New Data Science in Nucleic Acids Chemistry (15): Universal prediction of DNAzyme activity using new nearest neighbor parameters and AI

(*1FIBER, Konan Univ. *2FIRST, Konan Univ., *3Waseda Univ., Faculty of Science and Engineering, *4Waseda Univ. WIAS*)
○Shuntaro Takahashi,[1,2] Saptarshi Ghosh,[1] Hisae Tateishi-Karimata,[1,2] Tomohiro Nishimura.[3] Tsukasa Fukunaga,[4] Michiaki Hamada,[3] Naoki Sugimoto[1]
Keywords: Nearest neighbor (NN) model, Duplex, Stability prediction, AI, Functional prediction

For the functions of DNA and RNA based on base pairing such as Watson-Crick base pairs, the stability of forming base pairing is fundamental. Nearest neighbor (NN) model is the most successful method to predict the Gibbs free energy (ΔG^{o}_{37}) of the formation of duplexes. We have recently reported NN parameters for all duplexes (DNA, RNA, and RNA/DNA hybrid) for the stability prediction universally available diverse crowding conditions.[1-4] As the catalytic functions of nucleic acids require the suitable solution conditions with especially Mg^{2+} and molecular crowding, such versatile prediction of the duplex stability is of interest to utilize the nucleic acids functions. Here, we demonstrated the prediction of the activity of the RNA-cleaving 8-17 DNAzyme by using ΔG^{o}_{37} values for DNA duplexes predicted by our newly developed nearest neighbor (NN) parameters. As the enzymatic activity depends on Mg^{2+} concentrations and molecular crowding conditions, we improved the NN model to segmentate the thermodynamic parameters of each NN base pair into those affected by cation and cosolulte concentrations. By using the updated NN parameters and enzymatic activity in different Mg^{2+} concentrations and various molecular crowdings as a database, we characterized the relationship between the sequence information and enzymatic activity with machine learning techniques (Figure 1). Our results suggest that the NN parameters supported to efficiently and accurately predict the enzymatic reaction, which highlights the advantage of the NN parameters for predicting the functions of nucleic acids from the sequence information.
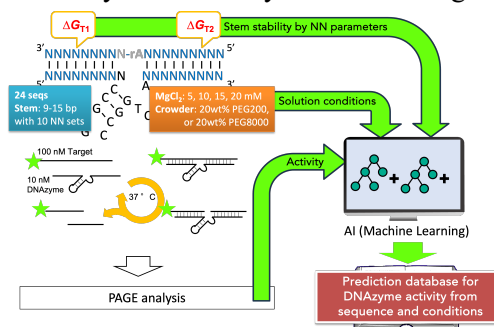

Figure 1. Schematic illustration of development of prediction database for DNAzyme activity from sequence and conditions.

1) S. Ghosh, S. Takahashi, T. Ohyama, T. Endoh, H. Tateishi-Karimata, and N. Sugimoto. *Proc. Natl. Acad. Sci. U. S. A.,* **2020,** *117*, 14194. 2) S. Ghosh, S. Takahashi, D. Banerjee, T. Ohyama, T. Endoh, H. Tateishi-Karimata and N. Sugimoto. *Nucleic Acids Res.*, **2023,** *51*, 4101. 3) D. Banerjee, H. Tateishi-Karimata, M. Toplishek, T. Ohyama, S. Ghosh, S. Takahashi, M. Trajkovski, J. Plavec, and N. Sugimoto. *J. Am. Chem. Soc.,* **2023,** *145*, 23503. 4) S. Ghosh, S. Takahashi, T. Ohyama, L. Liu, N. Sugimoto, *J. Am. Chem. Soc.,* **2024**, 146, 32479.