponding Author

Tel: +62 857-8130-3080
E-mail: alfi.sophian@pom.go.id

Abstract:

RNA-based therapeutics have achieved remarkable clinical validation, yet rational sequence design remains a fundamental bottleneck. Large language models (LLMs) trained on tens of millions of RNA sequences are beginning to reframe this challenge, offering a unified computational framework spanning structure prediction, immunogenicity optimization, and de novo sequence generation. Here, we survey the expanding landscape of RNA-focused foundation models — from early BERT-based encoders such as RNA-FM and ERNIE-RNA to large generative systems including GRAPE-LM — and map their end-to-end utility across six therapeutic modalities: messenger RNA, siRNA, antisense oligonucleotides, aptamers, circular RNA, and CRISPR guide RNAs. We identify three critical translation barriers: the RNA structural data famine, the absence of chemical-modification-aware architectures, and the lack of *in vivo* benchmarking datasets. We propose a roadmap — anchored by community data-generation initiatives and pre-competitive regulatory standards — for translating LLM-driven RNA design into clinical practice.

INTRODUCTION

The emergence of RNA as a clinically validated drug class represents one of the most consequential developments in modern medicine. From the landmark approval of patisiran in 2018 — the first RNA interference (RNAi) therapeutic for hereditary transthyretin amyloidosis — to the deployment of mRNA COVID-19 vaccines at billion-dose scale, RNA has demonstrated a capacity for programmable, sequence-directed engagement with virtually any biological target [1,2].

More than 20 RNA-based drugs have received regulatory approval globally as of early 2025, encompassing siRNAs for hypercholesterolaemia and rare liver diseases, antisense oligonucleotides (ASOs) for neurological disorders, and emerging modalities including circular RNA and self-amplifying RNA [3,4].

The programmability of RNA is its defining asset: the information content of an RNA drug resides in a four-letter nucleotide sequence that can, in principle, be specified computationally. Yet the sequence-to-function relationship of RNA is fiendishly complex. A linear sequence folds into an intricate three-dimensional architecture governed by thermodynamic and kinetic forces; that architecture determines target engagement, translational efficiency, innate immune recognition, nuclease stability, and off-target activity. Predicting these properties has historically required bespoke biophysical models for each RNA class, with limited generalizability across modalities and a persistent gap between in silico prediction and *in vivo* pharmacological activity [5-7].

Large language models (LLMs), pre-trained on the statistical regularities of biological sequences at massive scale, offer a fundamentally different paradigm. Rather than encoding explicit biophysical rules, these models learn latent representations of sequence–structure–function relationships directly from data. The transformer architecture [8], with its self-attention mechanism, captures long-range nucleotide dependencies inaccessible to recurrent networks — dependencies that determine RNA tertiary folding and function. The success of protein language models, most notably ESM-2, which achieved structural prediction accuracy rivalling physics-based approaches without evolutionary information [9], has catalyzed an analogous movement in RNA biology, yielding a rapidly expanding family of RNA-focused foundation models [10-14]. In this review, we synthesize this landscape, critically evaluate demonstrated utility across six therapeutic modalities, and chart the barriers and priorities for clinical translation.

MATERIALS AND METHODS

This review employed a structured narrative synthesis approach appropriate to a rapidly evolving interdisciplinary field. We searched PubMed/MEDLINE, Web of Science, Scopus, and the preprint servers bioRxiv and arXiv for records published between January 2018 and June 2026, using combinations of the terms ‘RNA language model’, ‘RNA foundation model’, ‘transformer’ AND ‘RNA’, ‘mRNA design’ AND ‘deep learning’, ‘RNA aptamer’ AND ‘generative model’, ‘siRNA’ AND ‘machine learning’, ‘CRISPR guide RNA’ AND ‘language model’, and ‘RNA structure prediction’ AND ‘self-supervised learning’. Reference lists of retrieved articles and forward-citation alerts were screened for additional relevant records (snowballing).

Records were included if they (i) described a transformer-based or other self-supervised language model trained on RNA sequence data, (ii) reported at least one quantitative benchmark or wet-laboratory validation, and (iii) addressed a therapeutic RNA modality (mRNA, siRNA, ASO, aptamer, circRNA, saRNA, or CRISPR guide RNA) or a barrier directly relevant to clinical translation. Records were excluded if they addressed purely diagnostic or agricultural RNA applications unrelated to human therapeutics, were conference abstracts without an accompanying full text, or duplicated results already reported in a peer-reviewed version of a preprint. Model selection for in-depth discussion (Tables **1** and **2**) prioritized models with publicly available benchmarks, wet-laboratory validation, or clear architectural novelty rather than exhaustive field coverage; this prioritization is itself a source of selection bias, which we discuss explicitly alongside the field's other limitations.

RESULTS AND DISCUSSION

Therapeutic RNA Modalities and the Design Problem

A Landscape of Approved and Emerging Modalities

Therapeutic RNAs are mechanistically diverse. ASOs — the most clinically mature class — modulate pre-mRNA splicing or recruit RNase H for mRNA degradation; eleven agents have received FDA approval, spanning Duchenne muscular dystrophy,

spinal muscular atrophy, and hypercholesterolaemia. Small interfering RNAs trigger RNA-induced silencing complex (RISC)-mediated cleavage of target mRNAs; six siRNA drugs are approved, with inclisiran and patisiran as the commercial vanguard [3]. Messenger RNA therapeutics — catalyzed by the COVID-19 pandemic — now encompass vaccines, protein-replacement therapies for rare diseases, and personalized neoantigen cancer vaccines [2]. RNA aptamers, single-stranded nucleic acids that fold into target-binding conformations, have achieved clinical application with pegaptanib (anti-VEGF) and avacincaptad pegol (anti-C5, geographic atrophy).

Emerging modalities extend the therapeutic space. Circular RNAs (circRNAs), covalently closed and exonuclease-resistant, enable prolonged protein expression from a single dose [4]. Self-amplifying RNAs (saRNAs) encode viral replicase machinery to amplify the therapeutic transcript intracellularly, offering effective immunization from nanogram-scale doses. RNA components of CRISPR systems — guide RNAs (gRNAs), base-editor scaffolds, and prime-editing pegRNAs — represent a distinct class in which sequence optimization directly governs editing specificity and efficiency. The aggregate clinical opportunity is substantial: the combined market for approved RNA therapeutics is projected to exceed USD 20 billion annually by 2030, with a pipeline spanning rare disease, oncology, cardiology, and infectious disease.

Shared Design Challenges

Despite mechanistic diversity, all therapeutic RNAs share a common set of design challenges. First, structure governs function: RNA secondary and tertiary architecture profoundly influences biological activity, but accurate three-dimensional structure prediction from sequence alone remains unsolved for long transcripts [5]. AlphaFold 3 [15] has advanced RNA structural prediction, yet its accuracy for RNA remains substantially lower than for proteins, and the RNA structure database in the Protein Data Bank contains fewer than 5,000 entries as of 2025 — more than 40-fold smaller than the protein structural database — a data asymmetry with direct consequences for learning-based approaches.

Second, immunogenicity constrains the design space. Innate immune pattern recognition receptors — including TLR3, TLR7, TLR8, and RIG-I — detect exogenous RNA and trigger inflammatory responses that compromise both safety and efficacy. The incorporation of modified nucleosides — most famously pseudouridine (Ψ) and N1-methylpseudouridine (m1Ψ), the innovation recognized by the 2023 Nobel Prize in Physiology or Medicine [20] — substantially attenuates immune activation but introduces a chemical modification landscape that current RNA LLMs, trained exclusively on canonical A, U, G, C sequences, cannot navigate. Third, off-target effects — siRNA seed-region silencing, ASO hybridization to non-target mRNAs, aptamer cross-reactivity — impose multi-objective optimization constraints that bespoke models address inefficiently. The aggregate challenge is a high-dimensional, multi-objective design problem uniquely suited to generative AI.

RNA Language Models: Architecture and Performance

From Masked Encoders to Large Generative Models

The analogy between natural language and biological sequences is mechanistically grounded. In both domains, meaning emerges from context: the functional consequence of a nucleotide is determined by its broader sequence neighborhood, just as the semantic content of a word is determined by sentence context. Transformer self-attention [8] captures these long-range dependencies — including distal nucleotide contacts that determine RNA tertiary folding — in a way that recurrent architectures cannot.

RNA-FM [10], trained using BERT-style masked language modelling on 23.7 million ncRNA sequences from RNAcentral, established the paradigm: its representations encode secondary structure, three-dimensional proximity, and evolutionary signals without any structural supervision, enabling improved performance on structure prediction, protein–RNA interaction, and gene expression regulation tasks.

ERNIE-RNA [11] introduced a key architectural innovation: structure-aware attention masking, incorporating base-pairing biases directly into the pre-training objective. Even without fine-tuning, ERNIE-RNA's attention maps capture RNA secondary structure with an F1 score of 0.873 on bpRNA-1m, outperforming dedicated structure-prediction tools and demonstrating that structural priors can be internalized during language model training. RiNALMo [12], with 650 million parameters pre-trained on 36 million ncRNA sequences from RNAcentral and Rfam, demonstrated a critical property: inter-family generalization. Applied to

mRNA translation efficiency prediction — a task outside its training distribution — RiNALMo outperformed models specifically pre-trained on 5' UTR sequences, suggesting that large-scale pre-training on diverse RNA families captures transferable sequence–function priors. This cross-family generalization mirrors the behaviour of large protein language models [9] and is a prerequisite for a universal RNA foundation model.

AIDO.RNA [13], with 1.6 billion parameters trained on 42 million curated ncRNA sequences, achieved state-of-the-art performance on 3D RNA inverse design, demonstrating that data quality is the dominant determinant of model performance — more so than scale alone. The frontier has shifted to generative architectures. UTR-LM [14], a BERT-based model augmented with secondary structure and minimum free energy supervision, was used to design a library of 211 novel 5' UTRs that achieved a 32.5% increase in protein production relative to established therapeutic 5' UTRs in wet-laboratory validation — providing proof-of-concept that LLMs can surpass empirically optimized clinical standards. GRAPE-LM [16] — a conditional autoencoder integrated with nucleic acid language model representations and guided by CRISPR–Cas-based intracellular aptamer screening data — achieved one-round evolution of high-affinity RNA aptamers against three clinically relevant targets (CD3ε, SARS-CoV-2 spike RBD, c-Myc), collapsing what conventionally requires 10–15 SELEX rounds into a single computational generation cycle. An overview of this end-to-end LLM framework is illustrated in Figure **1**.

Benchmarking and Persistent Performance Gaps

Systematic benchmarking by Bugnon *et al*.17 revealed that model scale and training data diversity are the dominant performance determinants for RNA secondary structure prediction — consistent with neural scaling laws in natural language processing. Key characteristics of representative RNA language models are summarised in Table **1**. However, the same analysis identified persistent limitations: pseudoknot prediction accuracy remains poor across all architectures; performance degrades sharply on RNA families underrepresented in training data; and existing benchmarks may be compromised by data leakage through structural homology. The most demanding evaluation scenario — where the test RNA family is entirely absent from the training set — exposes a generalization gap that large models partially but incompletely address [38]. Closing this gap requires RNA-family-aware data curation protocols analogous to those developed for protein benchmark design [49].

Critical Appraisal: Comparative Strengths, Limitations, and Practical Applicability

The models summarised in Table **1** differ not only in scale but in the practical trade-offs they present for therapeutic RNA design, and a purely descriptive

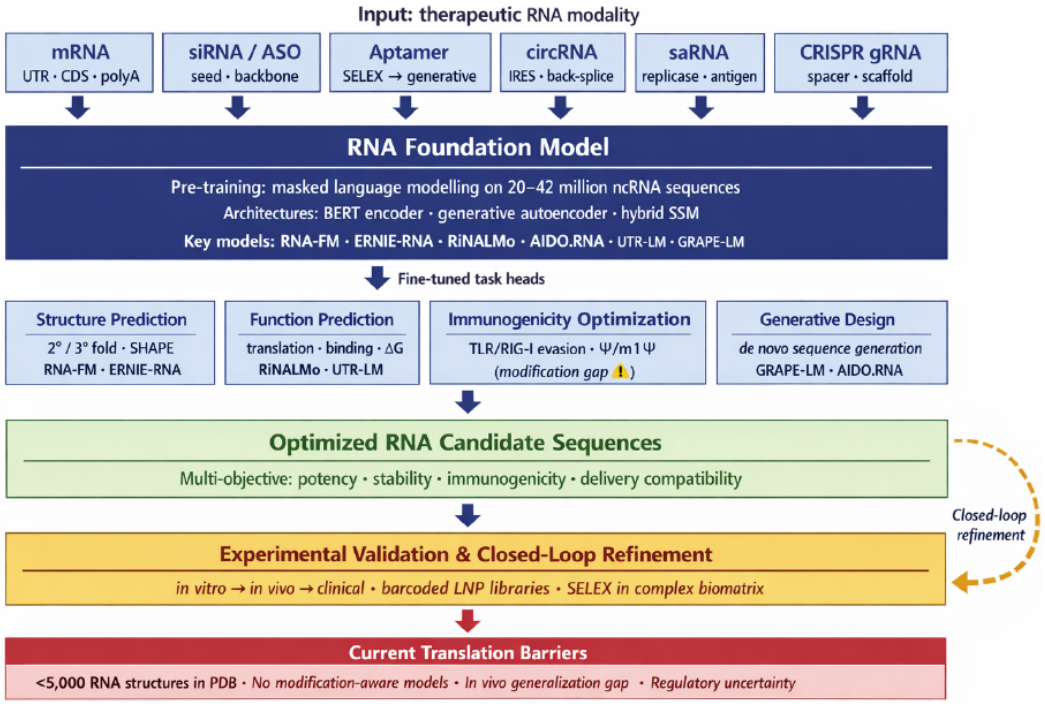

Figure 1: End-to-end LLM framework for therapeutic RNA design.

Table 1: Representative RNA Language Models and their Key Characteristics

Model	Parameters	Training data	Architecture	Key benchmark	Reference
RNA-FM	100 M	23.7 M ncRNA (RNAcentral)	BERT encoder	Structure, func. & RBP prediction	Chen *et al*. 2022 (Nat. Methods 2024)
ERNIE-RNA	86 M	20.4 M ncRNA	Structure-aware BERT	F1=0.873 on bpRNA-1m	Yin *et al*., Nat. Commun. 16, 10076 (2025)
RiNALMo	650 M	36 M ncRNA (RNAcentral + Rfam)	Large BERT encoder	Inter-family generalization; UTR translation efficiency	Penić *et al*., Nat. Commun. 16, 5671 (2025)
AIDO.RNA	1.6 B	42 M ncRNA (curated)	Encoder-only LM	State-of-art 3D RNA inverse design	Wang *et al*., bioRxiv 2024.11.28 (2024)
UTR-LM	~25 M	5' UTR sequences (multi-species)	BERT + struct. & MFE supervision	+32.5% protein yield vs. clinical benchmark	Chu *et al*., Nat. Mach. Intell. 6, 449–460 (2024)
GRAPE-LM	N/A	CRISPR–Cas intracellular screening data	Conditional autoencoder + nucleic acid LM	One-round aptamer evolution (CD3ε, SARS-CoV-2, c-Myc)	Zhang *et al*., Nat. Biotechnol. (2025). doi:10.1038/s41587-026-03007-5

BERT, bidirectional encoder representations from transformers; LM, language model; MFE, minimum free energy; ncRNA, non-coding RNA; N/A, not publicly reported.

inventory understates these differences. Table **2** extends this comparison with standardized evaluation dimensions — architecture class, training-data provenance, principal benchmark result, key strength, key limitation, practical applicability, and validation status — so that computational benchmark performance is explicitly distinguished from experimentally or clinically confirmed findings.

Three patterns emerge from this comparison. First, encoder-only masked language models (RNA-FM, ERNIE-RNA, RiNALMo, AIDO.RNA) are best suited to representation-learning tasks — structure annotation, function prediction, variant-effect scoring — in which a fixed-length embedding feeds a downstream predictor; none generates novel sequences de novo, which limits their direct applicability to design tasks without a separate generative module. Second, purpose-built generative systems (UTR-LM, GRAPE-LM) show that task-specific supervision integrated with language-model representations can outperform general-purpose scale, at the cost of narrower applicability outside the modality for which they were built. Third, the strength of evidence varies substantially across models: UTR-LM's and GRAPE-LM's central claims rest on wet-laboratory or cellular validation, whereas RNA-FM, ERNIE-RNA, RiNALMo, and AIDO.RNA report exclusively in silico benchmark performance against held-out sequence databases — a distinction that should inform how much weight each finding is given in translational decision-making, and that we return to throughout this review.

LLM-Driven Design Across Therapeutic RNA Modalities

mRNA Therapeutics

Messenger RNA therapeutics present the most architecturally complex design challenge: simultaneous optimization of the 5' cap compatibility, 5' UTR for ribosome recruitment, coding sequence (CDS) for translational efficiency, 3' UTR for mRNA half-life, and poly(A) tail architecture. UTR-LM's wet-laboratory validation [14] — demonstrating experimentally superior translation from LLM-designed 5' UTRs relative to established clinical sequences — provides proof-of-concept that foundation models can outperform empirical optimization.

Complementary analyses have established that CDS-level sequence composition, not UTR engineering alone, is the dominant determinant of translational efficiency [6], with direct implications for design effort prioritization. A critical unresolved challenge is integrating chemical modification chemistry — specifically m1Ψ patterns essential for clinical-grade mRNA — into LLM sequence representations; this limitation is addressed in Section 5.

Table 2: Comparative Critical Appraisal of Representative RNA Language Models

Model	Architecture & training data	Principal benchmark result	Key strength	Key limitation	Practical applicability	Validation status
RNA-FM	BERT encoder, 100 M params; 23.7 M ncRNA (RNAcentral)	Structure, function & RBP-interaction prediction	Learns structure and evolutionary signal without structural supervision; broad ncRNA family coverage	Encoder-only — no de novo sequence generation; canonical A/U/G/C tokens only	Upstream annotation/screening tool feeding a separate design module	Computational (in silico benchmarks)
ERNIE-RNA	Structure-aware BERT, 86 M params; 20.4 M ncRNA	F1 = 0.873 on bpRNA-1m (secondary structure)	Structural priors internalized without fine-tuning; interpretable attention maps	Benchmark restricted to secondary structure; pseudoknots and tertiary contacts remain poorly captured	Structure-informed target-site and accessibility screening	Computational
RiNALMo	650 M params; 36 M ncRNA (RNAcentral + Rfam)	Inter-family generalization; outperformed dedicated 5′UTR models on translation-efficiency prediction	Cross-family transfer; largest open-source encoder in this comparison	Compute-intensive; generalization gap persists for families wholly absent from pre-training	General-purpose embedding backbone for downstream fine-tuning across modalities	Computational
AIDO.RNA	1.6 B params; 42 M curated ncRNA	State-of-the-art performance on 3D RNA inverse-design benchmark subset	Demonstrates curated data quality outweighs raw parameter scale	Preprint status (not yet peer-reviewed at time of writing); largest compute/inference footprint	Research-stage; structure and inverse-design tasks	Computational
UTR-LM	~25 M params; multi-species 5′UTR sequences + structure/MFE supervision	+32.5% protein yield vs. established clinical 5′UTR benchmark	Only model in this comparison with prospective wet-laboratory confirmation of a therapeutically relevant output	Scope confined to 5′UTR design; does not address CDS, 3′UTR, or full-length mRNA	Directly deployable for 5′UTR optimization in mRNA drug candidates	Experimental (wet-laboratory confirmed)
GRAPE-LM	N/A (parameters not disclosed); CRISPR–Cas intracellular screening data + conditional autoencoder / nucleic-acid LM	One-round aptamer evolution against CD3ε, SARS-CoV-2 spike RBD, and c-Myc	Collapses multi-round SELEX into a single generative cycle; validated against a historically aptamer-resistant target (c-Myc)	Demonstrated for only three targets to date; full architecture and training data not publicly disclosed	Aptamer-discovery pipelines seeking to shorten SELEX timelines	Experimental (cellular/binding assays)

RBP, RNA-binding protein; SELEX, Systematic Evolution of Ligands by Exponential Enrichment; MFE, minimum free energy; N/A, not publicly reported. "Computational" indicates validation limited to in silico benchmarks against held-out sequence databases; "Experimental" indicates confirmation by wet-laboratory or cellular assay.

siRNA and Antisense Oligonucleotides

The design of potent siRNAs requires navigating thermodynamic, structural, and accessibility constraints while minimizing seed-region-mediated off-target silencing. LLM-based approaches offer qualitative advances over earlier machine-learning models by capturing long-range sequence context and enabling transfer learning from pre-trained RNA representations. A particularly productive pipeline integrates RNA LLM

structure predictions to identify accessible regions of target mRNA — sites where hybridization is not sterically impeded by secondary structure — with LLM-derived sequence activity models, creating an end-to-end workflow from target mRNA to optimal silencing agent. For ASOs, the diversity of backbone chemistries (phosphorothioate, LNA, 2'-OMe, 2'-fluoro, morpholino) and their differential effects on RNase H recruitment and nuclease stability constitute a multi-parameter optimization problem that modification-aware LLMs could, in principle, address holistically.

RNA Aptamers

RNA aptamers represent perhaps the clearest demonstration of LLM-driven design in this review period. Wong *et al*.18 demonstrated that a structure-to-sequence generative platform (RhoDesign) could produce Mango light-up aptamer variants with experimentally verified fluorescent activity — sequences that were sequence-dissimilar to all known Mango aptamers but structurally similar as predicted by RhoFold, demonstrating that structural predictions can guide targeted design of functional RNA sequences without traversing the known sequence landscape. GRAPE-LM [16] extended this paradigm to therapeutically relevant targets including the intracellular oncogenic transcription factor c-Myc, demonstrating that one-round generative aptamer evolution can produce high-affinity binders against targets that have historically resisted aptamer development. Earlier generative aptamer work by Iwano *et al*. 19, using variational autoencoders coupled with an *in silico* SELEX framework, established key methodological foundations that subsequent LLM-based approaches have built upon.

Circular RNA, saRNA, and CRISPR Guide RNAs

LLM applications to circular RNA design remain nascent. The key sequence engineering challenges — optimal internal ribosome entry site (IRES) selection, back-splicing junction design, and avoidance of internal structures that impede ribosome processivity — are accessible in principle to foundation models pre-trained on diverse RNA databases, but circRNA-specific training datasets remain sparse [4]. Self-amplifying RNAs pose compounded complexity: the replicase coding sequence is functionally constrained, and heterologous sequence compatibility with the replicase system imposes hard boundary conditions on the design space. For CRISPR guide RNAs, LLM-based approaches have demonstrated utility primarily in off-target activity prediction [24] — identifying 20-nucleotide spacer sequences that minimize Cas cleavage at genomic off-target sites — with more sophisticated applications in base-editor and prime-editor pegRNA engineering at early development stages.

Critical Barriers to Clinical Translation

The RNA Structural Data Famine

The most fundamental barrier to RNA LLM performance is data scarcity at the level of three-dimensional structure. Fewer than 5,000 RNA structures are deposited in the Protein Data Bank — compared with more than 220,000 protein structures — reflecting the practical difficulty of crystallizing conformationally flexible RNA molecules [5]. This 40-fold asymmetry fundamentally limits supervised learning for 3D RNA structure and constrains LLMs' ability to learn tertiary structural priors from data. Integration of chemical probing data — SHAPE-MaP, DMS-MaPseq, PARIS, and related high-throughput in-cell structure probing methods [36] — into LLM pre-training represents a promising strategy to supplement experimental tertiary structure data. However, chemical probing captures secondary structure propensities, not tertiary contacts, and models trained on probing data must be validated against crystallographic and cryo-EM structures to avoid artifacts of the probing methodology itself.

The Chemical Modification Gap

All clinically approved RNA therapeutics incorporate chemical modifications — phosphorothioate backbones, 2'-sugar modifications (2'-OMe, 2'-F, LNA), and nucleobase modifications (Ψ, m1Ψ, m6A) — that profoundly alter nuclease stability, immunogenicity, and pharmacokinetics [20]. Yet all current RNA LLMs are pre-trained exclusively on canonical four-letter RNA sequences. This is not a peripheral gap: a sequence designed by a canonical-nucleotide LLM and subsequently modified for clinical use may exhibit structural and immunogenic properties that differ substantially from those the model was trained to predict. Developing modification-aware RNA LLMs — incorporating modified nucleotide tokenization, pre-training on modification-annotated experimental datasets, and validation against clinical pharmacology endpoints — is a first-order priority for the field.

The clinical significance of this gap is illustrated by mRNA vaccines: replacing uridine with N1-methylpseudouridine (m1Ψ) not only dampens RIG-I-

and Toll-like-receptor-mediated innate immune sensing but also alters ribosome dwell time and local secondary-structure stability at each substituted position, so that a codon-optimization or UTR-design solution derived from an unmodified-sequence LLM does not necessarily transfer to the m1Ψ-substituted therapeutic product. Analogous effects apply to siRNA and ASO chemistries: 2′-O-methyl and 2′-fluoro substitutions alter duplex thermodynamics and nuclease resistance in ways invisible to a model trained only on unmodified sequence, and phosphorothioate backbone substitutions change protein-binding and pharmacokinetic behaviour independently of the underlying nucleotide sequence. Closing this gap will likely require an expanded tokenization scheme that assigns distinct tokens to modified nucleosides rather than treating modification as a post hoc annotation, pre-training corpora that pair canonical and modified versions of the same sequence family, and downstream validation against pharmacological endpoints — immunogenicity, half-life, and potency — rather than sequence-based benchmarks alone.

The In Vivo Generalization Problem

Current RNA LLMs are trained and evaluated on in vitro assay data: translation efficiency in cell-free systems, binding affinity by surface plasmon resonance, secondary structure from purified RNA preparations. Translation to *in vivo* pharmacological activity involves multiple layers of biological complexity — intracellular trafficking, endosomal escape, competition with endogenous RNA-binding proteins, RNA quality control mechanisms, and organ-level distribution determined by the delivery vehicle — that no current model captures. High-throughput *in vivo* mRNA translation datasets using barcoded LNP-delivered mRNA libraries with sequencing-based quantification would provide the training signal necessary to close this gap. Similarly, aptamer SELEX in complex biological matrices rather than purified buffer would generate training data more predictive of *in vivo* pharmacological performance [19].

Model Bias and Data Representativeness

A further limitation, distinct from data scarcity in absolute terms, is the compositional bias of existing training corpora. RNAcentral and Rfam — the principal data sources for RNA-FM, ERNIE-RNA, and RiNALMo — are dominated by well-characterized non-coding RNA families (tRNA, rRNA, snoRNA) from a comparatively narrow set of model organisms, while therapeutically relevant sequence space — engineered aptamers, synthetic UTRs, non-natural circRNA back-splice junctions — is systematically underrepresented. Models pre-trained predominantly on naturally occurring ncRNA may therefore encode priors poorly matched to the synthetic sequences characteristic of therapeutic RNA design, a distributional bias that compounds the data-scarcity and modification gaps discussed above rather than existing independently of them. Benchmark leakage compounds this concern further: because many evaluation sets are constructed from the same source databases as the pre-training corpora, reported benchmark performance can overstate real-world generalization — a caveat that applies to every encoder-only model in Tables 1 and 2, and one reason we treat benchmark scores as a lower, not sufficient, bound on clinical readiness.

A Roadmap for Clinical Translation

Near-Term: Data Infrastructure and Modification-Aware Models (2025–2027)

The most actionable near-term priorities are infrastructural. First, a coordinated, community-wide effort to generate and publicly deposit high-throughput RNA sequence–function data across all therapeutic modalities — analogous to the ProteinGym benchmark [49] — is required. This initiative must incorporate modification-annotated sequences to enable training of modification-aware models. Second, modification-native RNA LLMs must be developed: architectures that tokenize chemically modified nucleosides as first-class vocabulary elements, pre-trained on datasets including the modification patterns used in clinical RNA therapeutics. Third, RNA-family-aware benchmark protocols — with rigorous control of sequence homology between training and test sets [38] — must become the standard for model evaluation. The BEACON benchmark [48] represents an important step in this direction and should be extended to cover functional prediction tasks beyond structural annotation.

Translating these priorities into concrete deliverables, we recommend: (i) a federated, versioned RNA sequence–function data repository — spanning all six modalities discussed here and explicitly annotated for chemical modification, species, and experimental assay type — governed by a pre-competitive consortium of academic, industry, and regulatory stakeholders; (ii) release of at least one modification-aware RNA LLM per major therapeutic class (mRNA, siRNA/ASO), benchmarked directly against the

corresponding unmodified-sequence model to quantify the performance gap identified in this review; (iii) mandatory family-holdout evaluation — excluding any RNA family present in pre-training from the corresponding test set — as a condition of benchmark publication, extending the precedent set by BEACON [48]; and (iv) public deposition of negative results and failed design iterations alongside successful ones, since publication bias toward positive results is likely to inflate perceived model performance relative to real-world design success rates.

Medium-Term: The AI–RNA Closed Design Loop

The medium-term vision is a closed-loop design system in which LLM-guided computational generation, high-throughput experimental screening, and model refinement operate in an integrated cycle. GRAPE-LM's one-round aptamer generation [16] represents an early realization of this paradigm. Extending the closed loop to mRNA therapeutics — where LNP-delivered barcoded mRNA libraries screened *in vivo* provide feedback to iteratively refine generative models — and to siRNA design with *in vivo* efficacy and off-target readouts, is the central technical challenge for this period. The closed loop changes the economics of RNA drug discovery: each iterative cycle compresses the time-to-candidate, and the data generated in each cycle enriches the training set for the next generation of models.

Long-Term: Regulatory Science and Universal Design Platforms

Realizing the full potential of LLM-driven RNA therapeutics requires parallel investment in regulatory science. Fundamental questions remain open regarding validation evidence sufficient for clinical entry, how training data and model architecture should be disclosed in regulatory submissions, and how to assess latent immunogenic or toxic properties in generatively designed sequences not observed in biological organisms. Pre-competitive collaboration between the biotechnology industry, academic researchers, and regulatory agencies to establish standards for LLM validation in therapeutic RNA design is an urgent priority. The digital nature of RNA — four-letter alphabet, deterministic synthesis chemistry, programmable target engagement — makes it uniquely suited to AI-driven design; ensuring its clinical potential is safely realized is among the most consequential challenges at the intersection of artificial intelligence and medicine.

Model validation for regulatory purposes raises questions distinct from academic benchmarking. Regulators evaluating an LLM-designed RNA sequence must be able to assess not only the manufactured product but the design process that produced it, which requires harmonized disclosure standards for training-data provenance, model architecture, and version control that do not yet exist across major regulatory agencies. Practical barriers compound this uncertainty: the absence of an agreed validation framework discourages early adoption of generative design tools for GMP-track candidates, because a sequence generated by an undocumented or frequently updated model complicates the reproducibility record expected in a regulatory submission. Establishing a tiered validation framework — distinguishing exploratory use of LLMs in early discovery, where full provenance disclosure may be unnecessary, from use in generating or modifying a specific clinical candidate sequence, where full disclosure would be expected — is a concrete near-term step that pre-competitive consortia and regulatory science initiatives could pursue jointly.

CONCLUSION

RNA language models have moved, within a few years, from proof-of-concept representation learners to systems capable of designing individual components of clinically relevant RNA therapeutics — most convincingly in 5′UTR optimization and aptamer discovery, where wet-laboratory validation already exists. The comparative appraisal presented here (Table **2**) shows that this progress is real but uneven: most encoder-only models remain confined to in silico benchmarks, chemical modification chemistry is still absent from mainstream architectures, and structural training data remain scarce relative to the protein field. Closing these gaps will require coordinated action on data infrastructure, modification-aware model architectures, standardized and family-holdout benchmarking, and tiered regulatory validation frameworks. None of these barriers appears fundamental; each is addressable with the roadmap outlined above. If pursued, LLM-driven RNA design is well positioned to become a routine, end-to-end engine for therapeutic RNA development over the coming decade.

CONSENT FOR PUBLICATION

Not applicable. This is a narrative literature review; it did not involve human participants, human data, or

animal experimentation, and no patient consent was required.

CONFLICT OF INTEREST

The author declares no financial or non-financial conflicts of interest related to this work.

ACKNOWLEDGEMENTS

The author thanks the editor and the two anonymous reviewers for their constructive comments, which substantially improved this manuscript. No specific funding was received for this work.

APPENDIX: GLOSSARY OF KEY TERMS

Large Language Model (LLM)

A neural network with hundreds of millions to billions of parameters pre-trained on massive sequence datasets using self-supervised objectives (e.g., masked token prediction), capable of learning biophysically meaningful representations of sequence–structure–function relationships without explicit supervision.

Foundation Model

A large pre-trained model that serves as the base for diverse downstream tasks via fine-tuning, without retraining from scratch. Protein LM analogues include ESM-2 (ref. [9]) and ProtTrans (ref. [28]).

SELEX

Systematic Evolution of Ligands by Exponential Enrichment; an iterative in vitro selection procedure for nucleic acid aptamers, conventionally requiring 10–15 rounds over several weeks. LLM-based approaches aim to replace iterative SELEX with single-round generative design.

Pseudouridine (Ψ) / N1-methylpseudouridine (m1Ψ)

Modified RNA nucleosides that reduce innate immune activation by TLR7 and TLR8 while maintaining translational activity; incorporated into all clinically approved mRNA therapeutics.

GalNAc Conjugation

Chemical conjugation of RNA drugs (siRNA, ASO) to N-acetylgalactosamine, enabling receptor-mediated hepatic delivery via the asialoglycoprotein receptor — the dominant delivery strategy for approved liver-targeted RNA therapeutics.

REFERENCES

[1] Adams D, *et al*. Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. N Engl J Med 2018; 379: 11-21. https://doi.org/10.1056/NEJMoa1716153

[2] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines - a new era in vaccinology. Nat Rev Drug Discov 2018; 17: 261-279. https://doi.org/10.1038/nrd.2017.243

[3] Ray KK, *et al*. Two phase 3 trials of inclisiran in patients with elevated LDL cholesterol. N Engl J Med 2020; 382: 1507-1519. https://doi.org/10.1056/NEJMoa1912387

[4] Wesselhoeft RA, Kowalski PS, Anderson DG. Engineering circular RNA for potent and stable translation in eukaryotic cells. Nat Commun 2018; 9: 2629. https://doi.org/10.1038/s41467-018-05096-6

[5] Schneider B, *et al*. When will RNA get its AlphaFold moment? Nucleic Acids Res 2023; 51: 9522-9532. https://doi.org/10.1093/nar/gkad726

[6] Mauger DM, *et al*. mRNA structure regulates protein expression through changes in functional half-life. Proc Natl Acad Sci USA 2019; 116: 24075-24083. https://doi.org/10.1073/pnas.1908052116

[7] Childs-Disney JL, *et al*. Targeting RNA structures with small molecules. Nat Rev Drug Discov 2022; 21: 736-762. https://doi.org/10.1038/s41573-022-00521-4

[8] Vaswani A, *et al*. Attention is all you need. Adv Neural Inf Process Syst 2017; 30: 5999-6009.

[9] Lin Z, *et al*. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 2023; 379: 1123-1130. https://doi.org/10.1126/science.ade2574

[10] Chen J, *et al*. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. Nat Methods 2024.

[11] Yin W, *et al*. ERNIE-RNA: an RNA language model with structure-enhanced representations. Nat Commun 2025; 16: 10076. https://doi.org/10.1038/s41467-025-64972-0

[12] Penić RJ, *et al*. RiNALMo: general-purpose RNA language models can generalize well on structure prediction tasks. Nat Commun 2025; 16: 5671. https://doi.org/10.1038/s41467-025-60872-5

[13] Wang N, *et al*. AIDO.RNA: a large-scale foundation model for RNA function and structure prediction. bioRxiv [Preprint] 2024. https://doi.org/10.1101/2024.11.28.625345

[14] Chu Y, *et al*. A 5' UTR language model for decoding untranslated regions of mRNA and function predictions. Nat Mach Intell 2024; 6: 449-460. https://doi.org/10.1038/s42256-024-00823-9

[15] Abramson J, *et al*. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024; 630: 493-500. https://doi.org/10.1038/s41586-024-07487-w

[16] Zhang J, *et al*. GRAPE-LM: single-round evolution of RNA aptamers with a generative AI framework. Nat Biotechnol 2025. https://doi.org/10.1038/s41587-026-03007-5

[17] Bugnon LA, *et al*. Comprehensive benchmarking of large language models for RNA secondary structure prediction. Brief Bioinform 2025; 26: bbaf137. https://doi.org/10.1093/bib/bbaf137

[18] Wong F, *et al*. Deep generative design of RNA aptamers using structural predictions. Nat Comput Sci 2024; 4: 829-839. https://doi.org/10.1038/s43588-024-00720-6

[19] Iwano N, *et al*. Generative aptamer discovery using RaptGen. Nat Comput Sci 2022; 2: 378-386. https://doi.org/10.1038/s43588-022-00249-6

[20] Karikó K, *et al*. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity 2005; 23: 165-175. https://doi.org/10.1016/j.immuni.2005.06.008

[21] Jumper J, *et al*. Highly accurate protein structure prediction with AlphaFold. Nature 2021; 596: 583-589. https://doi.org/10.1038/s41586-021-03819-2

[22] Nguyen E, *et al*. Sequence modeling and design from molecular to genome scale with Evo. Science 2024; 386: eado9336. https://doi.org/10.1126/science.ado9336

[23] Fire A, *et al*. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 1998; 391: 806-811. https://doi.org/10.1038/35888

[24] Joshi CK, Liò P. gRNAde: a geometric deep learning approach for 3D RNA inverse design. Methods Mol Biol 2025; 2847: 121-135. https://doi.org/10.1007/978-1-0716-4079-1_8

[25] Wang X, *et al*. Uni-RNA: universal pre-trained models revolutionize RNA research. bioRxiv [Preprint] 2023. https://doi.org/10.1101/2023.07.11.548588

[26] Shen T, *et al*. Accurate RNA 3D structure prediction using a language model-based deep learning approach. Nat Methods 2024; 21: 1-12. https://doi.org/10.1038/s41592-024-02487-0

[27] Devlin J, *et al*. BERT: pre-training of deep bidirectional transformers for language understanding. Proc NAACL-HLT 2019: 4171-4186.

[28] Elnaggar A, *et al*. ProtTrans: toward understanding the language of life through self-supervised learning. IEEE Trans Pattern Anal Mach Intell 2022; 44: 7112-7127. https://doi.org/10.1109/TPAMI.2021.3095381

[29] Garner AL. Contemporary progress and opportunities in RNA-targeted drug discovery. ACS Med Chem Lett 2023; 14: 251-259. https://doi.org/10.1021/acsmedchemlett.3c00020

[30] Riley AT, *et al*. Generative and predictive neural networks for the design of functional RNA molecules. Nat Commun 2025; 16: 1-17. https://doi.org/10.1038/s41467-025-59389-8

[31] Autour A, *et al*. Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. Nat Commun 2018; 9: 656. https://doi.org/10.1038/s41467-018-02993-8

[32] Huang H, *et al*. RiboDiffusion: tertiary structure-based RNA inverse folding with generative diffusion models. Bioinformatics 2024; 40: i347-i356. https://doi.org/10.1093/bioinformatics/btae259

[33] Wang N, *et al*. Multi-purpose RNA language modelling with motif-aware pretraining and type-guided fine-tuning. Nat Mach Intell 2024; 6: 1-10. https://doi.org/10.1038/s42256-024-00836-4

[34] Wayment-Steele HK, *et al*. RNA secondary structure packages evaluated and improved by high-throughput experiments. Nat Methods 2022; 19: 1234-1242. https://doi.org/10.1038/s41592-022-01605-0

[35] Celaj A, *et al*. An RNA foundation model enables discovery of disease mechanisms and candidate therapeutics. bioRxiv [Preprint] 2023. https://doi.org/10.1101/2023.09.20.558508

[36] Strobel EJ, Yu AM, Lucks JB. High-throughput determination of RNA structures. Nat Rev Genet 2018; 19: 615-634. https://doi.org/10.1038/s41576-018-0034-x

[37] Ji Y, *et al*. DNABERT: pre-trained bidirectional encoder representations from transformers for DNA-language in genome. Bioinformatics 2021; 37: 2112-2120. https://doi.org/10.1093/bioinformatics/btab083

[38] Szikszai M, *et al*. Deep learning models for RNA secondary structure prediction (probably) do not generalize across families. Bioinformatics 2022; 38: 3892-3899. https://doi.org/10.1093/bioinformatics/btac415

[39] Tan C, *et al*. RDesign: hierarchical data-efficient representation learning for tertiary structure-based RNA design. Proceedings of ICLR 2024.

[40] Wang Z, *et al*. AptaDiff: de novo design and optimization of aptamers based on diffusion models. Brief Bioinform 2024; 25: bbae517. https://doi.org/10.1093/bib/bbae517

[41] Li S, *et al*. CodonBERT: large language model for mRNA vaccines. Genome Res 2024; 34: 1027-1035. https://doi.org/10.1101/gr.278870.123

[42] Dauparas J, *et al*. Robust deep learning-based protein sequence design using ProteinMPNN. Science 2022; 378: 49-56. https://doi.org/10.1126/science.add2187

[43] Su J, *et al*. RoFormer: enhanced transformer with rotary position embedding. Neurocomputing 2024; 568: 127063. https://doi.org/10.1016/j.neucom.2023.127063

[44] Singh J, *et al*. RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. Nat Commun 2019; 10: 5407. https://doi.org/10.1038/s41467-019-13395-9

[45] Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinform 2010; 11: 129. https://doi.org/10.1186/1471-2105-11-129

[46] Lorenz R, *et al*. ViennaRNA Package 2.0. Algorithms Mol Biol 2011; 6: 26. https://doi.org/10.1186/1748-7188-6-26

[47] Rives A, *et al*. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci USA 2021; 118: e2016239121. https://doi.org/10.1073/pnas.2016239118

[48] Tian K, *et al*. BEACON: a comprehensive benchmark for RNA tasks. Adv Neural Inf Process Syst 2024; 37.

[49] Notin P, *et al*. ProteinGym: large-scale benchmarks for protein fitness prediction and design. Adv Neural Inf Process Syst 2023; 36. https://doi.org/10.1101/2023.12.07.570727