De *novo* structure modeling for nucleic acids in cryo-EM maps using deep learning

Xiao Wang[1], Genki Terashi[2], and Daisuke Kihara[1,2,3*]

[1] Department of Computer Science, Purdue University, West Lafayette, Indiana, 47907, USA

[2] Department of Biological Sciences, Purdue University, West Lafayette, Indiana, 47907, USA

[3] Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN, 47907, USA

DNA and RNA play fundamental roles in various cellular processes, where the three-dimensional (3D) structure provides critical information to understand molecular mechanisms of their functions. Although an increasing number of structures of nucleic acids and their complexes with proteins are determined by cryogenic electron microscopy (cryo-EM), structure modeling for DNA and RNA is still often challenging particularly when the map is determined at sub-atomic resolution. Moreover, computational methods are sparse for nucleic acid structure modeling. Therefore, we developed a deep learning-based fully automated *de novo* DNA/RNA atomic structure modeling method, CryoREAD. CryoREAD first identifies phosphate, sugar, and base positions in a cryo-EM map using deep learning. Subsequently, sugar backbones are traced, and the nucleic acid sequence is mapped along the backbone considering our predictions along the backbone path. Finally, a full atomic model including nucleotide bases is reconstructed. When tested on cryo-EM maps determined at 2.0 to 5.0 Å resolution, CryoREAD built substantially accurate models than existing methods. We have further applied the method on cryo-EM maps of biomolecular complexes in SARS-CoV-2. The code is available at https://github.com/kiharalab/CryoREAD together with other cryo-EM software https://kiharalab.org/emsuites.