CryoID: A Deep Learning Framework for Enhanced Identification and Automated Model Construction of Unknown RNA-Protein Complexes from CryoEM Maps

Qibo Xu, Michael Rebelo, Rudy Narayan, Leon Wu, Neha Kulkarni, Grayson Feng, Neha Pai, Star Yu, Z. Hong Zhou

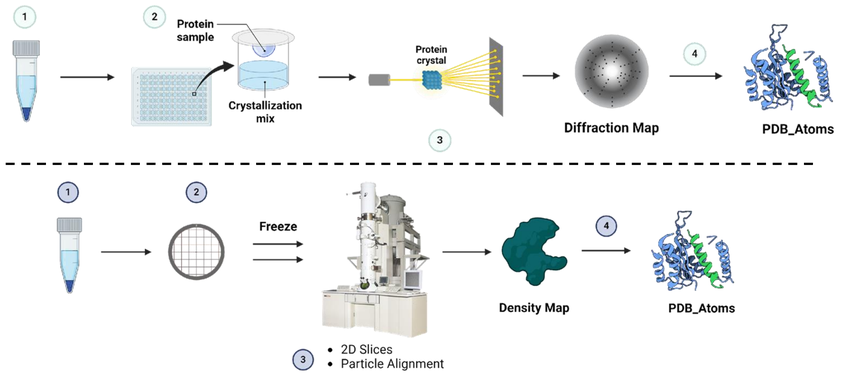
California NanoSystems Institute (CNSI) and Department of Bioengineering, University of California, Los Angeles, 607 Charles E. Young. Drive East Los Angeles, CA, 90095-1569, United States

Project Overview

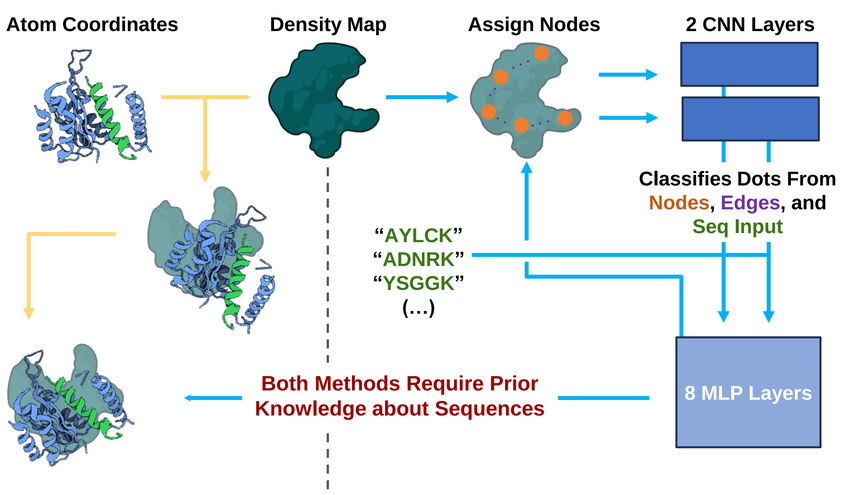
Protein-RNA complexes are crucial cellular components governing fundamental biological processes like gene regulation and DNA translation. It is essential to examine their structural details to understand their unique functions. Cryo-electron microscopy (Cryo-EM) has become widely used to observe protein structures at high resolutions, exemplified by the deposition of 7770 Cryo-EM structures in 2023. Despite its increased use, user accessibility is a significant barrier for Cryo-EM methods. Although many software tools claim to simplify workflows, they often convolute the process by requiring users to input an amino acid sequence for structure determination.

Here, the group created Cryo-ID, a user-friendly application with a robust U-Net algorithm predicting RNA and protein residues from input density maps. The application also contains access to commonly used protein software, including BLAST and AlphaFold. Through cryoID's holistic pipeline, the group hopes to provide equitable access to cryoEM to researchers, empowering them to make continued contributions in their field.

Protein Imaging: X-Ray Crystallography vs Cryo-EM

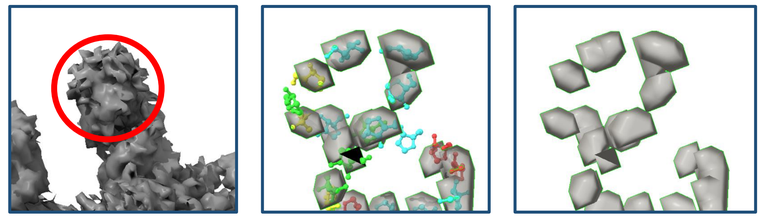


Current Cryo-EM Analysis Workflow: Phenix/ModelAngelo



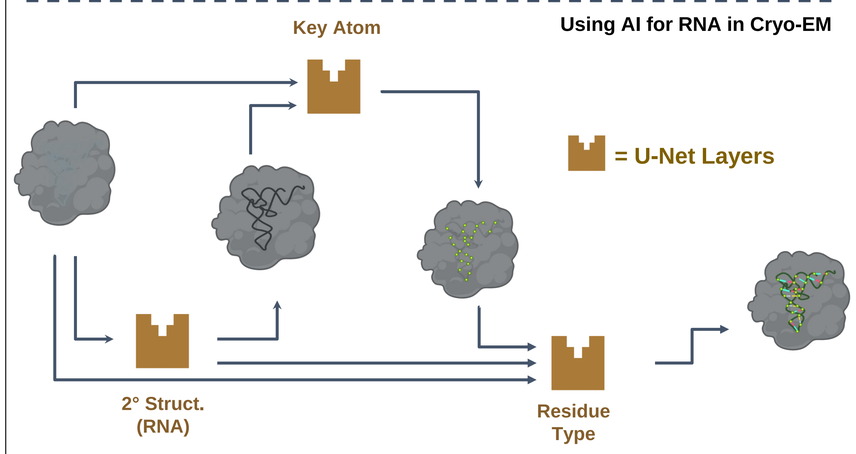
Our Methodology

Relating Points to Densities:

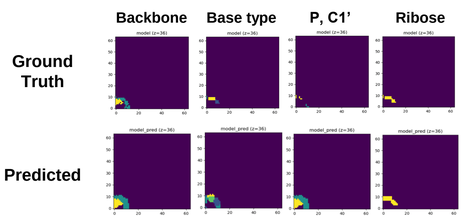


Translated Points in space to "Pseudo-Density Volumes" using Octahedral parameterization

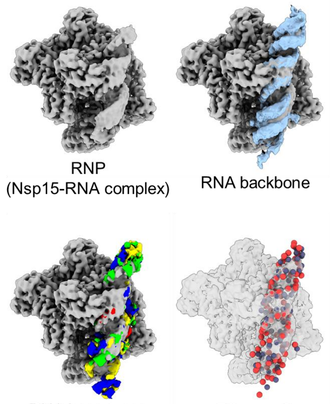
Using AI for RNA in Cryo-EM



RNA Prediction



	Backbone	Base type	P, C1'	Ribose
Accuracy	.9685	.9854	.9878	.9664
Early Stop Epoches	164	133	147	110



User Interface

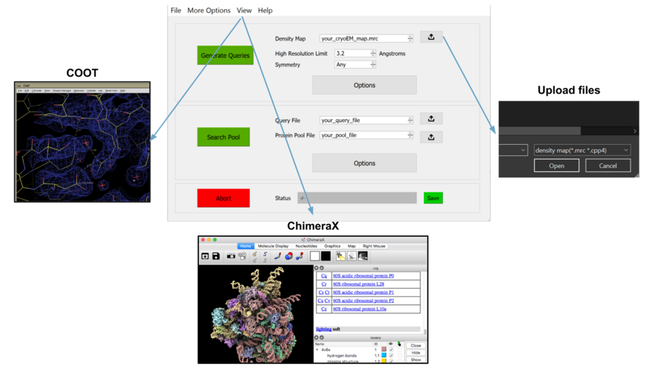


Figure showing user-friendly GUI with parameter flexibility, help function, option to add user files, and high resolution figures of RNA-protein structure

Implications

- **Streamlined** cryo-EM data analysis methodology reshapes the landscape of structural biology and paves the way for transformative discoveries in protein-RNA complexes
- Enhances the **accuracy, speed, and versatility** of predicting RNA and protein structures from cryo-EM maps
- Works towards providing **equitable** access to institutions and researchers that lack tools to analyze computationally expensive data
- **User-friendly GUI** enables the democratization of cryo-EM across **diverse computational backgrounds**, therefore integrating cryo-EM into routine practices, fostering collaboration and inclusivity in the scientific community

Conclusions & Future Work

- **High throughput AI** tool enables the structural identification and data evaluation of protein-RNA complexes
- Reduction in time and resource investments fosters the **integration of cryo-EM into routine research practices**
- Future work involves optimizing the algorithm to provide researchers with a robust and accessible tool for investigating molecular structures in protein-RNA complexes
- Extension to applications in other disciplines of structural biology such as gene regulation, cellular signaling, and virus-host interactions.
- Promising direction at the intersection of structural biology, cryo-EM, and AI, driving innovation and facilitating discoveries.

Acknowledgements

CNSI AT UCLA
UCLA Samueli Bioengineering
NSF