Institute of SCIENCE TOKYO

Computational Approaches to RNA Structure and Function
Jul 5 - 17, 2026 @ Benasque, Spain

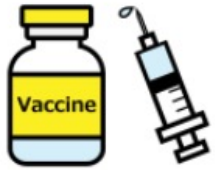
mRNA design with deep generative models

Kengo Sato, Chuankai Dai

School of Life Science and Technology, Institute of Science Tokyo

mRNA sequence design

- **Applications of mRNA are attracting attention.**
  - Great expectations for medical applications due to the success of mRNA vaccines for COVID-19
  - Production of useful biomaterials by bacteria
    - New functions unique to biotechnology, reduction of environmental stress
    - Control and optimization of mRNA transcription and translation for metabolic pathway optimization



Biopharmaceuticals

Bioplastics

Biofuels

Alternative Proteins

High-performance fibers

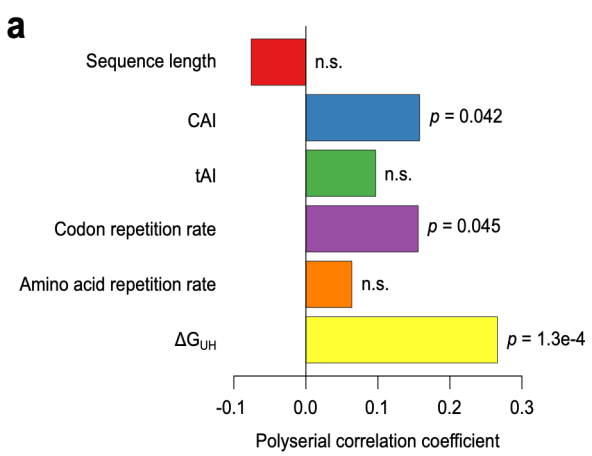
Cosmetic materials

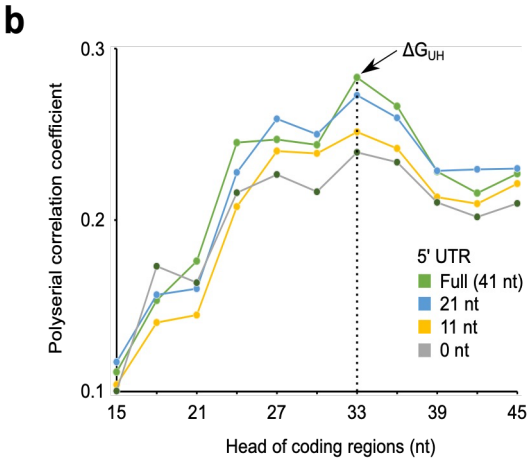
mRNA sequence design

- **Applications of mRNA are attracting attention.**
  - Great expectations for medical applications due to the success of mRNA vaccines for COVID-19
  - Production of useful biomaterials by bacteria
    - New functions unique to biotechnology, reduction of environmental stress
    - Control and optimization of mRNA transcription and translation for metabolic pathway optimization

- **mRNA transcription, translation efficiency, and stability** are affected by codon adaptation as well as primary sequences and secondary structures of promoters, UTRs, terminators, etc.

- Existing mRNA sequence design techniques based on codon optimization are insufficiently optimized.
  - Not fully utilizing the high degree of design flexibility



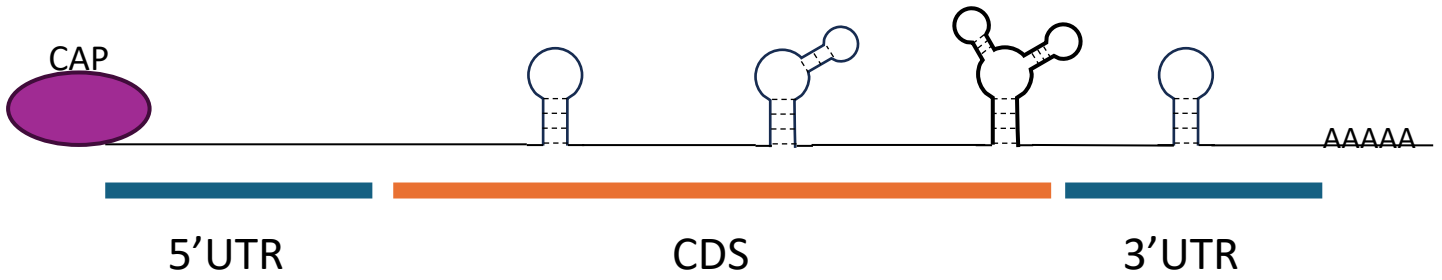


[Saito et al., 2019]

mRNA sequence design with deep generative models

Institute of SCIENCE TOKYO

- We reasonably design highly efficient and high-performance mRNA using deep generative models such as **diffusion models**.

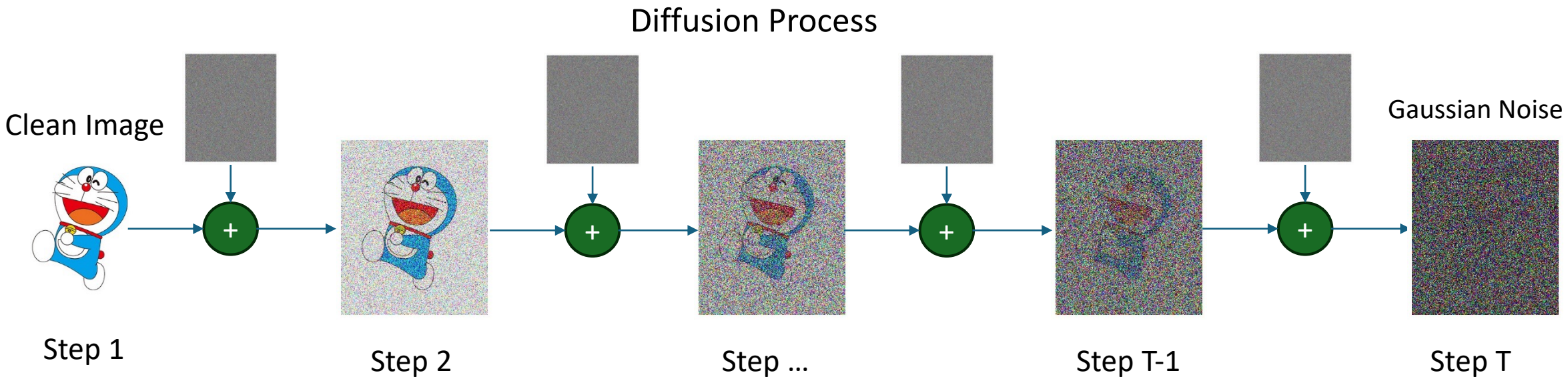


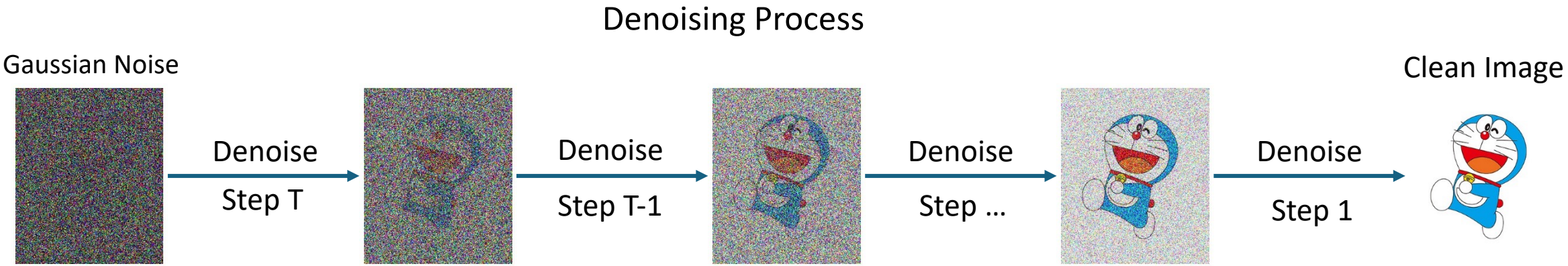
- Learning mRNA features from large-scale data enables designing sequences that maximize protein production with:
  - Weak secondary structures around the start codons.
  - High stability in the CDS and 3′ UTR regions.
  - An appropriate translation speed in the host.

Maximize the production of the target protein

Diffusion models

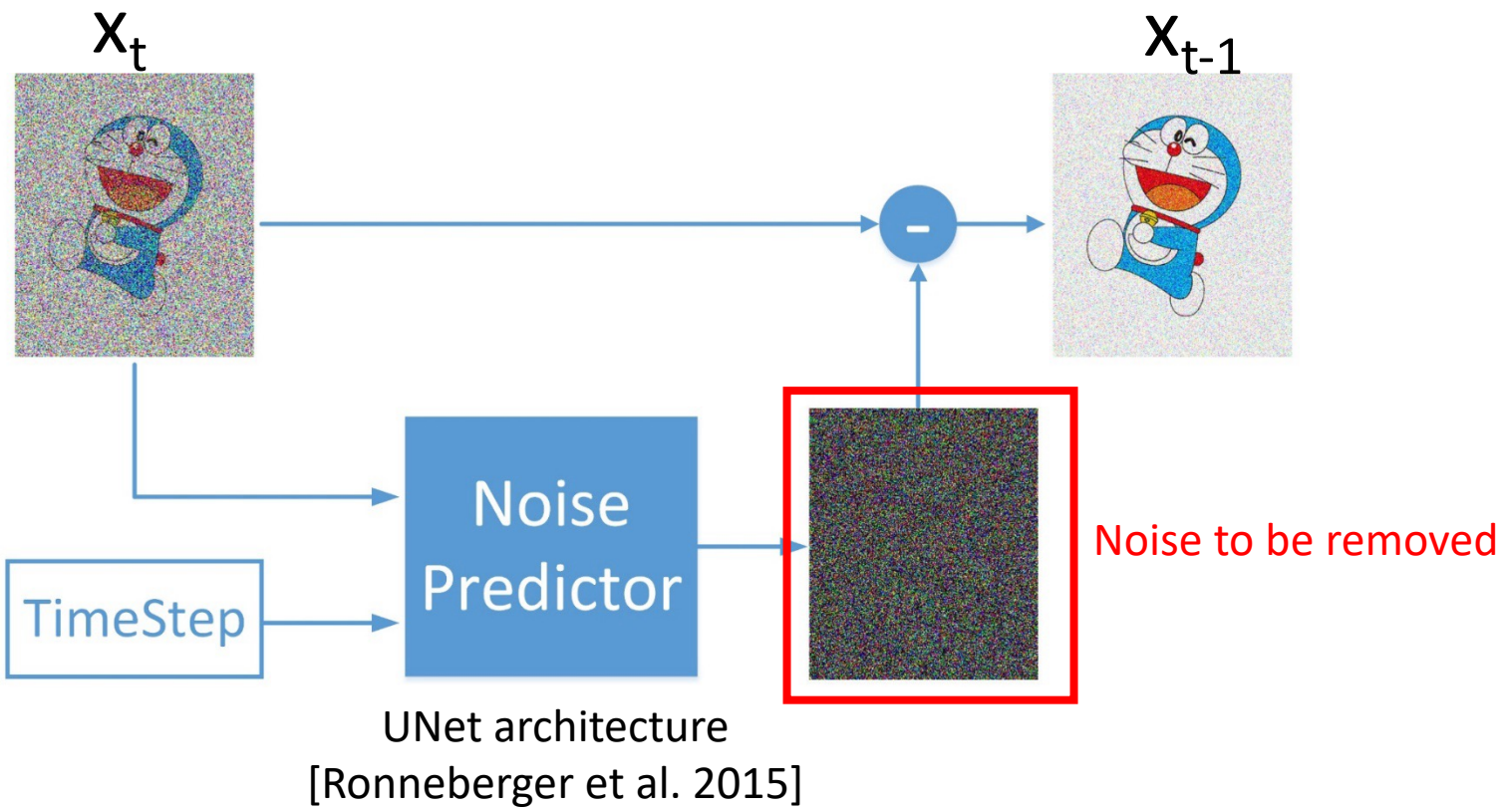
Institute of SCIENCE TOKYO





Diffusion models

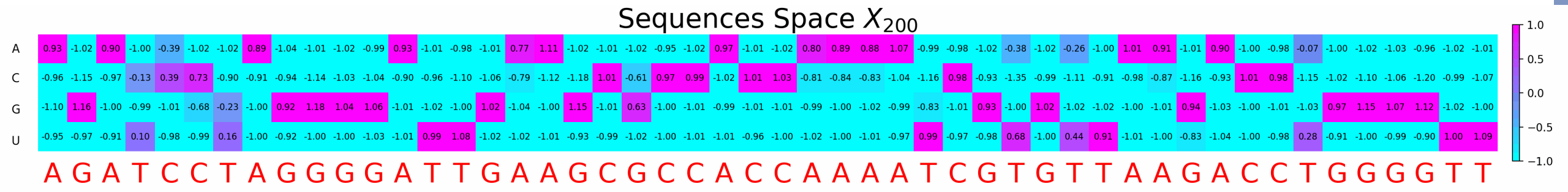
- The deep neural network predicts the noise to be removed at each time step.



UNet architecture
[Ronneberger et al. 2015]

Diffusion models

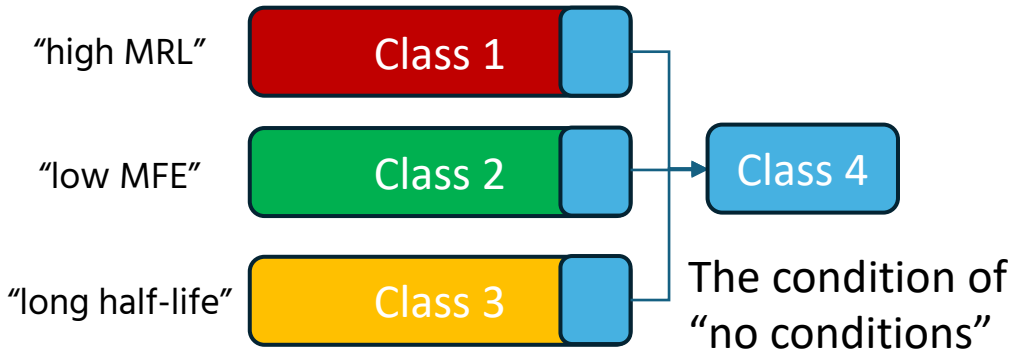
- Apply to RNA sequences



Guiding the model toward the target condition

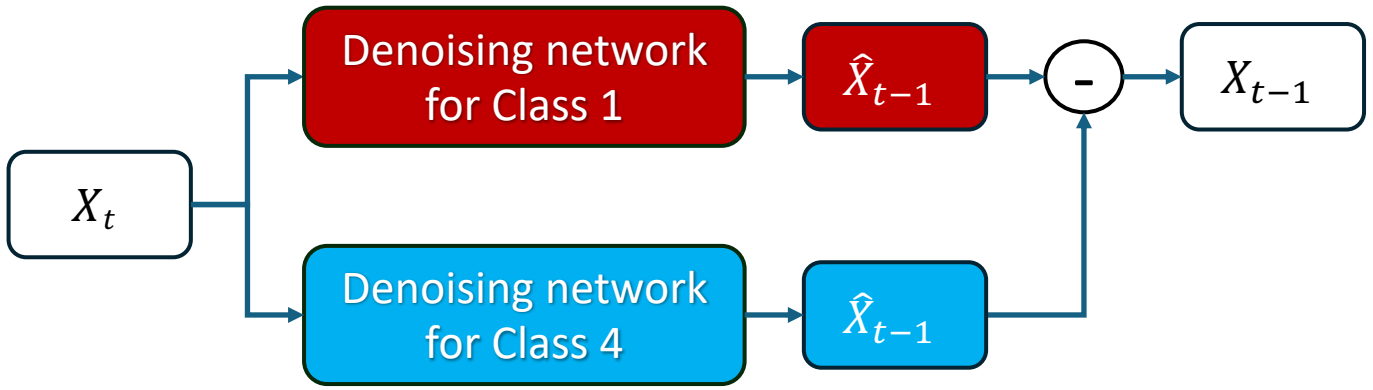
- **Classifier-free guidance** [1] generates RNA sequences that match target conditions such as “high expression”, “low energy”, or “long stability”.

Training



Train the denoising networks for each class

Conditional generation



Move predictions **toward the target condition**, and **away from the unconditional direction**

[1] Nichol et al. "Improved denoising diffusion probabilistic models." *International conference on machine learning*. PMLR, 2021.

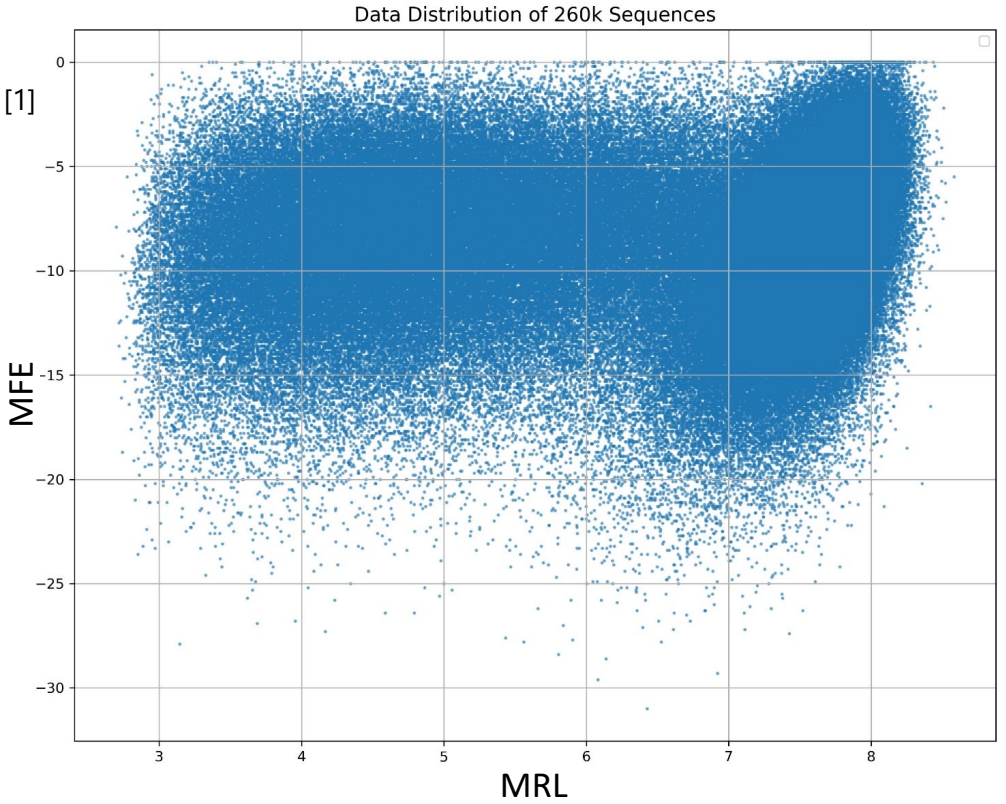
Controlling RNA sequence generation by two target conditions

- We trained a diffusion model to generate 5′ UTR RNA sequences conditioned on two target conditions, **translation efficiency (Mean Ribosome Load; MRL)** and **minimum free energy (MFE)**, and evaluated whether the generated sequences reflect the specified target conditions.

- Dataset:
  - Randomly-generated 260k eGFP 5′ UTR sequences (GSM3130435) [1]
  - Length: 50bp
  - MRL predicted by UTR-LM [2]
  - MFE predicted by RNAfold [3]
- Labels: combinations of two conditions (3x3=9 groups)

Label	MRL
low	4
medium	6
high	8

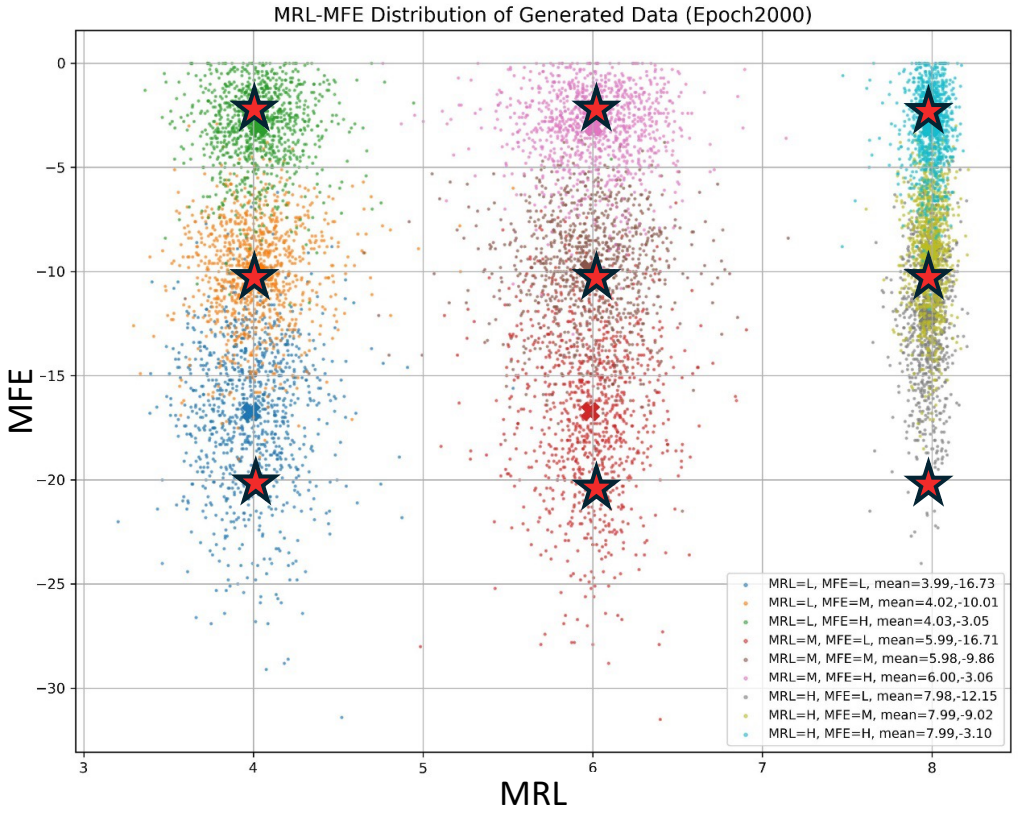
Label	MFE
low	-20
medium	-10
high	-2



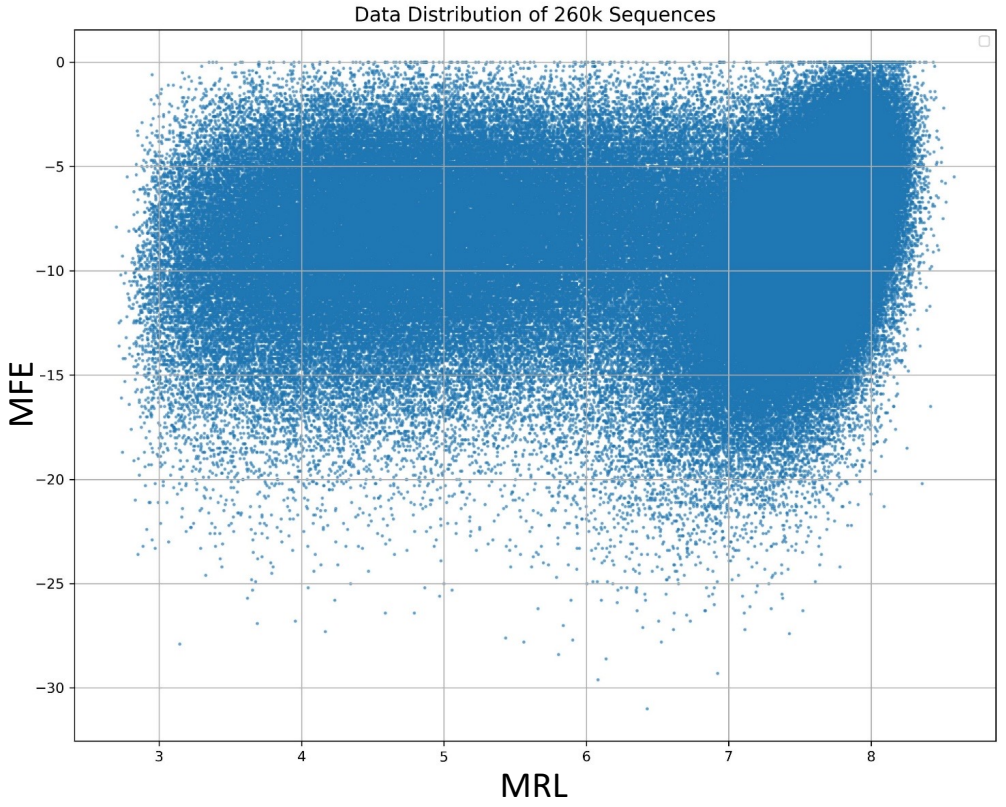
[1] Sample, P. J. *et al.* Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* **37**, 803–809 (2019).
[2] Chu, Y. *et al.* A 5′ UTR language model for decoding untranslated regions of mRNA and function predictions. *Nat. Mach. Intell.* **6**, 449–460 (2024).
[3] Lorenz, R. *et al.* "ViennaRNA Package 2.0." *Algorithms Mol. Biol.* **6**, 26 (2011).

Controlling RNA sequence generation by two target conditions

- We trained a diffusion model to generate 5′ UTR RNA sequences conditioned on two target conditions, **translation efficiency (Mean Ribosome Load; MRL)** and **minimum free energy (MFE)**, and evaluated whether the generated sequences reflect the specified target conditions.

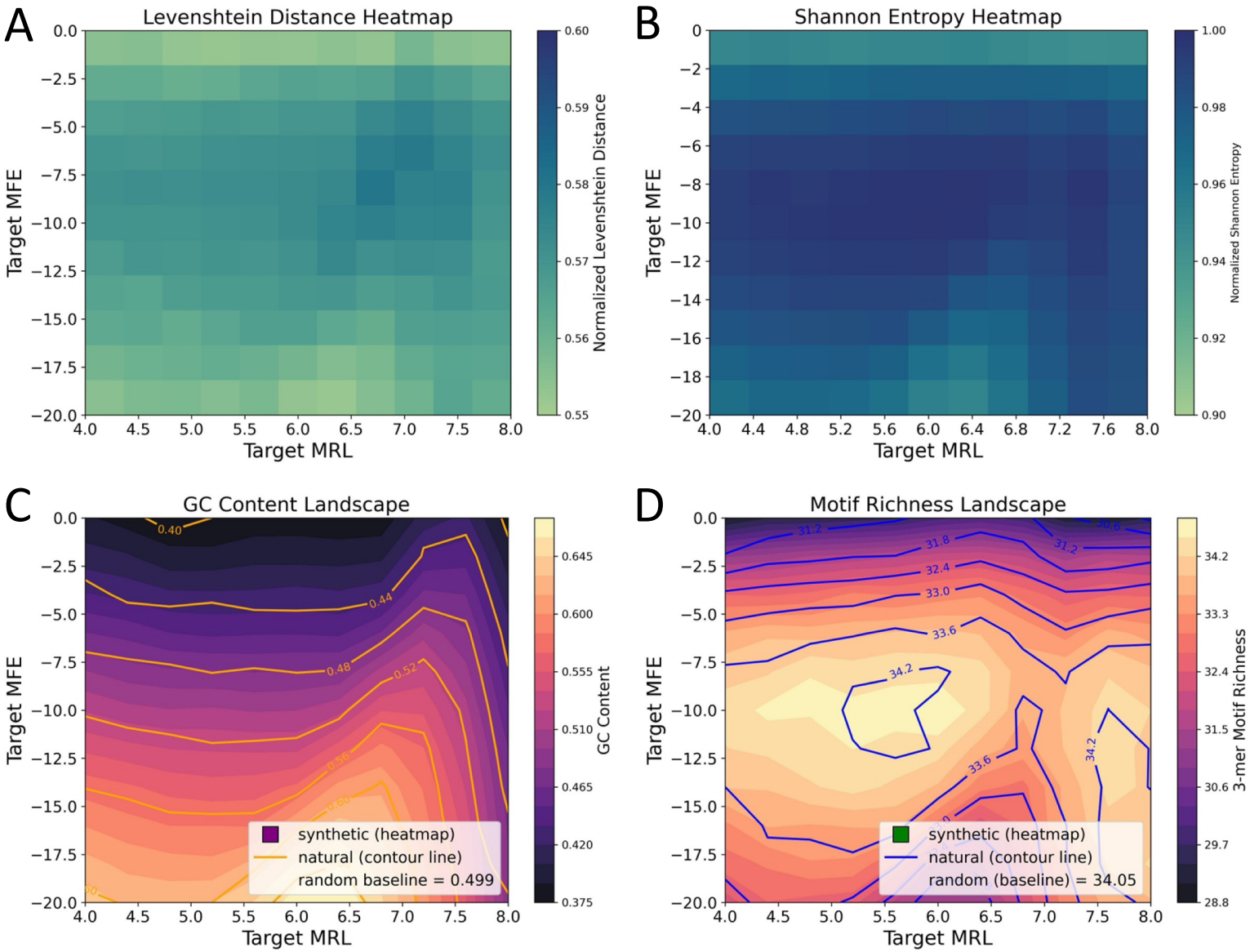


Distribution of generated data (1k samples)



Distribution of training data

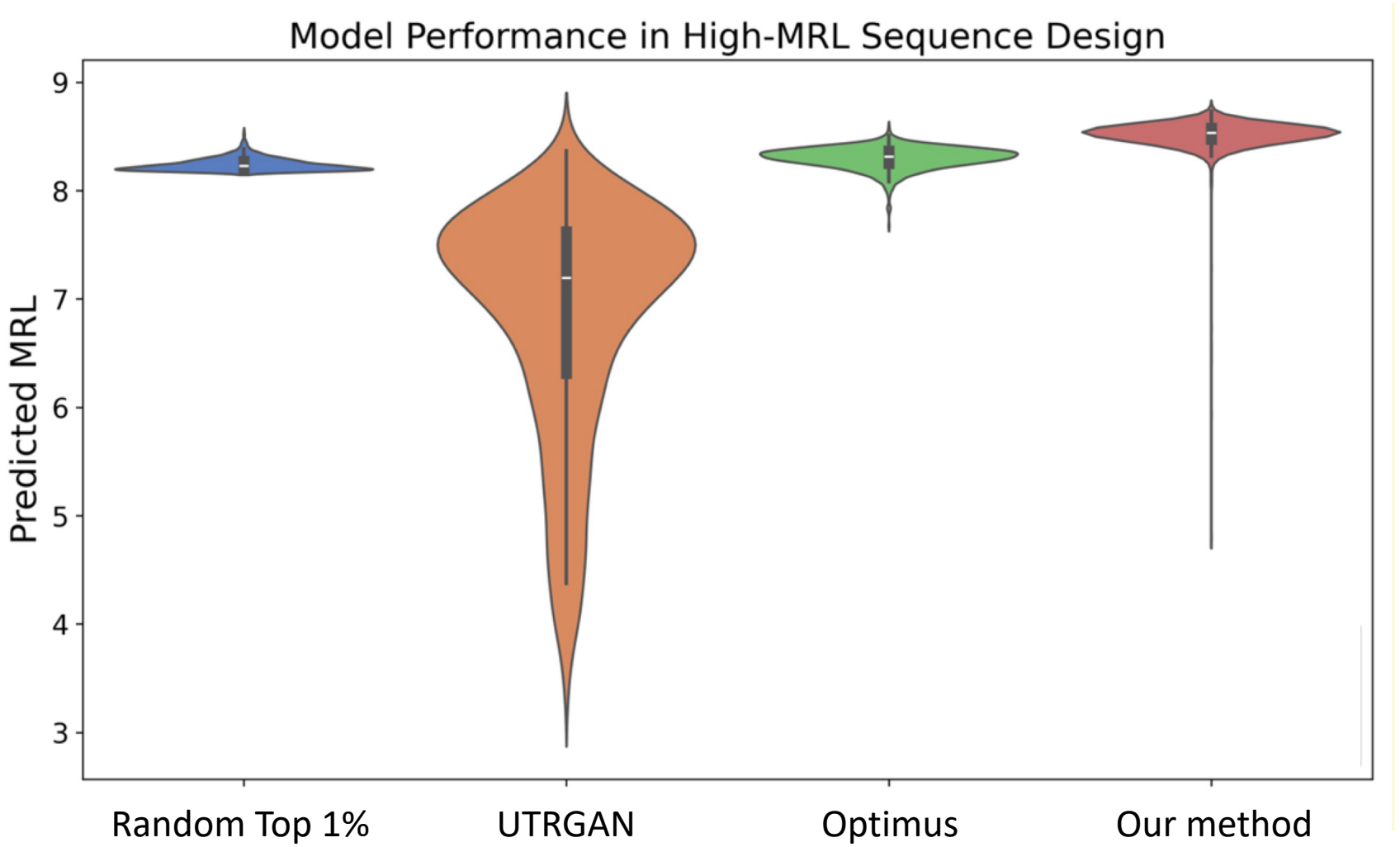
Quality of generated sequences



- Panel A: The generated sequences are not mere copies of the training data.
- Panel B: They are highly diverse.
- Panels C and D: Their distributions closely resemble the training data.

Our method outperforms other methods

- We compare our method with UTRGAN [Barazandeh et al. 2025] and Optimus 5-Prime [Sample et al. 2019].



Localized conditional generation via Repaint technique[1]

Institute of SCIENCE TOKYO

Masked input | Denoising process | Completed images

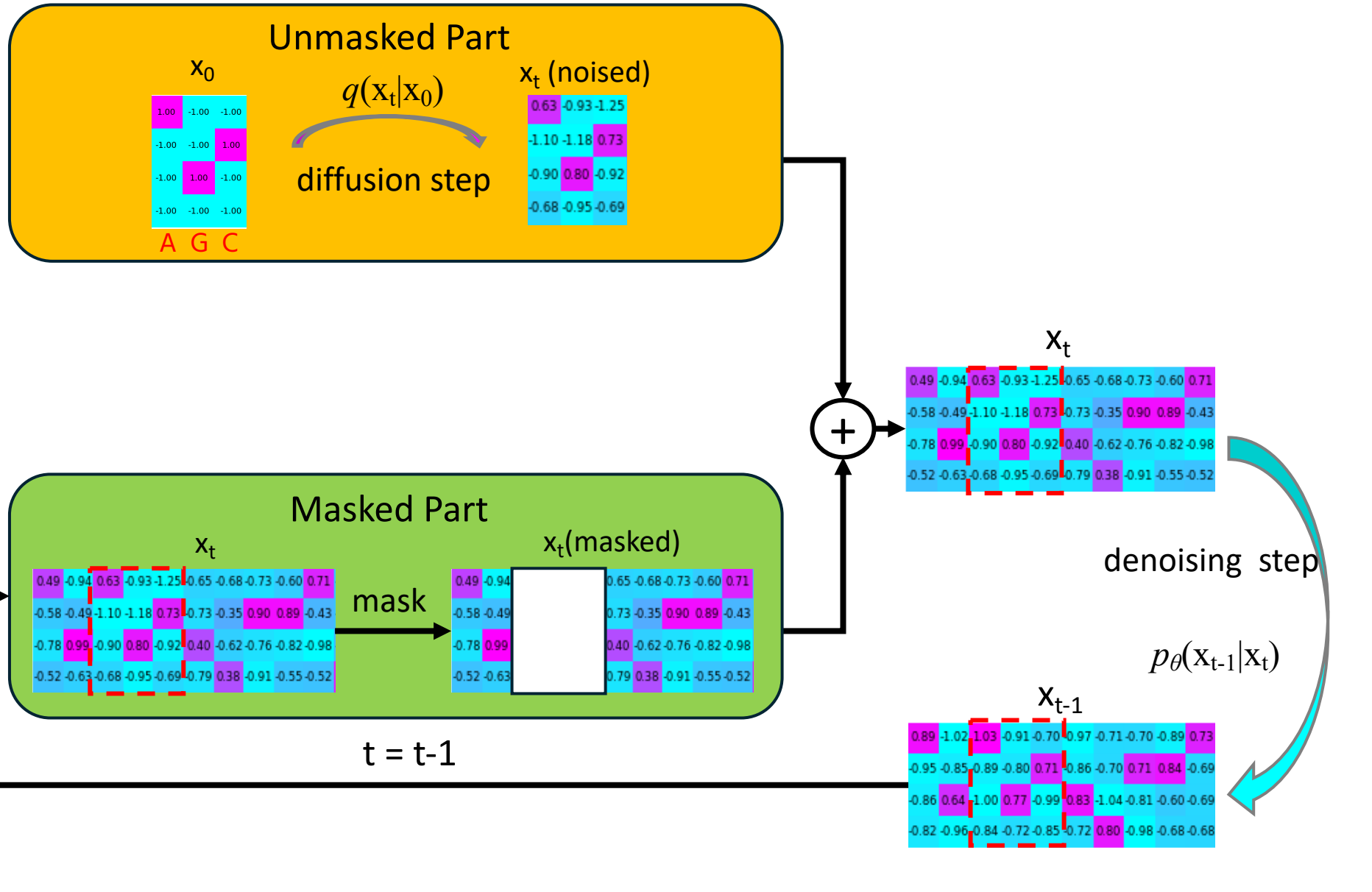
[1]: Lugmayr, Andreas, et al. "Repaint: Inpainting using denoising diffusion probabilistic models." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

RNA design with fixed nucleotides (Repaint extension)

Institute of SCIENCE TOKYO

Unmasked part
fixed nucleotides
(given condition)

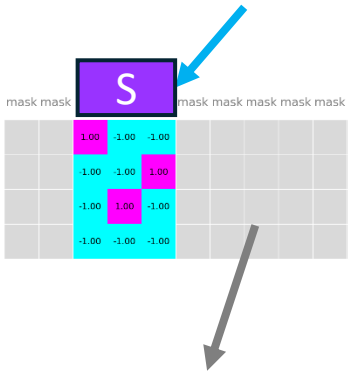
Masked part
are generated from
Gaussian noise



RNA design with fixed amino acids (Repaint extension)

Institute of SCIENCE TOKYO

Unmasked part
fixed amino acids
(given condition)



Masked part
are generated from
Gaussian noise

Unmasked Part

x_0

$q(x_t|x_0)$

diffusion step

x_t (noised)

S

UCU
UCC
UCA
UCG
AGU
AGC

select the most probable codon

Masked Part

x_t

mask

x_t(masked)

x_t

denoising step

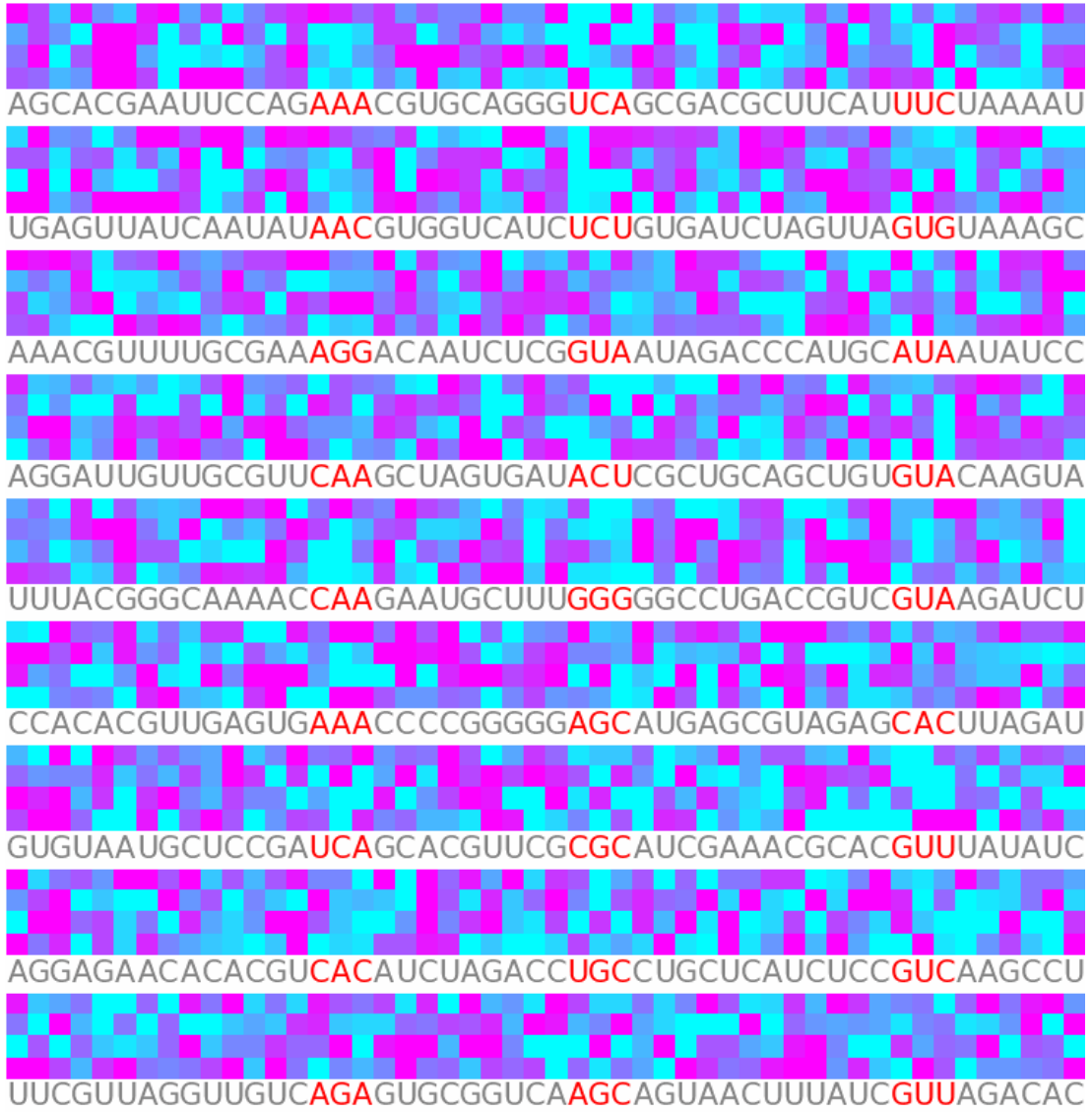
$p_\theta(x_{t-1}|x_t)$

x_{t-1}

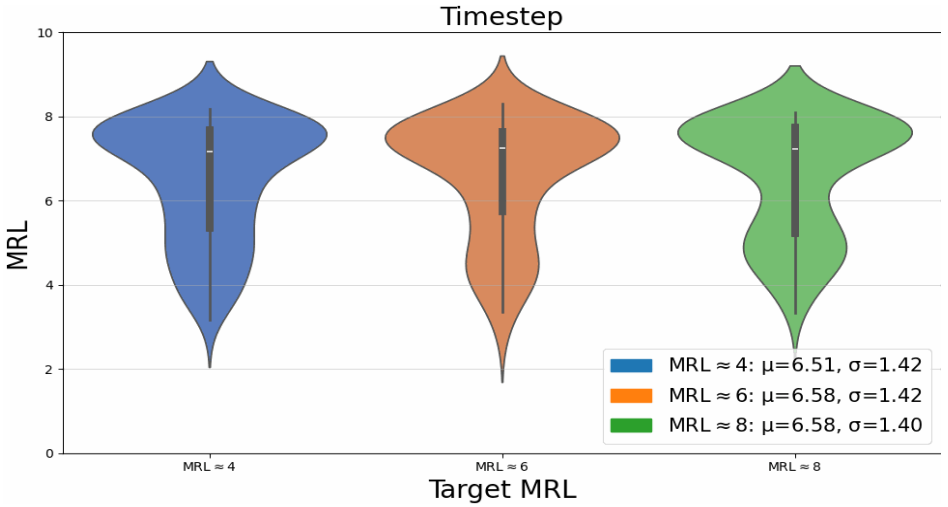
t = t-1

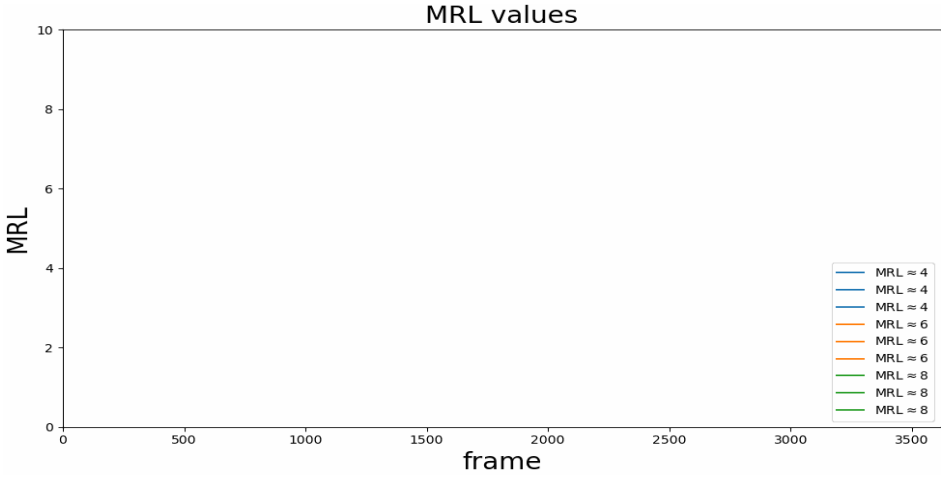
Conditional RNA generation with fixed amino acids (by target MRL)

- Target MRL: 4, 6, 8



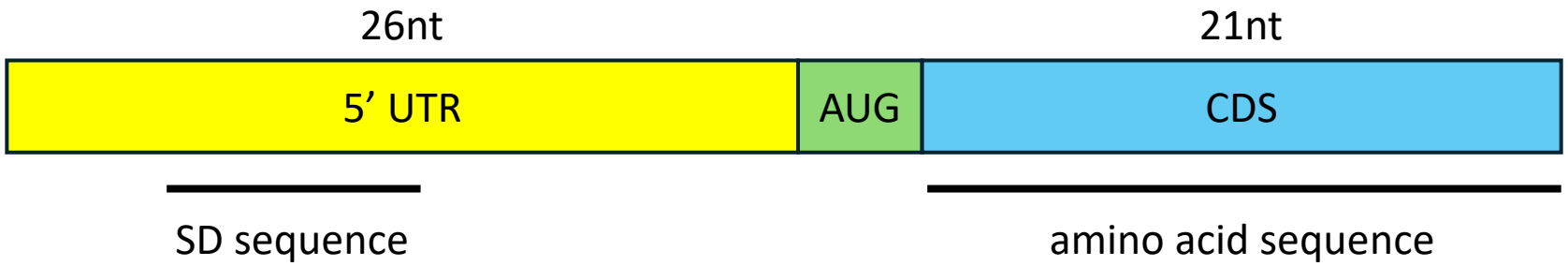

```
'K': ['AAA', 'AAG']
'S': ['UCU', 'UCC', 'UCA', 'UCG', 'AGU', 'AGC']
'V': ['GUU', 'GUC', 'GUA', 'GUG']
```





UTR-Diffusion: Designing mRNAs near 5′ UTR

- Enables the design of mRNA sequences near the 5′ UTR with high translational efficiency (high MRL and MFE)
- Designs can also specify the SD sequence and the amino acid sequence around the start codon

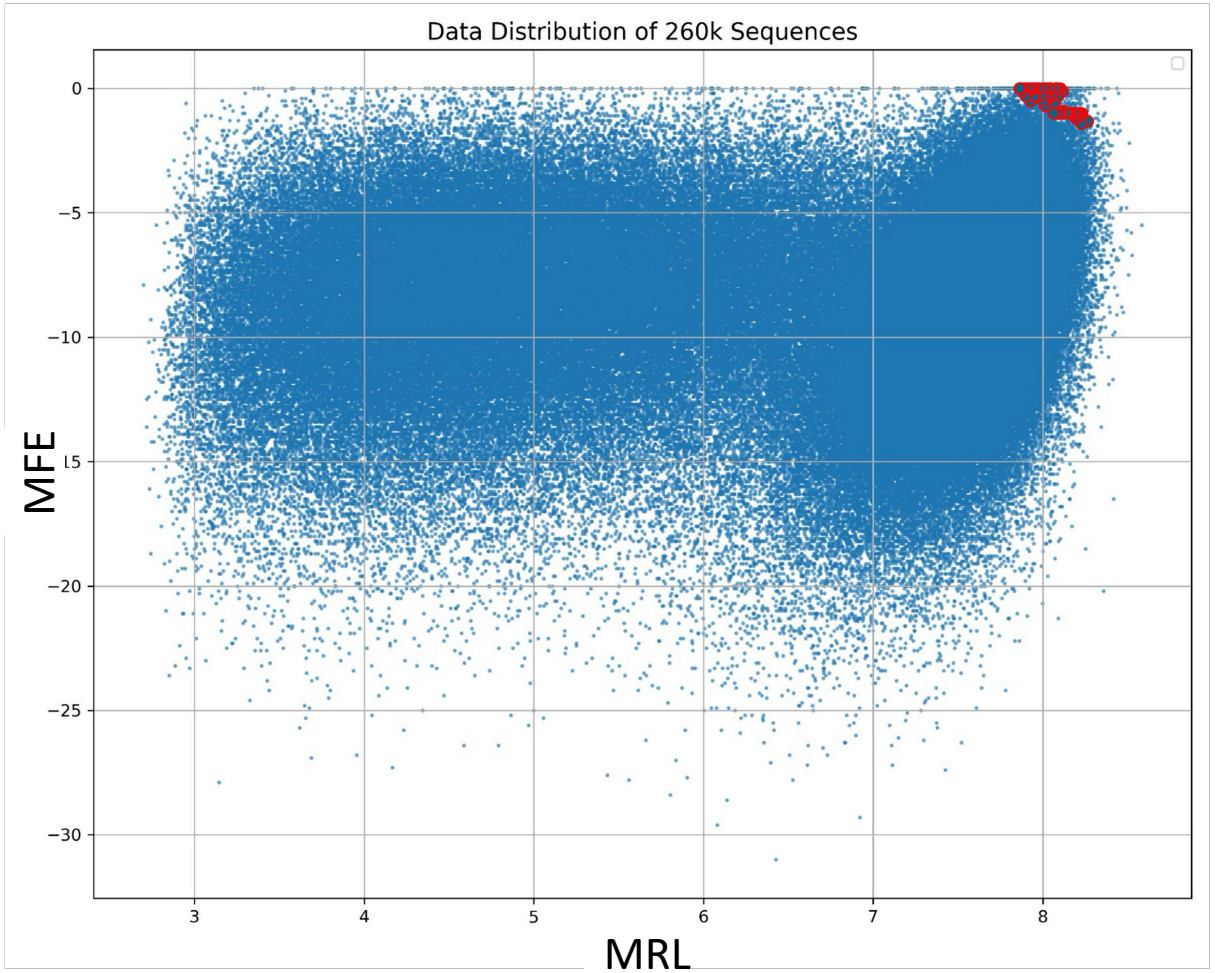


- An integrated pipeline with LinearDesign [Zhang et al., 2023] has been established for full-length mRNA design

UTR-Diffusion: Designing the NanoLuc 5′ UTR–CDS Junction

Institute of SCIENCE TOKYO

- We jointly designed the NanoLuc 5′ UTR and the first eight CDS codons (encoding MVFTLEDF), targeting high MRL and high MFE.



Current limitations and next steps

Institute of SCIENCE TOKYO

- **Fixed-length generation**
  - 50-nt windows do not directly scale to full-length mRNA.
  - Next: use of transformer-based discrete diffusion models and integration with CDS-design tools.
- **Data and domain dependence**
  - Control relies on a large HEK293-derived, predictor-labeled dataset.
  - Next: broader datasets and more data-efficient guidance.
- **Experimental validation**
  - MRL and MFE are evaluated in silico.
  - Next: collaborative wet-lab validation and closed-loop refinement.

Summary

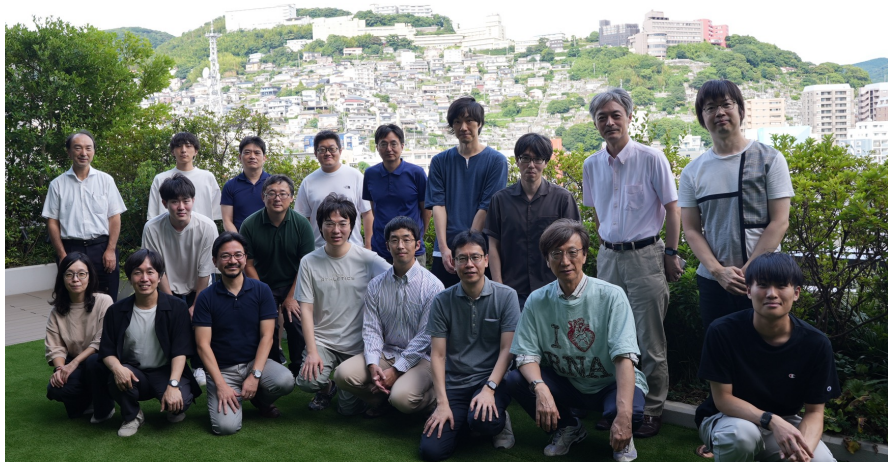
- **Diffusion models make mRNA design controllable.**
- **Goal-directed design:** Diffusion models can generate mRNA sequences for specific design goals.
- **Conditional generation:** The model can control predicted MRL and MFE together.
- **Biological constraints:** Repaint keeps required nucleotides, SD sequences, or amino acids while redesigning the other positions.
- **Computational proof of concept:** NanoLuc candidates show high predicted MRL and MFE while preserving the required amino-acid sequence.

- **Next step:** Test selected generated sequences in wet-lab experiments.

Acknowledgements

- Sato Lab (Institute of Science Tokyo)
  - Dr. Sho Tsukiyama
  - **Dr. Chuankai Dai**
- Asai Lab (ALIS/ROIS)
  - Prof. Kiyoshi Asai
  - Dr. Goro Terai
- Saito Lab (Kitasato University)
  - Prof. Yutaka Saito
  - Dr. Bian Bian
- Abe Lab (Nagoya University)
  - Prof. Hiroshi Abe
  - Dr. Fumitaka Hashiya

Sato Lab Party
(2026 Feb)



RNA Dojo
(2025 Aug)