

参赛学生姓名： 邵子熙

中 学： 南京外国语学校

省 份： 江苏省

国家/地区： 中 国

指导老师姓名： 肖易倍，黄天意

指导老师单位： 中国药科大学

论文题目： ProMORNA: An End-to-End Generative

Framework for Full-Length mRNA Design

ProMORNA: An End-to-End Generative Framework for Full-Length mRNA Design

Zixi Shao

Abstract

Designing therapeutic messenger RNA (mRNA) requires generating full-length transcripts that optimally balance intracellular stability and translation efficiency. While generative models have advanced mRNA design, existing approaches rely on fragmented region-wise generation, computationally expensive inference-time search, or target-specific model updates. To overcome these limitations, we introduce ProMORNA, a zero-shot, end-to-end generative framework for full-length mRNA design. ProMORNA aligns the design process with modern Large Language Model (LLM) pipelines: it utilizes a BART-style architecture pre-trained on over 6 million natural protein-mRNA pairs, followed by a therapeutic supervised fine-tuning stage. Crucially, we propose Multi-Objective Group Relative Policy Optimization (MO-GRPO) to directly align the generative policy with conflicting biological constraints, avoiding direct scalarization of raw objective scores. *In silico* evaluations demonstrate that ProMORNA pushes the Pareto frontier of predicted stability and translation efficiency beyond standard baselines and GEMORNA, a leading state-of-the-art method. In three independent Fluc2P experiments, Candidate 2 achieved a 12–72 h reporter-activity AUC 3.33 times that of the conventional Benchmark (Holm-adjusted $P = 0.01285$), while its estimated 1.29-fold advantage over the reported best GEMORNA candidate remained exploratory. These results support ProMORNA as a framework for generating functionally active full-length mRNA candidates.

Keywords: mRNA Design, Generative Biology, Large Language Models, Multi-Objective Optimization, Reinforcement Learning

Contents

1	Introduction	3
2	Methods	5
2.1	Supervised Training Stages	5
2.2	Biological Objectives and Design Constraints	6
2.3	Multi-Objective GRPO	7
3	Results	9
3.1	<i>In silico</i> mRNA Generation Performance	9
3.2	<i>In Vitro</i> Validation	10
4	Conclusion	12
	References	13
A	Experimental Setup	16
A.1	Model Training	16
A.2	<i>In Silico</i> Experimental Setup	17
A.3	<i>In Vitro</i> Validation	17
B	Reinforcement Learning Dynamics	19
C	MO-GRPO Ablation	21
D	Additional mRNA Property Analysis	22

1 Introduction

The success of messenger RNA (mRNA) vaccines against SARS-CoV-2 has highlighted the broad potential of mRNA-based medicines.^{1,2,3,4} Creating effective mRNA medicines requires computational design of mRNA sequences based on the amino acid sequence of the desired protein.⁵

Full-length mRNA design can be viewed as a conditional generation and multi-objective optimization problem: given a protein sequence of amino acids, the goal is to generate an mRNA sequence of nucleotides that, (i) translates back into that protein, and (ii) optimizes for certain competing biological objectives. Specifically, a full-length mRNA sequence has a central coding sequence (CDS) bordered by untranslated regions (UTRs) at both ends. The CDS is divided into codons of three adjacent nucleotides, where each codon naturally translates into a certain amino acid determined by the genetic code. Thus, the number of codons in the CDS must be equal to the number of amino acids in the target protein, and each codon must be selected from a fixed set of synonymous codons that translate into the amino acid in the corresponding position. In contrast, the UTRs are largely unconstrained but work with the CDS to shape biomolecular dynamics and finally protein output.^{6,7}

The primary challenge of mRNA design is the large design space, which makes exhaustive search infeasible. As illustrated in Fig. 1, the number of possible CDSs alone for a given protein grows exponentially with the protein length, not even considering the unconstrained compositions of the UTRs.

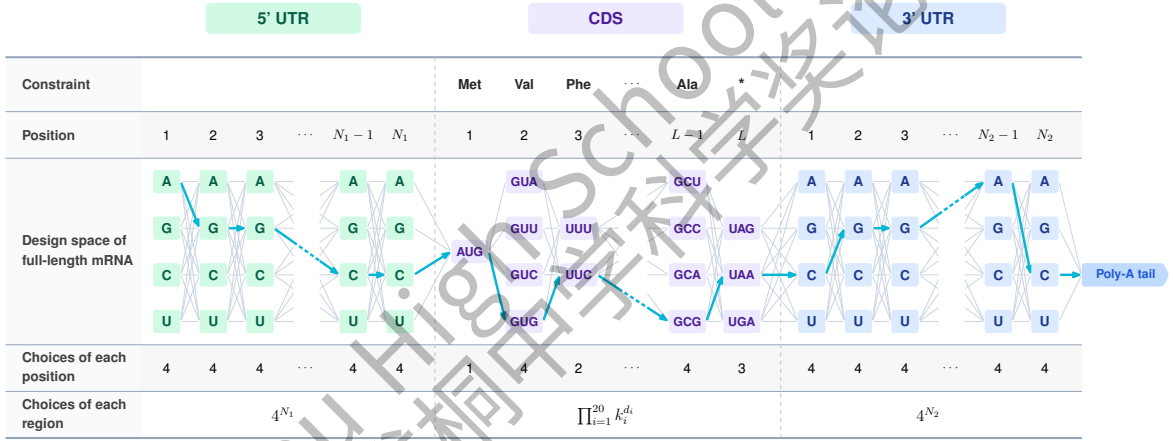


Figure 1: The vast potential mRNA search space of a target protein.

A primary objective in mRNA sequence design is to maximize functional protein expression, which is shaped by stability and translation efficiency.^{7,8} However, the complex biological mechanisms that affect these properties are still not fully understood by human research, making it difficult for rule-based algorithms to accurately predict and optimize for downstream biological properties. Conventional methods, such as LinearDesign,⁹ approach this challenge by jointly optimizing simple heuristic proxy metrics like Minimum Free Energy (MFE) and Codon Adaptation Index (CAI).¹⁰ Although these algorithms can achieve mathematical optimality within their defined search spaces, their real-world biological performance is fundamentally bounded by the limitations of these rigid, rule-based heuristics.

Recently, generative models have emerged to address limitations of traditional heuristic algorithms. For example, mRNAutilus¹¹ treated mRNA design as a discrete diffusion process guided by inference-time tree search; GEMORNA¹² used separate Transformer models to independently generate the 5' UTR, CDS, and 3' UTR. Most recently, mRNA-GPT¹³ proposed jointly optimizing these regions through full-length sequential generation using a unidirectional, decoder-only language model.

However, these existing generative frameworks are not strictly end-to-end and are limited by their architecture. mRNAutilus requires a CDS template and two UTR lengths as inputs to initialize the

masked diffusion process, which is constrained by human prior; GEMORNA does not directly optimize for downstream biological objectives, which limits the ability to explore candidates beyond the training distribution, and concatenates outputs from three independent models, which may overlook long-range context-dependency between different regions;^{7,14} mRNA-GPT is partially or fully blind to the protein target when generating the 5' UTR and CDS regions, being limited on planning ahead of the current amino acid.

Also, these methods require different forms of inference-time scaling: mRNAutilus runs tree search at every diffusion step; GEMORNA requires wet-lab screening on a large number of generated candidates and optionally utilizes beam search for CDS generation; mRNA-GPT applies Iterative Supervised Fine-Tuning (SFT) or Reinforcement Learning (RL) for each specific target from a fixed base model, generating thousands of candidates per iteration per target. The vast resources required for these models at inference time limit their application to practical scenarios such as personalized mRNA vaccines,^{15,16} where rapid computational turnaround is critically required.

To overcome these challenges, we propose **ProMORNA (Protein-conditioned Multi-Objective RNA generation)**, an end-to-end generative framework for full-length mRNA design. First, we pre-train a BART-style encoder-decoder generative model on public data. Next, we employ post-training strategies inspired by modern LLM training¹⁷ to optimize for biological metrics. *In silico* experiments illustrate that ProMORNA improves the Pareto frontier for predicted stability and translation efficiency relative to standard supervised baselines and the existing generative mRNA design method, GEMORNA.¹² Wet-lab evaluation of seven Pareto-selected candidates on Fluc2P provides functional evidence for the generated designs.

We summarize the core contributions of our framework as follows:

- **Protein-conditioned end-to-end mRNA generation:** We introduce an encoder-decoder framework that generates complete mRNA transcripts, including the 5' UTR, CDS, and 3' UTR, directly from a target protein sequence, without requiring additional inputs at inference time. The model is given maximum freedom in generation and optimization, and only a minimal set of basic biological rules is enforced throughout the whole pipeline.
- **Target-agnostic model alignment:** Our target-agnostic RL post-training generalizes to unseen proteins, enabling rapid, high-quality zero-shot generation without requiring target-specific model updates. Furthermore, the autoregressive model natively supports inference-time scaling, allowing users to optionally generate and filter large candidate pools to further explore the Pareto frontier when computational budget permits.
- **Effective multi-objective optimization:** We introduce a multi-objective variant of Group Relative Policy Optimization (GRPO),¹⁷ MO-GRPO, used in our model post-training. By computing relative advantages at the metric level prior to aggregation, MO-GRPO avoids direct scalarization of raw biological scores with different scales, stabilizing the RL process.
- **Pushing the Pareto frontier *in silico*:** We evaluate ProMORNA *in silico* against existing design strategies and assess generated candidates across translation efficiency, stability, normalized Minimum Free Energy (MFE), uridine content (U-content), and UTR length plausibility. Notably, we find that ProMORNA significantly improves the performance over supervised baselines and GEMORNA on the two primary predicted objectives: predicted stability and predicted translation efficiency.
- **Prospective biological validation:** We synthesize and evaluate seven Pareto-selected ProMORNA candidates alongside a conventional mRNA Benchmark and the GEMORNA design GMR-FL-F5.¹² Three independent experiments identify candidates with higher time-integrated reporter activity than Benchmark.

2 Methods

In this section, we detail the **ProMORNA** framework. First, we outline the supervised pre-training and fine-tuning stages of model. Next, we define the biological metrics used to evaluate transcript quality. Finally, we introduce Multi-Objective GRPO (MO-GRPO), which aligns the model with these multi-dimensional biological constraints.

2.1 Supervised Training Stages

Training Dataset. We curated vertebrate transcripts from NCBI RefSeq,¹⁸ prioritizing shorter sequences to ensure computational tractability and to remain within a practical length regime for future *in vitro* synthesis. We filtered out sequences labeled as “pseudo” and retained transcripts with canonical start and stop codons, a verified N-terminal methionine, and minimum 5’ and 3’ UTR lengths of 20 nt and 30 nt, respectively. Additionally, all retained sequences were required to fit within our 1,024-token limit under our hybrid tokenization scheme. This pipeline yielded 6,089,519 natural protein-mRNA pairs. From these data, the model learns the conditional distribution $P(y|x)$ of a full-length mRNA transcript y given a target protein sequence x , implicitly capturing natural vertebrate codon usage biases and plausible UTR compositions.

Model Architecture. Our model uses an encoder-decoder architecture based on the BART-small configuration¹⁹ with a context limit of 1,024 tokens applied to both the encoder and the decoder, comprising 45.2M trainable parameters in total. As shown in Fig. 2, the encoder processes the input amino acid sequence of the target protein, while the decoder autoregressively generates the corresponding full-length mRNA sequence step by step. Because the encoder processes the full protein sequence before decoding begins, the decoder can condition every generated mRNA token on the complete target protein.

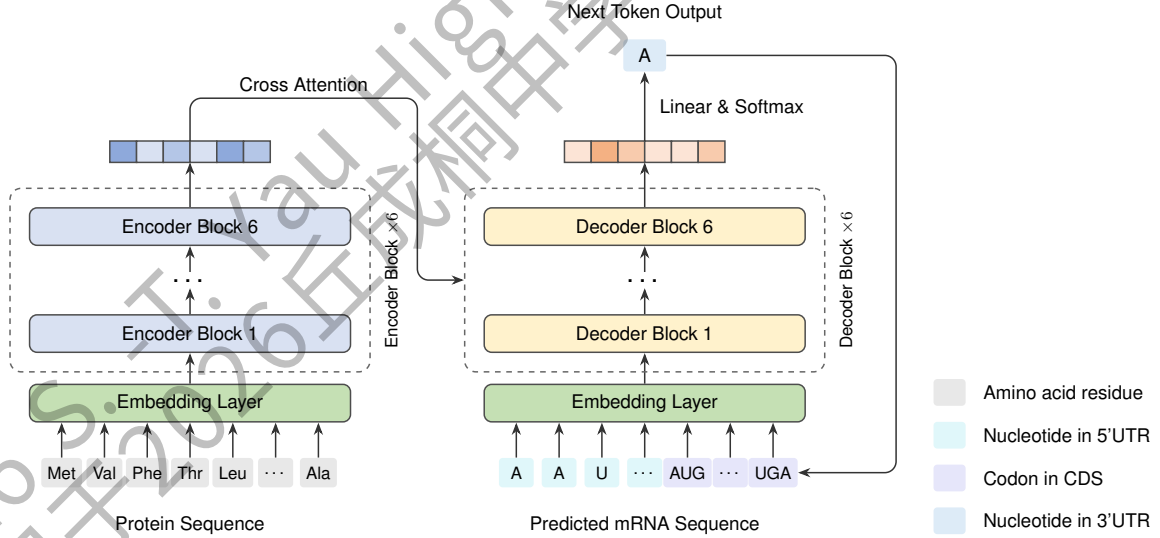


Figure 2: Modeling protein-to-mRNA with BART.

Tokenization. We apply per-amino-acid tokenization for the source protein sequences. For target mRNA sequences, we use a hybrid vocabulary: per-nucleotide tokenization for both UTRs, and per-codon tokenization for the CDS. Special separator tokens are inserted to mark region boundaries and prompt the model to switch between nucleotide-level and codon-level vocabularies.

Training Objective. The model is pre-trained by minimizing the standard sequence-to-sequence negative log-likelihood:

$$\mathcal{L}(\theta) = -\frac{1}{T} \sum_{t=1}^T \log P_{\theta}(y_t | y_{<t}, \mathbf{x}), \quad (1)$$

where θ represents the trainable parameters of the model, T is the total length of the target mRNA sequence, y_t is the ground-truth target token at time step t , $y_{<t}$ denotes the sequence of preceding target tokens, and $\mathbf{x}$ is the tokenized source input amino acid sequence.

Supervised Fine-Tuning. Aligning with modern LLM training paradigms, we employ SFT to establish a high-quality cold start for reinforcement learning. While pre-training captures broad natural evolutionary distributions, therapeutic mRNAs require maximized protein expression. Therefore, we curate a purified subset of 37,369 NM-prefixed RefSeq mRNAs strictly lacking upstream open reading frames (uORFs), which are natural regulatory elements that suppress translation.²⁰ Fine-tuning the base model on this subset shifts its generative prior toward therapeutic efficacy. Crucially, this step provides a biologically aligned reference policy for the subsequent RL phase, ensuring that the KL-divergence penalty does not force the model to mimic undesirable natural traits.

2.2 Biological Objectives and Design Constraints

We define a set of $K = 5$ separate biological objectives for RL optimization. Let $r_{i,k}$ denote the raw score of the i -th sequence on the k -th objective. The following metrics are used:

Predicted Stability ($k = 1$): Stability is an inherent biochemical property of an mRNA molecule that governs its resistance to enzymatic degradation and thus defines its functional persistence within the cellular environment. It is predicted by an additional ridge-regression predictor pre-trained on latent embedding features extracted via mRNABERT,²¹ as shown in Fig. 3. This objective is maximized.

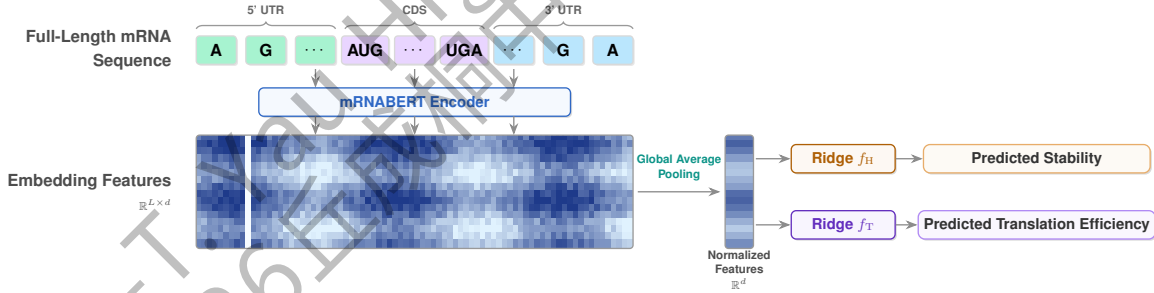


Figure 3: Predicting mRNA stability and translation efficiency with mRNABERT.

Predicted Translation Efficiency ($k = 2$): The efficiency of translating an mRNA into protein, predicted with mRNABERT via another prediction head (Fig. 3). This objective is also maximized.

Normalized Minimum Free Energy (MFE, $k = 3$): We minimize the folding energy predicted by LinearFold,²² divided by sequence length. This objective serves as a robust proxy for the thermodynamic stability of the secondary structure.

U-content ($k = 4$): The fractional proportion of Uridine (U) nucleotides within the entire sequence. We explicitly monitor this, as AU-rich sequence elements are associated with mRNA turnover²³ and

uridine-containing RNA can activate innate immune sensing pathways.^{24,25} Thus, this objective is minimized.

UTR Length Plausibility ($k = 5$): We encourage biologically plausible UTR lengths with Gaussian-shaped priors over the 5' and 3' UTR lengths:

$$r_{i,5} = \sum_{j \in \{5,3\}} \exp \left(-\frac{1}{2} \left(\frac{u_{i,j} - \mu_j}{\sigma_j} \right)^2 \right), \quad (2)$$

where μ is the length of the corresponding wild-type UTR observed in our training dataset, and σ is the standard deviation of all UTR lengths in our dataset. This is only a regularization term used during training and is not evaluated when selecting candidates for a novel protein.

Generation Constraints. We use constrained decoding during generation so that each partial output remains a valid prefix. For the CDS, the decoder is restricted to synonymous codons of the corresponding target amino acid; for UTRs, it generates nucleotide tokens or allowed special tokens. Since this only guarantees prefix-level validity, we further apply a post-hoc check to the complete transcript, verifying region boundaries, frame consistency, start and stop codons, and exact translation into the target protein. Following LinearDesign,⁹ sequences with long double-stranded regions that include 33 or more base pairs are also deemed invalid to reduce unwanted innate immune responses. Invalid sequences bypass predictor evaluation, receive maximal penalty, and are excluded from standard deviation calculation.

2.3 Multi-Objective GRPO

While standard RL algorithms collapse multiple evaluation metrics into a single explicit scalar reward, this static scalarization can be unstable for biological sequence design. Biological metrics have different scales and their variances shift asynchronously as the policy evolves.

To address this issue, we propose Multi-Objective GRPO (MO-GRPO). MO-GRPO first standardizes each objective within a rollout group and then aggregates the standardized scores, avoiding a hand-crafted scalar reward over raw objective values (Fig. 4).

For a given prompt, the policy samples a group of G complete mRNA sequences. Let μ_k and σ_k be the empirical mean and standard deviation of the k -th objective scores within the current group. The aggregated baseline advantage A_i for sequence i is formulated as:

$$A_i = \frac{1}{\sqrt{\sum_{k=1}^K w_k^2}} \sum_{k=1}^K w_k \cdot \text{clip} \left(\frac{r_{i,k} - \mu_k}{\sigma_k + \delta_k}, -c, c \right), \quad (3)$$

where w_k denotes the relative importance and optimization direction of the k -th objective ($w_k > 0$ for maximization and $w_k < 0$ for minimization). Normalizing the weights by their L_2 norm mathematically guarantees that, assuming the objectives are mutually orthogonal, the standard deviation of the aggregated advantages $\text{std}(\{A_1, \dots, A_G\})$ remains approximately 1.

To improve training stability, MO-GRPO uses three additional mechanisms. First, we introduce a dynamic damping factor δ_k for each objective. When the group variance σ_k drops significantly ($\sigma_k \ll \delta_k$), δ_k smoothly pushes the normalized advantage contribution of this metric toward zero. This naturally prevents numerical instability and shifts the model’s exploration focus toward other under-optimized metrics.

Second, we apply extreme-value clipping with bound $c = 3.0$ to limit the influence of outlier objective values. This is especially useful when reward components are computed from heuristic or learned proxy predictors, whose outputs may occasionally contain large deviations.

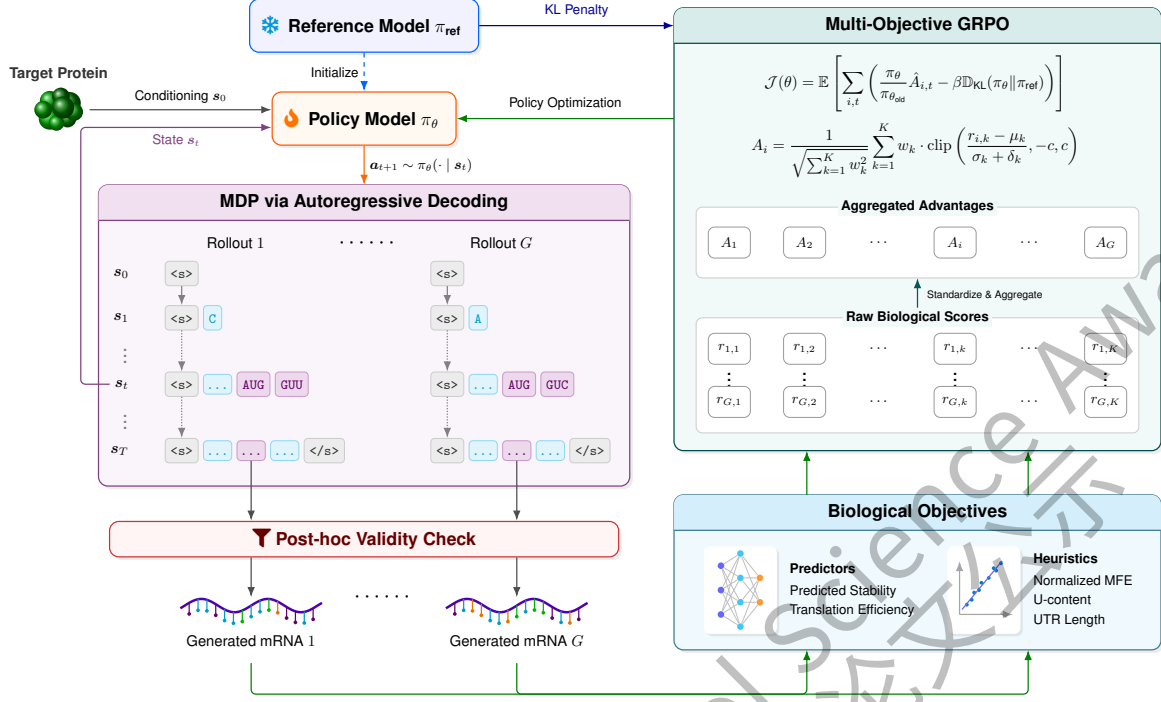


Figure 4: ProMORNA’s post-training framework. $\langle s \rangle$ is the start token and $\langle /s \rangle$ is the end token. Note that this is only a demonstration of the rollout process; in practice, mRNA sequences of varying lengths can be generated for the same target protein.

Third, because clipping and damping can shift the group-level mean away from zero, we apply a post-centralization step to derive the final token-level advantage $\hat{A}_{i,t}$:

$$\hat{A}_{i,t} = A_i - \frac{1}{G} \sum_{j=1}^G A_j. \quad (4)$$

Importantly, we do not re-divide by the standard deviation in this step. This preserves the adaptive magnitude derived from the damping factor δ_k , allowing the advantage to decay naturally as the policy converges. For invalid sequences, we directly assign a maximal penalty $A_i = -c \sum |w_k| / \sqrt{\sum w_k^2}$ before post-centralization, thereby incorporating invalid samples into the policy update.

Finally, the policy π_θ is updated to maximize:

$$\mathcal{J}(\theta) = \mathbb{E}_{\substack{\mathbf{s}_0 \sim P(S_0) \\ \{\mathbf{a}_i\}_{i=1}^G \sim \pi_{\text{old}}(\cdot | \mathbf{s}_0)}} \left[\frac{1}{\sum_{i=1}^G |\mathbf{a}_i|} \sum_{i=1}^G \sum_{t=1}^{|\mathbf{a}_i|} \left(\frac{\pi_\theta(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})}{\pi_{\text{old}}(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})} \hat{A}_{i,t} - \beta \mathbb{D}_{\text{KL}}[\pi_\theta || \pi_{\text{ref}}] \right) \right], \quad (5)$$

where the KL divergence is approximated analytically as

$$\mathbb{D}_{\text{KL}}(\pi_\theta || \pi_{\text{ref}}) = \frac{\pi_\theta(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})}{\pi_{\text{old}}(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})} \left(\frac{\pi_{\text{ref}}(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})}{\pi_\theta(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})} - \log \frac{\pi_{\text{ref}}(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})}{\pi_\theta(\mathbf{a}_{i,t} | \mathbf{s}_{i,t})} - 1 \right). \quad (6)$$

Here, different from the original GRPO algorithm,¹⁷ we (i) use a token-level policy gradient loss,²⁶ assigning equal weights to every generated token; (ii) estimate the KL penalty with an importance-weighted estimator,²⁷ which eliminates bias introduced by evaluating samples drawn from the old policy.

3 Results

This section presents *in silico* and *in vitro* results of ProMORNA on the widely used Fluc2P target. We compare ProMORNA with GEMORNA and supervised baselines to evaluate its mRNA generation performance. Section A presents the experimental setup, including the target protein, baselines, parameters, generation protocol, and the reward predictor training. Section B presents RL learning curves. Ablation studies demonstrate that standard scalarized GRPO suffers from rapid reward hacking and KL-collapse, underscoring the necessity of MO-GRPO’s metric-level advantage formulation (Section C). Our additional analyses (see Section D) demonstrate that ProMORNA automatically learned several known biological rules, such as optimized GC content and structures more conducive to regulating protein expression. These findings may further indicate that the model enhanced its understanding of sequence structure in the absence of explicit supervisory signals.

3.1 *In silico* mRNA Generation Performance

In this subsection, we evaluate ProMORNA on Fluc2P using the reward-related metrics. We compare ProMORNA with GEMORNA and two supervised ablations, BASE and SFT.

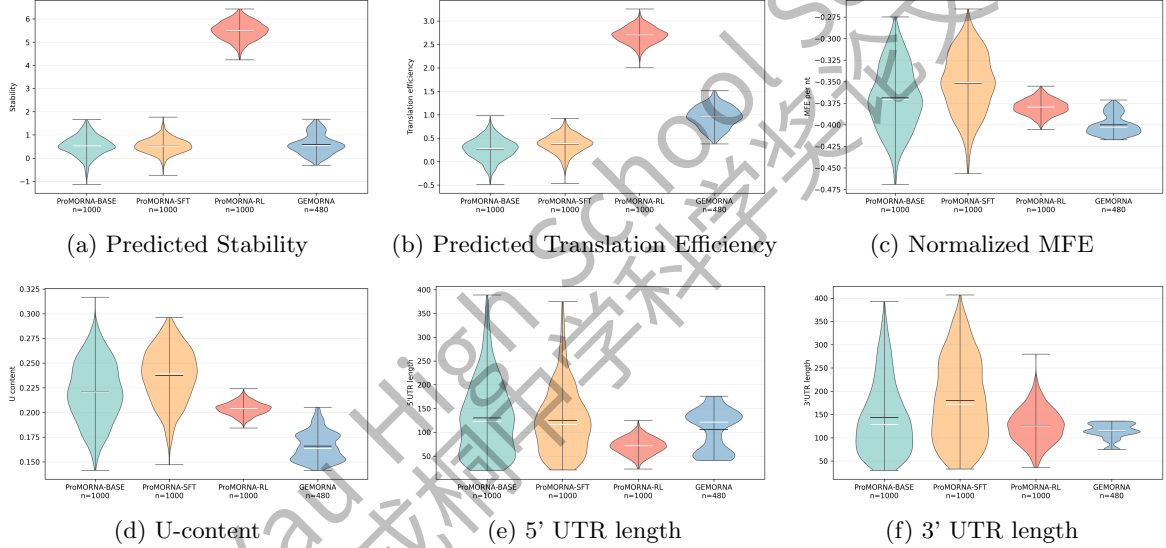


Figure 5: Comparison of three ProMORNA checkpoints and GEMORNA on reward-related metrics.

The comparison results and ablation study are presented in Fig. 5. As shown in Figs. 5a and 5b, ProMORNA scores highest on the two primary functional objectives: predicted stability and translation efficiency. This indicates that MO-GRPO effectively guides the generator toward candidates with improved functional proxy scores. While GEMORNA also outperforms the purely supervised BASE and SFT models, consistent with its strong region-wise mRNA generation design, ProMORNA builds on this by directly optimizing the full-length transcript using multi-objective reinforcement learning, rather than recombining independently generated UTR and CDS elements.

Figures 5c and 5d show that ProMORNA scores slightly higher than GEMORNA scores on normalized full-length MFE and U-content, but lower than BASE and SFT scores. Because mRNA expression relies on complex interactions among regional structure, codon usage, UTR context, innate immune sensing, and degradation, these secondary metrics should be evaluated alongside the primary functional objectives.⁵ ProMORNA’s moderate scores on these subsidiary metrics reflect MO-GRPO’s core approach: prioritizing a balanced, full-length trade-off over maximizing individual proxy metrics independently.

Regarding UTR length distributions (Figs. 5e and 5f), ProMORNA tends to generate shorter 5’ UTRs and longer 3’ UTRs compared to the supervised baselines. Since UTR length effects are highly context-

dependent,¹⁴ this distribution suggests that the length-plausibility reward and functional objectives work together to guide ProMORNA toward an optimal, target-specific UTR length for Fluc2P instead of defaulting to a universally fixed length.



Figure 6: Pareto frontiers of ProMORNA, GEMORNA, BASE, and SFT on predicted stability and translation efficiency.

Figure 6 illustrates the Pareto frontiers for predicted stability and translation efficiency across ProMORNA, BASE, SFT, and GEMORNA using the external evaluator Saluki.²⁸ The upper-right shift of the ProMORNA frontier provides additional computational support for the multi-objective design strategy.

3.2 *In Vitro* Validation

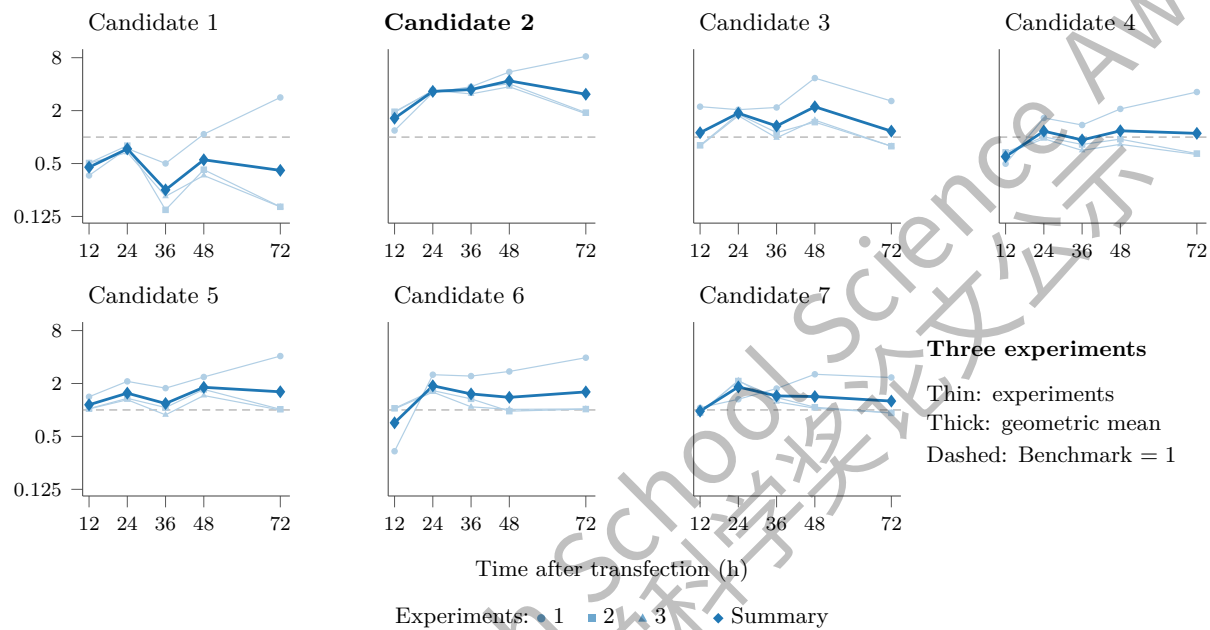
We sample 1,000 sequences from the optimized ProMORNA-RL policy and selected all seven candidates on its Pareto frontier for experimental evaluation. The candidates are tested alongside a conventional Benchmark and GMR-FL-F5¹² in three independent Fluc2P experiments at 12, 24, 36, 48, and 72 h after transfection. The Benchmark, named Benchmark-FL2 in the GEMORNA study,¹² combines BNT162b2-derived UTRs with a CDS from pGL4.11. See Section A.3 for more details regarding assay procedures and data analysis methods.

Figure 7a illustrates the Fluc activity normalized to the benchmark at each time point and Fig. 7b shows the AUC ratios of each candidate to Benchmark. Candidates 2 and 7 showed higher AUCs after seven-candidate Holm correction: Candidate 2 had a geometric-mean ratio of 3.33 (95% CI, 2.67–4.15; $P_{\text{Holm}} = 0.01285$), and Candidate 7 had a ratio of 1.59 (1.43–1.77; $P_{\text{Holm}} = 0.01767$). Candidate 2 is thus selected for closer comparison with the controls.

Figure 8a shows the raw activity trajectories of Candidate 2 and the two controls. Candidate 2 is numerically above both controls at every sampled time point in each experiment, peaking at 24 h and declining by 72 h. Using the same experiments, its AUC ratio relative to GMR-FL-F5 was 1.29 (95% CI, 1.09–1.52; unadjusted $P = 0.02319$; $P_{\text{Holm}} = 0.09294$; Fig. 8b). This estimated 28.7% gain remains exploratory after correction across the seven GMR-FL-F5 comparisons.

(a) Time courses of the seven candidates

Fluc activity / matched Benchmark (fold; log scale)



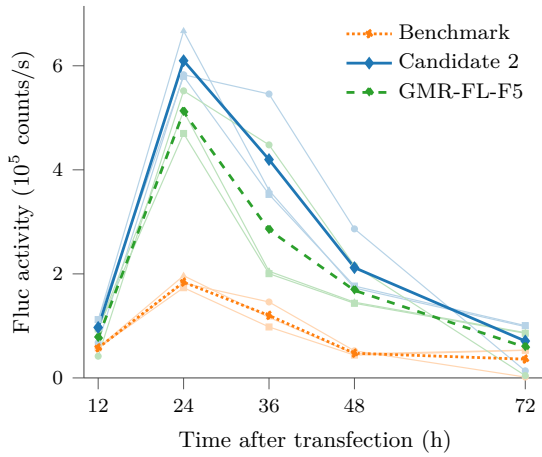
(b) AUC screening against Benchmark

Candidate	AUC ratio	Geometric mean [95% CI]	P (raw)	P_{Holm}
Candidate 1		0.53 [0.29, 0.98]	0.04770	0.23852
Candidate 2		3.33 [2.67, 4.15]	0.00184	0.01285
Candidate 3		1.70 [0.69, 4.16]	0.12559	0.50238
Candidate 4		1.04 [0.43, 2.51]	0.85488	0.85488
Candidate 5		1.44 [0.70, 2.96]	0.15964	0.50238
Candidate 6		1.61 [0.68, 3.79]	0.14077	0.50238
Candidate 7		1.59 [1.43, 1.77]	0.00295	0.01767

Fold relative to Benchmark

Figure 7: Screening ProMORNA candidates against Benchmark in three independent Fluc2P experiments.

(a) Raw reporter activity



(b) Integrated reporter activity

AUC over 12–72 h

Candidate 2 / Benchmark

Geometric mean: 3.33 [2.67, 4.15]

 $P = 0.00184$; $P_{\text{Holm}} = 0.01285$ (B7)

Candidate 2 / GMR-FL-F5

Geometric mean: 1.29 [1.09, 1.52]

 $P = 0.02319$; $P_{\text{Holm}} = 0.09294$ (G7)

Paired AUC ratio (fold)

Experiments: • 1 ■ 2 ▲ 3 ◆ Summary

Figure 8: Candidate 2 compared with Benchmark and GMR-FL-F5 in the same three experiments. Holm-adjusted P values are shown for the corresponding seven-candidate families: B7 for Benchmark and G7 for GMR-FL-F5.

4 Conclusion

In this paper, we propose **ProMORNA**, an end-to-end generative framework for full-length mRNA design. ProMORNA generates complete mRNA transcripts, including the 5' UTR, CDS, and 3' UTR, directly from a target protein sequence. Using a BART-style encoder-decoder architecture, the model conditions every generated region on the full protein context, enabling joint design of coding and non-coding regions without requiring template inputs or updating model weights at inference time.

To align the generative model with multiple biological objectives, we introduce MO-GRPO. Unlike conventional scalar-reward optimization, it computes sequence advantages at the level of individual biological metrics before aggregation. This formulation mitigates the instability caused by objectives with different scales and variances, and provides a flexible mechanism for balancing these objectives.

In silico experiments on Fluc2P demonstrate that ProMORNA improves generated mRNA candidates over supervised baselines and GEMORNA, particularly on the two primary predicted objectives: predicted stability and predicted translation efficiency. Then, the Pareto frontier analysis further shows that MO-GRPO can optimize the model toward generating mRNA candidates with better multi-objective trade-offs. In the *in vitro* screen, Candidates 2 and 7 show significantly higher time-integrated reporter activity than Benchmark after Holm adjustment. These results support that ProMORNA is capable of generating functionally active full-length mRNA candidates.

A primary limitation of the current study is that our prospective biological validation was restricted to seven Pareto-selected candidates for a single reporter protein, Fluc2P. The retrospective AUC analysis and prioritization of Candidate 2 require independent validation, particularly against GMR-FL-F5. Reporter activity does not directly measure mRNA stability or translation efficiency. Extensive *in vitro* experiments with longer transcripts across diverse therapeutic contexts are required to rigorously establish the generalizability and practical applicability of ProMORNA. This future work will provide a stronger foundation for applying ProMORNA to therapeutic mRNA design. Moreover, improving the accuracy and experimental robustness of the stability and translation efficiency predictors will be essential for more reliable reward modeling.

References

- [1] F. P. Polack et al., “Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine,” *New England Journal of Medicine*, vol. 383, no. 27, pp. 2603–2615, 2020. DOI: [10.1056/NEJMoa2034577](https://doi.org/10.1056/NEJMoa2034577)
- [2] L. R. Baden et al., “Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine,” *New England Journal of Medicine*, vol. 384, no. 5, pp. 403–416, 2021. DOI: [10.1056/NEJMoa2035389](https://doi.org/10.1056/NEJMoa2035389)
- [3] N. Pardi, M. J. Hogan, F. W. Porter, and D. Weissman, “mRNA vaccines—a new era in vaccinology,” *Nature Reviews Drug Discovery*, vol. 17, no. 4, pp. 261–279, 2018. DOI: [10.1038/nrd.2017.243](https://doi.org/10.1038/nrd.2017.243)
- [4] S. Qin et al., “mRNA-based therapeutics: Powerful and versatile tools to combat diseases,” *Signal Transduction and Targeted Therapy*, vol. 7, no. 1, 2022, Art. no. 166. DOI: [10.1038/s41392-022-01007-w](https://doi.org/10.1038/s41392-022-01007-w)
- [5] M. Metkar, C. S. Pepin, and M. J. Moore, “Tailor made: The art of therapeutic mRNA design,” *Nature Reviews Drug Discovery*, vol. 23, no. 1, pp. 67–83, 2024. DOI: [10.1038/s41573-023-00827-x](https://doi.org/10.1038/s41573-023-00827-x)
- [6] U. Şahin, K. Karikó, and Ö. Türeci, “mRNA-based therapeutics—developing a new class of drugs,” *Nature Reviews Drug Discovery*, vol. 13, no. 10, pp. 759–780, 2014. DOI: [10.1038/nrd4278](https://doi.org/10.1038/nrd4278)
- [7] K. Leppek et al., “Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics,” *Nature Communications*, vol. 13, no. 1, 2022, Art. no. 1536. DOI: [10.1038/s41467-022-28776-w](https://doi.org/10.1038/s41467-022-28776-w)
- [8] D. M. Mauger et al., “mRNA structure regulates protein expression through changes in functional half-life,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 48, pp. 24 075–24 083, 2019. DOI: [10.1073/pnas.1908052116](https://doi.org/10.1073/pnas.1908052116)
- [9] H. Zhang et al., “Algorithm for optimized mRNA design improves stability and immunogenicity,” *Nature*, vol. 621, no. 7978, pp. 396–403, 2023. DOI: [10.1038/s41586-023-06127-z](https://doi.org/10.1038/s41586-023-06127-z)
- [10] P. M. Sharp and W.-H. Li, “The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications,” *Nucleic Acids Research*, vol. 15, no. 3, pp. 1281–1295, 1987. DOI: [10.1093/nar/15.3.1281](https://doi.org/10.1093/nar/15.3.1281)
- [11] S. Patel et al., *mRNAutilus: Multi-objective-guided discrete generation of mRNA with optimized therapeutic properties*, arXiv preprint, 2026. DOI: [10.48550/arXiv.2605.31296](https://doi.org/10.48550/arXiv.2605.31296) arXiv: [2605.31296v1 \[q-bio.BM\]](https://arxiv.org/abs/2605.31296).
- [12] H. Zhang et al., “Deep generative models design mRNA sequences with enhanced translational capacity and stability,” *Science*, vol. 390, no. 6773, 2025, Art. no. eadr8470. DOI: [10.1126/science.adr8470](https://doi.org/10.1126/science.adr8470)
- [13] S. Li, P. Chauvin, O. Gross, M. Bailey, and S. Jager, *mRNA-GPT: A generative model for full-length mRNA design and optimization*, bioRxiv preprint, 2026. DOI: [10.64898/2026.03.31.715707](https://doi.org/10.64898/2026.03.31.715707) biorxiv: [2026.03.31.715707](https://doi.org/10.1101/2026.03.31.715707).
- [14] A. Medjmedj et al., “Evaluation of synthetic mRNA with selected UTR sequences and alternative poly(A) tail, in vitro and in vivo,” *Molecular Therapy Nucleic Acids*, vol. 36, no. 3, 2025, Art. no. 102648. DOI: [10.1016/j.omtn.2025.102648](https://doi.org/10.1016/j.omtn.2025.102648)
- [15] U. Şahin and Ö. Türeci, “Personalized vaccines for cancer immunotherapy,” *Science*, vol. 359, no. 6382, pp. 1355–1360, 2018. DOI: [10.1126/science.aar7112](https://doi.org/10.1126/science.aar7112)
- [16] E. Blass and P. A. Ott, “Advances in the development of personalized neoantigen-based therapeutic cancer vaccines,” *Nature Reviews Clinical Oncology*, vol. 18, no. 4, pp. 215–229, 2021. DOI: [10.1038/s41571-020-00460-2](https://doi.org/10.1038/s41571-020-00460-2)
- [17] Z. Shao et al., *DeepSeekMath: Pushing the limits of mathematical reasoning in open language models*, arXiv preprint, 2024. DOI: [10.48550/arXiv.2402.03300](https://doi.org/10.48550/arXiv.2402.03300) arXiv: [2402.03300 \[cs.CL\]](https://arxiv.org/abs/2402.03300).

- [18] N. A. O’Leary et al., “Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation,” *Nucleic Acids Research*, vol. 44, no. D1, pp. D733–D745, 2016. DOI: [10.1093/nar/gkv1189](https://doi.org/10.1093/nar/gkv1189)
- [19] M. Lewis et al., “BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2020, pp. 7871–7880. DOI: [10.18653/v1/2020.acl-main.703](https://doi.org/10.18653/v1/2020.acl-main.703)
- [20] S. E. Calvo, D. J. Pagliarini, and V. K. Mootha, “Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 18, pp. 7507–7512, 2009. DOI: [10.1073/pnas.0810916106](https://doi.org/10.1073/pnas.0810916106)
- [21] Y. Xiong, A. Wang, Y. Kang, C. Shen, C.-Y. Hsieh, and T. Hou, “mRNABERT: Advancing mRNA sequence design with a universal language model and comprehensive dataset,” *Nature Communications*, vol. 16, no. 1, 2025, Art. no. 10371. DOI: [10.1038/s41467-025-65340-8](https://doi.org/10.1038/s41467-025-65340-8)
- [22] L. Huang et al., “LinearFold: Linear-time approximate RNA folding by 5’-to-3’ dynamic programming and beam search,” *Bioinformatics*, vol. 35, no. 14, pp. i295–i304, 2019. DOI: [10.1093/bioinformatics/btz375](https://doi.org/10.1093/bioinformatics/btz375)
- [23] C. Barreau, L. Paillard, and H. B. Osborne, “AU-rich elements and associated factors: Are there unifying principles?” *Nucleic Acids Research*, vol. 33, no. 22, pp. 7138–7150, 2005. DOI: [10.1093/nar/gki1012](https://doi.org/10.1093/nar/gki1012)
- [24] K. Karikó, M. Buckstein, H. Ni, and D. Weissman, “Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA,” *Immunity*, vol. 23, no. 2, pp. 165–175, 2005. DOI: [10.1016/j.immuni.2005.06.008](https://doi.org/10.1016/j.immuni.2005.06.008)
- [25] R. Verbeke, M. J. Hogan, K. Loré, and N. Pardi, “Innate immune mechanisms of mRNA vaccines,” *Immunity*, vol. 55, no. 11, pp. 1993–2005, 2022. DOI: [10.1016/j.immuni.2022.10.014](https://doi.org/10.1016/j.immuni.2022.10.014)
- [26] Q. Yu et al., “DAPO: An open-source LLM reinforcement learning system at scale,” in *Advances in Neural Information Processing Systems*, vol. 38, Curran Associates, Inc., 2025, pp. 113 222–113 244. DOI: [10.52202/085713-3775](https://doi.org/10.52202/085713-3775)
- [27] DeepSeek-AI et al., *DeepSeek-V3.2: Pushing the frontier of open large language models*, arXiv preprint, 2025. DOI: [10.48550/arXiv.2512.02556](https://doi.org/10.48550/arXiv.2512.02556) arXiv: [2512.02556](https://arxiv.org/abs/2512.02556) [cs.CL].
- [28] V. Agarwal and D. R. Kelley, “The genetic and biochemical determinants of mRNA degradation rates in mammals,” *Genome Biology*, vol. 23, no. 1, 2022, Art. no. 245. DOI: [10.1186/s13059-022-02811-x](https://doi.org/10.1186/s13059-022-02811-x)
- [29] J. Liu et al., *Muon is scalable for LLM training*, arXiv preprint, 2025. DOI: [10.48550/arXiv.2502.16982](https://doi.org/10.48550/arXiv.2502.16982) arXiv: [2502.16982](https://arxiv.org/abs/2502.16982) [cs.LG].
- [30] K. Jordan et al. “Muon: An optimizer for hidden layers in neural networks,” Accessed: Sep. 15, 2026. [Online]. Available: <https://kellerjordan.github.io/posts/muon/>
- [31] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *International Conference on Learning Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=Bkg6RiCqY7>
- [32] D. Zheng et al., “Predicting the translation efficiency of messenger RNA in mammalian cells,” *Nature Biotechnology*, vol. 44, no. 5, pp. 783–796, 2026. DOI: [10.1038/s41587-025-02712-x](https://doi.org/10.1038/s41587-025-02712-x)
- [33] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2019, pp. 2623–2631. DOI: [10.1145/3292500.3330701](https://doi.org/10.1145/3292500.3330701)

- [34] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison III, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature Methods*, vol. 6, no. 5, pp. 343–345, 2009. DOI: [10.1038/nmeth.1318](https://doi.org/10.1038/nmeth.1318)
- [35] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979. [Online]. Available: <https://www.jstor.org/stable/4615733>
- [36] L. McInnes, J. Healy, and J. Melville, *UMAP: Uniform manifold approximation and projection for dimension reduction*, arXiv preprint, 2018. DOI: [10.48550/arXiv.1802.03426](https://doi.org/10.48550/arXiv.1802.03426) arXiv: [1802.03426](https://arxiv.org/abs/1802.03426) [stat.ML].
- [37] M. Kozak, "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs," *Nucleic Acids Research*, vol. 15, no. 20, pp. 8125–8148, 1987. DOI: [10.1093/nar/15.20.8125](https://doi.org/10.1093/nar/15.20.8125)
- [38] M. Kozak, "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes," *Cell*, vol. 44, no. 2, pp. 283–292, 1986. DOI: [10.1016/0092-8674\(86\)90762-2](https://doi.org/10.1016/0092-8674(86)90762-2)
- [39] G. Kudla, L. Lipinski, F. Caffin, A. Helwak, and M. Zylicz, "High guanine and cytosine content increases mRNA levels in mammalian cells," *PLoS Biology*, vol. 4, no. 6, 2006, Art. no. e180. DOI: [10.1371/journal.pbio.0040180](https://doi.org/10.1371/journal.pbio.0040180)
- [40] T. Tuller and H. Zur, "Multiple roles of the coding sequence 5' end in gene expression regulation," *Nucleic Acids Research*, vol. 43, no. 1, pp. 13–28, 2015. DOI: [10.1093/nar/gku1313](https://doi.org/10.1093/nar/gku1313)
- [41] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990. DOI: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)

Appendix A Experimental Setup

A.1 Model Training

Supervised training details. The BASE model is trained with the scalable Muon variant proposed in Moonlight,²⁹ which augments the original Muon optimizer³⁰ with weight decay and parameter-wise update-scale adjustment. We use learning rate 10^{-3} , weight decay 0.1, equivalent batch size 1,024, and linear warmup over 1,000 steps followed by cosine annealing to 1% of the peak learning rate. We train for 20 epochs and use the final checkpoint as BASE; SFT is initialized from the BASE checkpoint and trained on 37,369 NM-prefixed RefSeq mRNAs without upstream open reading frames (uORFs), with batch size 128 and peak learning rate 5×10^{-5} for 3 epochs.

MO-GRPO training details. ProMORNA-RL is initialized from the SFT checkpoint, and the reference policy π_{ref} is fixed as a frozen copy of the SFT model. Then we employ MO-GRPO as a general protein-conditioned optimization procedure: we randomly sample 4,000 protein prompts from the training set (firefly luciferase target excluded), pulling a batch of 4 protein prompts from this pool at each RL step. For each prompt, the policy generates a group of $G = 32$ complete mRNA rollouts for MO-GRPO updates. We train for 1,000 RL steps using AdamW³¹ with constant learning rate 10^{-5} and linear warmup over 50 steps. The KL penalty coefficient is set to $\beta = 0.02$ and the advantage clipping bound is set to $c = 3.0$. The policy ratio clipping term is naturally omitted since we only perform one optimizer step per batch. The objective weights are set to $(w_{\text{stability}}, w_{\text{TE}}, w_{\text{MFE}}, w_{\text{U}}, w_{\text{UTR}}) = (1.0, 1.0, -0.5, -0.5, 0.5)$, where negative weights indicate minimization objectives. The damping factors are set to $(\delta_{\text{stability}}, \delta_{\text{TE}}, \delta_{\text{MFE}}, \delta_{\text{U}}, \delta_{\text{UTR}}) = (0.2, 0.2, 0.01, 0.01, 0.01)$. Invalid sequences receive the maximal penalty before post-centralization. The stability and translation efficiency predictors are frozen throughout RL training.

Reward predictor training. Predicted stability and predicted translation efficiency serve as the two primary functional objectives in our reward design. Since these quantities cannot be directly computed from sequence alone, we train two supervised proxy predictors to estimate them from mRNA sequences. Each mRNA sequence is first encoded by a frozen mRNABERT model,²¹ and the resulting token-level representations are aggregated into a 768-dimensional sequence representation by global average pooling. This reduces the two prediction tasks to independent mappings from mRNABERT sequence embeddings to scalar property values. The translation-efficiency and stability predictors are then trained with ridge regression on human measurements from the transcriptome-wide atlas³² and the mammalian mRNA degradation compendium,²⁸ respectively.

The ridge regularization strength α is optimized separately for each task using Optuna.³³ For each predictor, we run 100 hyperparameter trials and sample α from the logarithmic range $[10^{-2}, 10^5]$. The hyperparameter search primarily maximizes validation Spearman’s rank correlation coefficient (SRCC). When two trials have validation SRCC values differing by no more than 10^{-4} , we select the one with the lower validation mean absolute error (MAE). To reduce variance in hyperparameter selection, we use repeated 5-fold cross-validation with three repeats.

Prediction task	SRCC	MAE	α
Stability	0.635	0.608	134.27
Translation efficiency	0.744	0.364	82.80

Table 1: Cross-validation performance of the mRNABERT-based proxy predictors.

A.2 *In Silico* Experimental Setup

Target protein. For a case study, we evaluate all methods on firefly luciferase (Fluc2P), following the reporter choice in the GEMORNA study.¹² Only the amino-acid sequence of the target protein is provided to ProMORNA at inference time, and no wild-type mRNA template is used. To avoid potential data leakage, transcripts corresponding to the target protein, if present, are removed from all training datasets and the MO-GRPO prompt pool. It is worth noting that the framework is protein-agnostic and can generate full-length mRNAs for any given protein sequence.

Baselines. We compare ProMORNA-RL with three baselines. BASE denotes the supervised encoder-decoder model trained on the full RefSeq training set. SFT denotes the BASE model further fine-tuned on the filtered NM_ subset. The GEMORNA baseline is constructed by combining the 12 GEMORNA-designed 5' UTRs, the four GEMORNA firefly-luciferase CDSs, and the 10 GEMORNA-designed 3' UTRs from their released supplementary data.¹² Their Cartesian product yields $12 \times 4 \times 10 = 480$ full-length GEMORNA mRNA candidates. Notably, the GEMORNA-derived sequences used for comparison are selected through wet-lab screening in their original experimental pipeline, and thus represent a carefully selected, high-performing set. For a fair comparison, all candidates from different models are evaluated using the same validity checker and the same evaluation metrics.

Candidate generation. For the three ProMORNA variants, sequences are generated using constrained stochastic decoding with temperature $\tau = 1.0$, top- p filtering disabled, and a maximum decoder length of 1,024 tokens. Each sequence is forced to begin with AGG to ensure fair comparison with other methods. We repeat this generation procedure until 1,000 valid candidates are produced. A generated sequence is accepted only if it passes these basic biological filters:

- $|5' \text{ UTR}| \geq 20$ and $|3' \text{ UTR}| \geq 30$;
- the CDS correctly translates to the target protein;
- no uORF is present, i.e. 'AUG' $\notin$ 5' UTR.

For GEMORNA, 480 candidates constructed from the Cartesian product of its released 5' UTR, CDS, and 3' UTR elements.

Evaluation protocol. All reported metric distributions are computed on valid and unique candidates. For each method, we evaluate (i) reward-related metrics, including predicted stability, predicted translation efficiency, normalized MFE, U-content, 5' UTR length, and 3' UTR length; (ii) diagnostic sequence-property metrics, including GC content, the normalized MFE of the 5' leader region, the normalized MFE of the remaining mRNA body, and external GEMORNA UTR scores, 5' UTR Max Identity, and 3' UTR Max Identity. For RL learning curves, we report the average metric value over all valid generated candidates in every training step.

A.3 *In Vitro* Validation

Plasmid Construction. The plasmids used in this study are constructed via restriction enzyme cloning and Gibson assembly.³⁴

Reaction Assembly. In vitro transcription (IVT) reactions are assembled as described in Table 2. After brief centrifugation, the reaction is incubated at 37°C for 1.5–2 hours.

DNA Digestion. DNase I (10 U) is added to each reaction, followed by incubation at 37°C for 30 minutes.

Component	Volume (μL)
RNase-free H_2O	Up to 20
10 $\times$ Transcription Buffer	2
ATP / CTP / GTP / N1-Me-pUTP (100 mM each)	2
CAP GAG 3011 (100 mM)	2
Template DNA	1 μg
T7 RNA Polymerase Mix	2
Total Volume	20

Table 2: Composition of the *in vitro* transcription reaction.

RNA Purification. The RNA product is purified via lithium chloride precipitation. An equal volume of lithium chloride precipitation solution (5 M) is added to the reaction product and mixed thoroughly. The mixture is precipitated at -20°C for 30 minutes and then centrifuged at 12,000 RPM at 4°C for 15 minutes. The supernatant is discarded. The RNA pellet is washed with 200 μL of 70% ice-cold ethanol and centrifuged at 12,000 RPM at 4°C for 15 minutes. The supernatant is discarded and the washing step is repeated 2–3 times. The pellet is air-dried with the lid open for 5–10 minutes. Once completely dry, the RNA pellet is dissolved in 100–200 μL of RNase-free ddH₂O or an appropriate buffer. The RNA concentration is measured, aliquoted, and stored at -80°C .

mRNA Transfection. HEK293T cells are cultured in DMEM (Gibco, 6126249) supplemented with 10% FBS (Vazyme, 7E951A5), 100 units/mL penicillin, and 100 $\mu\text{g}/\text{mL}$ streptomycin. One day prior to transfection, cells are seeded into 96-well plates at a density of 20,000 cells per well. Transfection is performed using Lipofectamine[®] RNAiMAX Reagent (Invitrogen, 13778-030) following the manufacturer’s protocol.

Cell Lysis. To measure firefly luciferase activity, 100 μL of reporter cell lysis buffer is added to each well to ensure complete cell lysis. After lysis, the lysate is centrifuged at 10,000–15,000 $\times g$ for 3–5 minutes, and the supernatant is collected for subsequent measurement.

Detection Reagent Preparation. Firefly luciferase detection reagent is thawed and equilibrated to room temperature. The chemiluminescence reader (or a multifunctional microplate reader with chemiluminescence detection capability) is operated according to the manufacturer’s instructions, with appropriate interval and measurement times set based on experimental requirements.

Luminescence Measurement. For each sample, 80 μL of supernatant is transferred to the assay plate. An equal volume of reporter cell lysis buffer serves as a blank control. Subsequently, 100 μL of firefly luciferase detection reagent is added to each well, mixed thoroughly by pipetting, and the luminescence signal is immediately measured.

Data Processing and Statistical Analysis. We average first-cycle readings from technical wells within each construct, experiment, and time point (four wells in experiment 1 and three in experiments 2 and 3). In this retrospective analysis, the principal activity measure is the trapezoidal AUC of the raw counts/s curve over 12–72 h, calculated before normalization. All five time points are retained, with no additional background subtraction. Candidate/control AUC ratios are paired within experiment. We report geometric means and pointwise 95% confidence intervals obtained from the three log ratios using the t distribution ($n = 3, df = 2$). These intervals assume approximately normal log ratios and are not adjusted for multiplicity or candidate selection. Two-sided t tests assess whether the mean log ratio differs from zero. Holm correction³⁵ controls family-wise error separately within the seven candidate–Benchmark

comparisons (B7; screening) and the seven candidate-GMR-FL-F5 comparisons (G7; exploratory). Time courses are descriptive.

Appendix B Reinforcement Learning Dynamics

In this section, we investigate how MO-GRPO alters the generator during reinforcement learning. Figure 9 presents the learning curves for the five reward-related metrics over 1,000 RL steps, averaged across valid rollouts from protein prompts sampled during general MO-GRPO training. The light traces represent step-level variance, while the bold curves depict the smoothed trends.

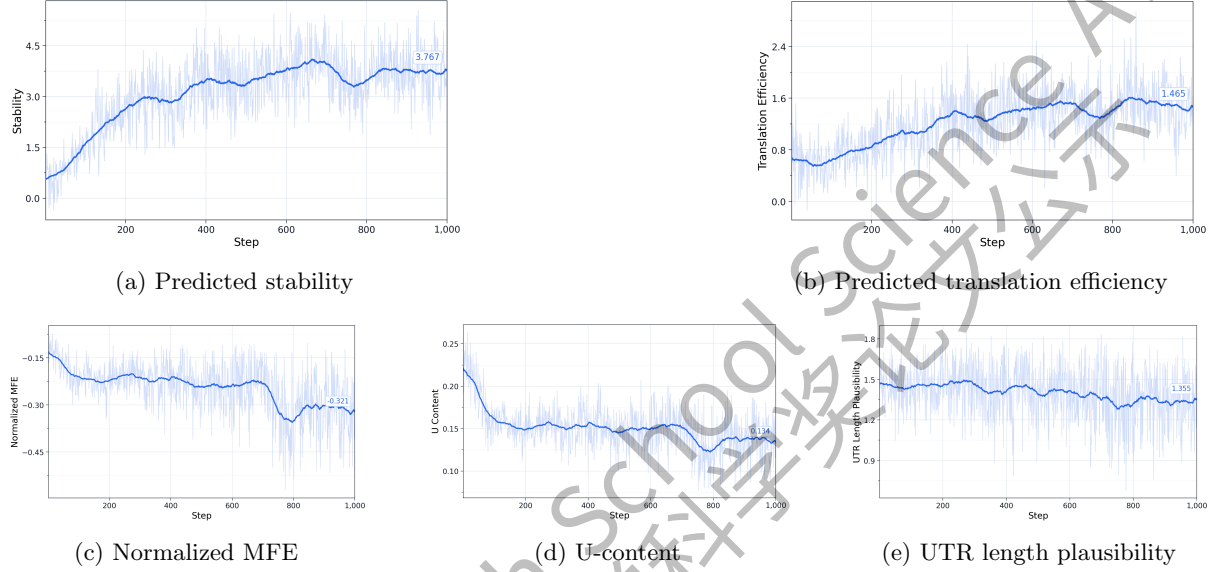


Figure 9: The learning curves of MO-GRPO on five reward metrics.

MO-GRPO successfully optimizes the two primary functional proxy objectives. Notably, predicted stability increases rapidly and steadily throughout most of the training. Predicted translation efficiency experiences an initial decline due to early trade-offs but subsequently recovers, rising steadily before gradually approaching a plateau. Concurrently, normalized MFE decreases, indicating the formation of more stable full-transcript secondary structures, and U-content drops sharply, aligning consistently with the reward design. The UTR length plausibility score slightly lowered, which matches our expectations for it as a regularization term. Overall, these trajectories demonstrate that MO-GRPO can effectively improve multiple rewards simultaneously, despite the noise introduced by stochastic full-length rollouts.

To understand the evolution of sequence composition during training, Fig. 10 visualizes the generated mRNAs using a UMAP projection³⁶ of 1,024-dimensional 5-mer frequency features. The projection reveals a striking bifurcation in the latent space: early RL samples cluster tightly around the original SFT baseline on the right, whereas progressing training steps drive the candidates to migrate into a completely separate, vastly expanded region on the left. This distinct spatial shift underscores that MO-GRPO does not simply exploit the supervised model’s local vicinity, but actively and extensively explores novel sequence-composition paradigms. While this compositional shift does not by itself establish biological quality, pairing this structural migration with the previously discussed learning curves demonstrates that on-policy optimization effectively navigates away from the SFT distribution to uncover highly distinct sequences with superior reward profiles.

The Kozak sequence is a well-established cis-regulatory element in eukaryotic mRNA that strongly influences translation initiation efficiency.^{37,38} According to the canonical Kozak consensus sequence, the nucleotide at position -3 relative to the start codon should preferably be a purine, with adenine (A)

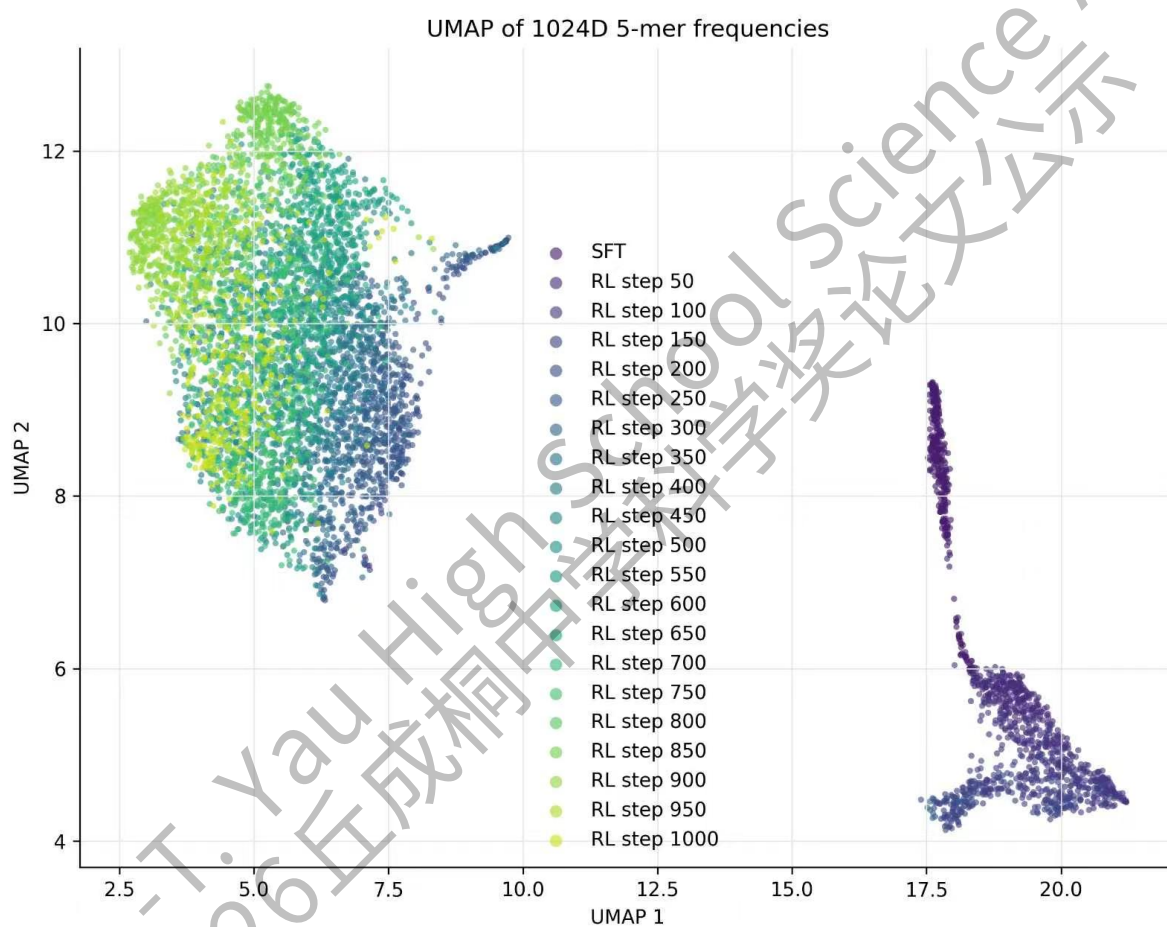


Figure 10: UMAP visualization of 5-mer features from mRNAs generated during MO-GRPO training.

being the most optimal, while the -1 position is ideally a cytosine (C). We investigated whether our RL-optimized models autonomously learned to respect these biological constraints, despite receiving no explicit Kozak-related inductive bias during training. The results are illustrated in Fig. 11.

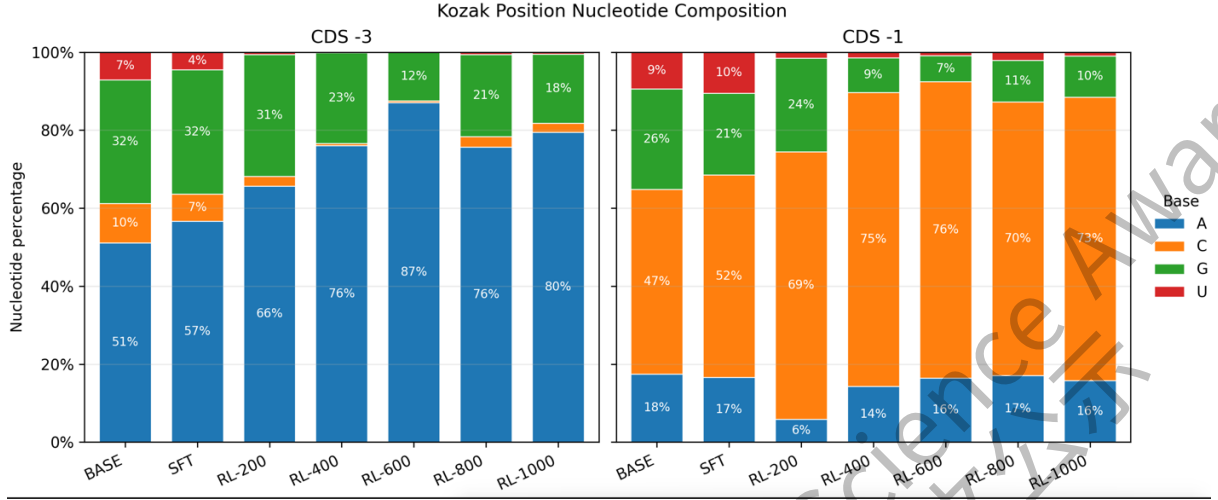


Figure 11: RL shifts nucleotide compositions at Kozak positions.

At the -3 position, the model exhibits a clear progressive enrichment for purines (A/G) across training stages, with a striking shift toward the optimal adenine. The BASE model demonstrates a moderate baseline preference ($A+G = 83\%$), which is further reinforced by supervised fine-tuning (SFT: 89%). Strikingly, RL training not only drives the combined purine content near $97\%-99\%$ (from RL-400 onward), but it also autonomously discovers the superiority of adenine: the frequency of 'A' alone surges from 57% in SFT to a peak of 87% at RL-600, while 'G' concurrently drops from 32% to 12% .

Similarly, at the -1 position, the trend toward Kozak compliance is highly pronounced. The BASE model places a cytosine at this position only 47% of the time, and SFT improves this to 52% . However, RL training yields a dramatic shift: C content jumps to 69% at RL-200, peaks at 76% at RL-600, and remains robustly high across all subsequent checkpoints. Together, these results provide evidence of biological interpretability within our RL pipeline.

Appendix C MO-GRPO Ablation

We replace MO-GRPO in our model training with a scalarized GRPO-style ablation, in which the reward is calculated by the weighted sum of the five metrics defined in Section 2.2, using exactly the same metric weights for MO-GRPO reported in Section A.1, and the advantage is given by standardized rewards. Invalid sequences are assigned a reward of zero. Other hyperparameters remain unchanged. The comparison of the two learning curves is shown in Fig. 12.

Because metric weights aren't divided by their group standard deviations in the scalarized GRPO, predicted stability and translation efficiency dominated the reward calculation at early stages of RL, resulting in severe reward hacking (Figs. 12a and 12b). This caused the policy to diverge largely from the reference SFT policy, and finally training collapsed because of the exploding KL divergence penalty (Fig. 12c). Due to reward hacking, only the two mRNABERT predictions improved, while all other metrics deteriorated. (Figs. 12d to 12f). On the other hand, MO-GRPO training was very stable, benefiting from its metric-level advantage calculation, dynamic damping, and extreme-value clipping methods.

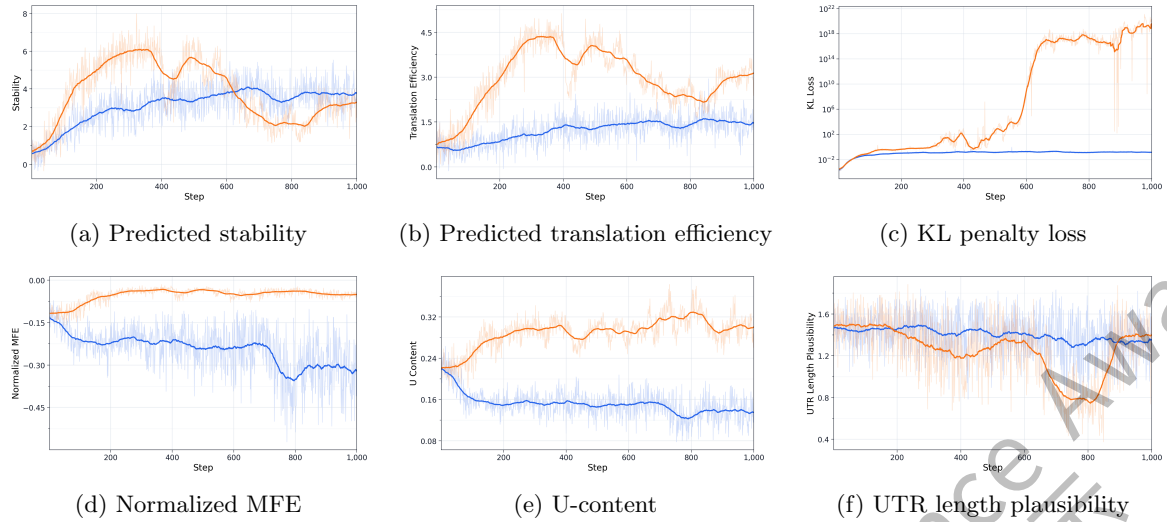


Figure 12: MO-GRPO versus scalarized GRPO on ProMORNA post-training. The orange curves are from scalarized GRPO, while the blue curves are exactly the MO-GRPO curves reported in Section B.

Appendix D Additional mRNA Property Analysis

To better understand the sequence-level mechanics driving these results, Fig. 13 details additional sequence properties. Although not primary reward objectives, these metrics provide valuable mechanistic insights into how ProMORNA shapes nucleotide and codon usage and alters regional mRNA secondary structure.

As illustrated in Fig. 13a, ProMORNA-RL produces a moderate increase in median GC content relative to both the BASE and SFT models, accompanied by a dramatic reduction in variance. While higher GC content is linked to increased mRNA abundance in mammalian cells,³⁹ excessive GC enrichment can deleteriously impact local secondary structure, perturb optimal codon context, and hinder downstream sequence manufacturability. ProMORNA’s moderate increase aligns with a balanced optimization strategy, avoiding the extremes of a single-proxy focus.

Figures 13b and 13c map region-specific mRNA secondary structure by assessing normalized MFE of different regions. Biologically, a higher MFE in the 5’ leader region (the 5’ UTR and first 10 CDS nucleotides) indicates looser secondary structure, which is known to facilitate efficient ribosome scanning.^{8,40} Conversely, tighter secondary structures in the remainder of the mRNA body (the rest of the CDS and the 3’ UTR) broadly enhance transcript stability and protein output, corresponding to a lower MFE.⁷ Compared to the BASE and SFT models, ProMORNA-RL closely aligns with these biological design principles: it shifts the 5’ leader MFE upward to minimize translation initiation barriers, while simultaneously driving the body MFE downward to stabilize the rest of the transcript. Although GEMORNA achieves more extreme MFE values in both regions—likely due to its reliance on a heavily curated sequence pool—ProMORNA-RL autonomously learns to optimize these regional folding topologies while retaining structural diversity.

Figures 13d and 13e evaluate the generated sequences using external GEMORNA UTR scoring models, which estimate UTR quality based on predictors trained for high Mean Ribosome Load in 5’ UTRs and high stability in 3’ UTRs.¹² While GEMORNA exhibits concentrated, high-scoring distributions due to sampling UTRs from its own pre-optimized pool, ProMORNA-RL achieves a highly competitive 5’ UTR score distribution. Notably, it demonstrates significant improvement and reduced variance relative to the BASE and SFT models, all while maintaining broader sequence diversity.

We also assessed the sequence similarities between generated and natural UTRs using BLAST.⁴¹ A low max identity score (MIS) indicates minimal homology between the AI-generated UTRs and naturally occurring UTRs in the reference database. As shown in Figs. 13f and 13g, the MIS distributions for the 5’

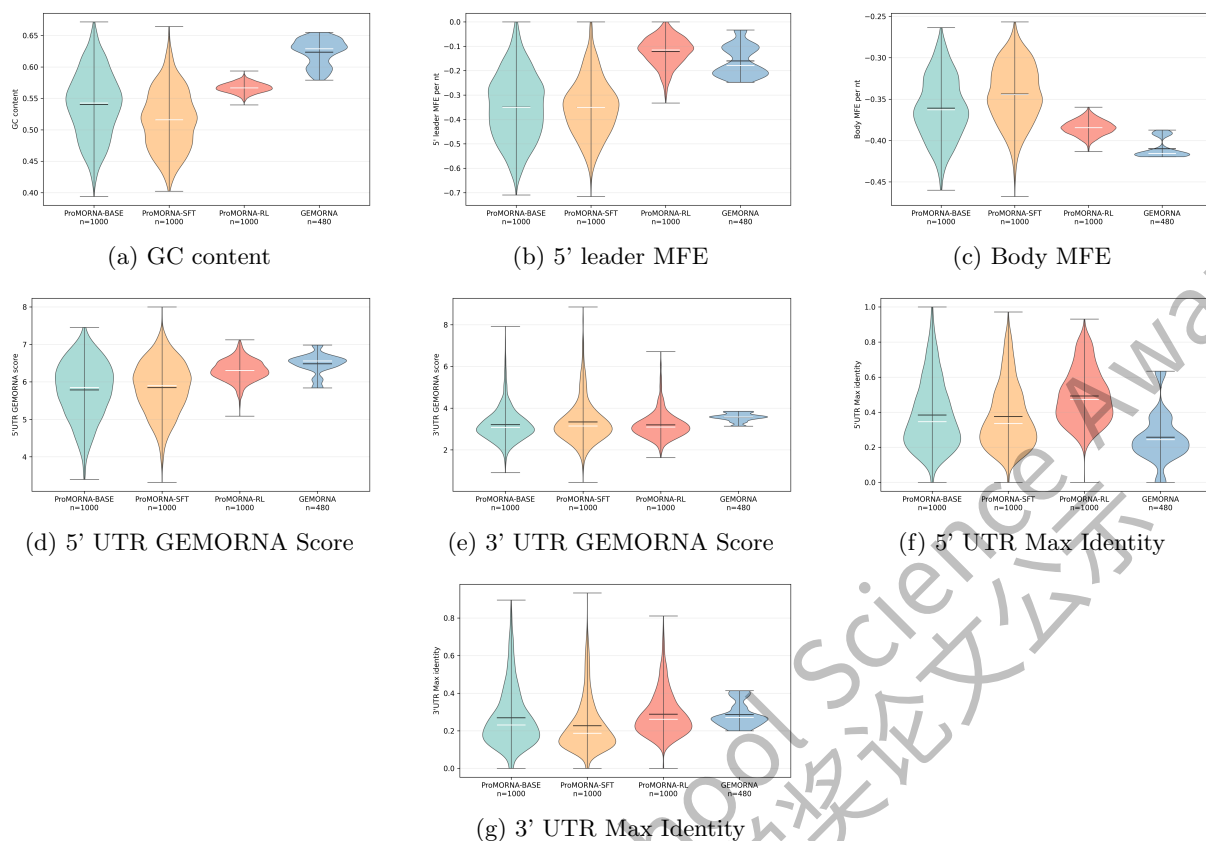


Figure 13: Additional sequence-property analysis of valid, unique firefly-luciferase mRNA candidates. The GEMORNA UTR scores are computed using the external UTR scoring models from GEMORNA.¹²

and 3' UTRs reveal that the ProMORNA models generate sequences with low to moderate homology to known natural transcripts. This confirms that the pipeline performs genuine *de novo* generation rather than merely memorizing sequences from reference databases.

Furthermore, ProMORNA-RL's two regulatory regions display distinct sequence-space dynamics. For the 5' UTR (Fig. 13f), ProMORNA-RL demonstrates a moderate upward shift in median MIS compared to the BASE and SFT models; for the 3' UTR (Fig. 13g), it maintains a persistently low sequence similarity. Taken together, these regional differences may highlight a sophisticated, region-specific optimization strategy within ProMORNA-RL: it autonomously converges toward natural-like motifs where translational initiation constraints are strict (5' UTR), while broadly exploring entirely novel, artificial sequence spaces where structural flexibility permits (3' UTR).