

Deep Learning-Enabled Approach for the Identification and Atomic Model Building of Large RNA-Protein Complexes from Cryogenic Electron Microscopy Maps

Cryogenic electron microscopy (cryoEM) has revolutionized structural biology by allowing the visualization of complex molecular structures and engineered nano-particles at atomic resolution. Reliable atomic models constructed from cryoEM maps contribute to the understanding of their functionalities and can inform bioengineering applications. CryoEM now can accommodate heterogeneous samples enriched directly from endogenous sources without the need of tedious purification process. However, this approach introduces an intriguing challenge: if a near atomic-resolution (3.0-4.0 Å) cryoEM map of a nucleic acid-protein complex is obtained from a heterogeneous sample, how to identify the complex this map represents?

Moreover, given the challenges and time-consuming nature of manually constructing a reliable atomic model from a cryoEM map of a nucleic acid-protein complex, could artificial intelligence (AI) assist in distinguishing between DNA, RNA and protein and in automatically constructing the atomic model? **To address these challenges, this project aims to develop a deep learning-enabled approach, named cryoID, for the identification of RNA-protein complex from a cryoEM map and building their atomic models automatically.** The team will primarily engage in three research components:

1. **Local database construction:** scour public databases for entries containing nucleic acids and construct a local database for AI training and testing the deep learning model
2. **Graphical user interface (GUI) design:** design user-friendly GUI for the identification and model building

Phase 1 Requirements

- The database should be sufficiently extensive to include various RNA-protein complexes, such as ribosomes, spliceosomes, polymerases, helicases, and more.
- The local database should ensure accurate alignment between cryoEM maps and their related atomic models.
- Redundant or similar structures should be eliminated from the local database.

The project mentors will provide instruction on Python and offer assistance as needed.

Phase 2 Requirements

- Develop a user-friendly GUI that facilitates the identification and model-building processes, even for new users.

The project mentors will supply the team with a deep learning model for training and testing using the locally built database. If the team possesses adequate machine learning capabilities and has a desire to learn and undertake more, additional AI-related tasks will be incorporated.

Depending on the level of project success resulting from team efforts, the results could be published in a professional journal and members have the potential to be listed as co-authors on such a publication.

Mentor: Prof. Z. Hong Zhou (Bioengineering, CNSI and MIMG)

TA: Dr. Qibo Xu