

RNA 3D STRUCTURE PREDICTION USING GENERATIVE ADVERSARIAL NETWORKS

M. Antczak^{1,2}, M. Justyna¹, M. Szachniuk^{1,2}

¹*Institute of Computing Science, Poznan University of Technology, Poznan, Poland.*

²*Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland.*

Abstract: Understanding the various functions that RNAs play [1] depends significantly on the knowledge of their 3D folds [2]. Unfortunately, the structures of most RNAs are still unknown. Because of the difficulties in the experimental determination of RNA structures, there is a great interest in new, high-accuracy 3D structure prediction methods.

Nowadays, deep neural networks successfully acquire hidden knowledge in extremely large-size data volumes [3]. They can successfully compete with experts in solving the biologically-inspired problems arising in medical image analysis [4], bioinformatics [5], and drug design [6]. For example, the deep-learning-based method (AlphaFold2) has shown high performance in the 14th edition of Community-Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction [7].

A similar achievement in the RNA structures field is a matter of time. However, it requires access to a large volume of diverse structural data used for training. Not much of this data is available yet. And this is one of the reasons why traditional prediction methods still win the competition with AI-based models, e.g., in the last round of CASP.

Here, we present a new concept integrating fragment recombination with Generative Adversarial Networks (GANs), template-based, and template-free modelling.

The main advantage of GANs [8] is the ability to infer and learn patterns from training data in such a way as to generate plausible structures not included in the original dataset. GANs generally consist of two sub-models: a generator and a discriminator. Both models are trained together in an adversarial game until the generator returns plausible 3D motifs/structures for at least half cases. The latter distinguishes native from irregular folds.

We are currently working on a benchmark set of RNA 3D structures and their fragments for successful training AI-based models for RNA 3D structure prediction. Moreover, we experiment with several machine learning techniques, e.g., Denoising Diffusion Probabilistic Models.

References:

- [1] Mortimer et al. (2014) Insights into RNA structure and function from genome-wide studies, *Nature Rev Gen* 15:469-479.
- [2] Leontis and Westhof (2012) RNA 3D Structure Analysis and Prediction. In Leontis, N. & Westhof, E. (Eds.), *Nucleic Acids and Molecular Biology Series: 27*, Springer-Verlag, Berlin and Heidelberg.

- [3] Stahl and Jordanov (2012) An overview of the use of neural networks for data mining tasks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(3), 193-208.
- [4] Tajbakhsh et al. (2016) Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5), 1299-1312.
- [5] Min, S., Lee, B., & Yoon, S. (2017) Deep learning in bioinformatics. *Briefings in bioinformatics*, 18(5), 851-869.
- [6] Popova et al. (2018) Deep reinforcement learning for de novo drug design. *Science advances*, 4(7), eaap7885.
- [7] Pereira et al. (2021) High-accuracy protein structure prediction in CASP14, *Proteins: Structure, Function, and Bioinformatics*, 89(12), 1687-1699.
- [8] Goodfellow et al. (2020) Generative adversarial networks, *Communications of the ACM* 63:139-144.