# —Supplementary Material— CryoFastAR: Fast Cryo-EM *Ab initio* Reconstruction Made Easy

## A. Additional Results

We compare the visual results of our method with baselines methods on the simulated Spliceosome [5] dataset. The results are shown in Figure 2, and the quantitative results are presented in Table 1 of the paper.

## B. Details of Dataset

The generation process for the simulated dataset follows the procedure outlined in Section E, under the paragraph titled **Simulated Particle Image Generation**, with some differences in the number of structures used and the number of projections.

### B.1. Simulated dataset.

We generate 3 simulated datasets for evaluation of the baselines. Each dataset has the same simulation procedure: 1) generating 50000 particles with uniformly sampled rotations, 2) adding CTF corruption 3) translating image in $[-10, 10]$ pixels along $x$ and $y$ axes, respectively, 4) add Gaussian noise to adjust the signal-to-noise ratio (SNR) of the image to 0.1. Each structure in the dataset has a different spatial resolution in terms of Å per pixel(Apix) when the PDB [1] structure is converted into a volume density map using EMAN2 [11]. The PDB ID and Apix for each dataset is as follows:

- The spliceosome structure (**Spliceosome**, PDB ID: 5nrl) [5]. Apix: 4.00.
- A variant of the SARS-CoV-2 spike protein (**Spike**, PDB ID: 7sbr) [12]. Apix: 2.03.
- the human Fanconi anaemia core complex (**FA**, PDB ID: 7kzp) [13]. Apix: 2.54.

### B.2. Experimental dataset.

For the experimental evaluation, we test on three datasets (EMPIAR-10049, EMPIAR-10076, and EMPIAR-10180) from EMPIAR [3]. After processing these datasets using the scripts provided by CryoDRGN-EMPIAR [14], we obtain the filtered particle stacks, along with the pre-computed accurate particle poses and 2D in-plane translations. We assume that the reconstructed structures, using the pre-computed poses and translations provided by the datasets, represent the ground-truth volume. The particles are then split into two sets: 30,000 particles for training and the remaining particles for evaluation. The number of images in the evaluation set and the Apix for each dataset are as follows:

- The RAG1-RAG2 complex (**RAG**, EMPIAR-10049) [8]. Apix: 1.845, number particles in the evaluation set: 78544.
- The assembling bacterial 50S ribosome (**50S**, EMPIAR-10076) [2]. Apix: 3.275, number particles in the evaluation set: 57327.
- The pre-catalytic spliceosome (**Spliceosome**, EMPIAR-10180) [4]. Apix: 4.25, number particles in the evaluation set: 109722.

### B.3. Contrast Transfer Function

In cryo-electron microscopy (cryo-EM), the imaging process is influenced by the point spread function (PSF), which characterizes the system's spatial response to a point source. The PSF encapsulates the effects of diffraction, aberrations, and other instrumental imperfections, thereby determining how the interactions between the high-energy electron beam and the specimen are distributed in the final image in real space.

For image processing and analysis, it is often advantageous to work in the frequency domain, typically by applying a Fourier or Hartley transform. In this domain, the Fourier transform of the PSF is referred to as the Contrast Transfer Function (CTF), which describes how different spatial frequency components are modulated by the microscope's optics.

In our work, instead of applying the PSF directly, we incorporate its effects through the CTF. This approach simplifies the image processing workflow by enabling us to operate entirely in the frequency domain. Following the methodology implemented in CTFFIND4 [7], the CTF is defined as:

$$
\begin{aligned}
\text{CTF}(w, \lambda, \mathbf{g}, \Delta f, C_s, \Delta\varphi) = \\
-\sqrt{1 - w^2} \sin[\chi(\lambda, |\mathbf{g}|, \Delta f, C_s, \Delta\varphi)] \\
-w \cos[\chi(\lambda, |\mathbf{g}|, \Delta f, C_s, \Delta\varphi)]
\end{aligned}
\tag{1}
$$

where

$$
\chi(\lambda, |\mathbf{g}|, \Delta f, C_s, \Delta\varphi) = \pi\lambda|\mathbf{g}|^2 \left(\Delta f - \frac{1}{2}\lambda^2|\mathbf{g}|^2 C_s\right) + \Delta\varphi.
\tag{2}
$$

In these equations, $w$ represents the relative phase contrast factor, while $\chi$ is a frequency-dependent phase shift function. The function $\chi$ incorporates key parameters, including the electron wavelength $\lambda$, the spatial frequency vector $\mathbf{g}$, the objective defocus $\Delta f$, the spherical aberration $C_s$, and the phase shift $\Delta\varphi$. The parameters $w$, $\lambda$, $C_s$, and $\Delta\varphi$ are intrinsic to the cryo-EM hardware.
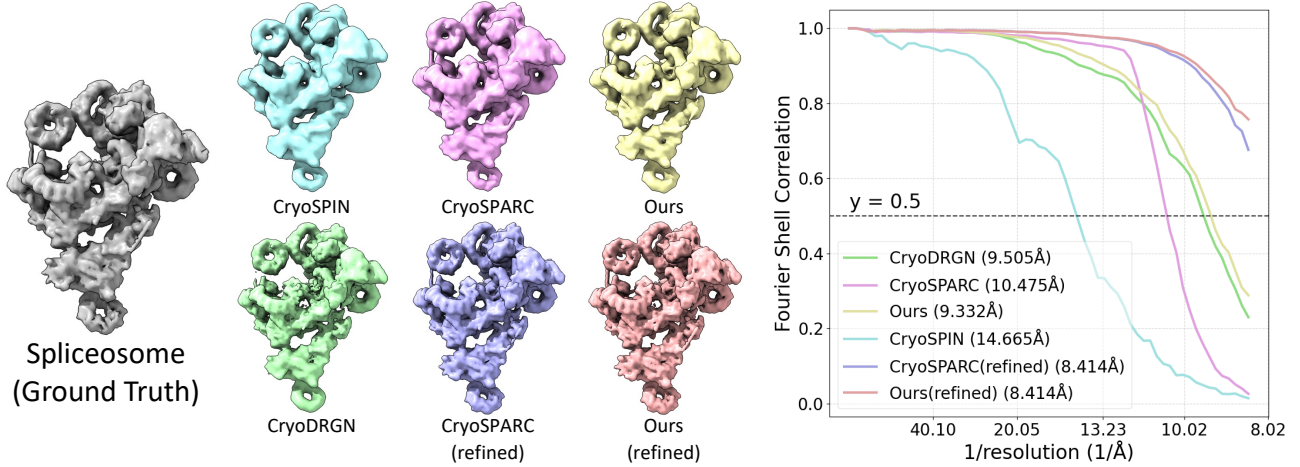
Figure 2. **Qualitative Result.** We compare our visual quality with all other baselines before and after the refinement for Spliceosome's simulated dataset. The results show that our method is comparable to them before refinement and achieves the best performance after the refinement.

## C. Details of Baselines

**CryoSPIN [9].** We use the official implementation of CryoSPIN in Github. In our experiments, we run CryoSPIN with its default setting while using our customized datasets as described in the main paper. Also, the official implementation does not include the estimation of the in-plane translation, we omit the random shift when training CryoSPIN. As CryoSPIN often falls to local minima, we run it three times per experiment and report the best result.

**CryoDRGN2 [15].** We use the official implementation of CryoDRGN v3.4.3 in Github, with all default parameter values, except setting the batch size to 32 while using the 'abinit_homo' command for *ab initio*.

**CryoSPARC [6].** We use the CryoSPARC software package v4.6.2. with all default parameter values. We followed the typical workflow: Import particle stacks and then perform *ab initio* reconstruction.

**CryoSPARC(refined).** We use the CryoSPARC software package v4.6.2. with all default parameter values for refinement. We use the result mentioned in paragraph **CryoSPARC**, continuously performs Homogeneous Refine, and finally a Local Refinement.

**Ours(refined).** We use the CryoSPARC software package v4.6.2 with all default parameter values for refinement. We use Import Particles, and then Reconstruction Only to generate volumes and masks for the future usage, and then run Local Refinement.

## D. Details of Evaluation Metrics

**Rotation F-norm error.** Given a sequence of ground-truth 3D orientations $R_1, \ldots, R_N$ and a sequence of estimated 3D orientations $\hat{R}_1, \ldots, \hat{R}_N$, we randomly select one view, indexed as $i$, to serve as the reference view. To report the F-norm rotation errors for each dataset, we randomly sample 5,000 views as reference views and select the minimum error as the final result. For each random selection $i$, all ground-truth poses are transformed into the coordinate system of this reference view: $R'_{i,j} \leftarrow R_i R_j^\top, \forall j = \{1, \ldots, N\}$. Similarly, all predicted poses are transformed into the coordinate system of the reference view of the $i$-th predicted pose: $\hat{R}'_{i,j} \leftarrow \hat{R}_i \hat{R}_j^\top, \forall j = \{1, \ldots, N\}$. The error for each view is then computed by taking the Frobenius norm of the difference between the transformed ground-truth pose and the corresponding transformed predicted pose. Finally, the average of these errors across all views is computed:

$$\mathcal{L}_{\text{rot},i} = \frac{1}{N} \sum_{j=1}^{N} \| R'_{i,j} - \hat{R}'_{i,j} \|_F. \tag{3}$$

**In-plane translation error.** Given a ground-truth 2D in-plane translation $\mathbf{t} = (t_x, t_y)$ and an estimated translation $\hat{\mathbf{t}} = (\hat{t}_x, \hat{t}_y)$, we compute the mean of the L2-norm:

$$\mathcal{L}_2(\mathbf{t}, \hat{\mathbf{t}}) = \| \mathbf{t} - \hat{\mathbf{t}} \|_2. \tag{4}$$

We report the average of the L2 translation errors over each dataset.

**Resolution.** The reconstruction resolution is calculated using the Fourier Shell Correlation (FSC) between the

aligned reconstructed and ground-truth volumes, with thresholds of 0.5 for simulated data and 0.143 for experimental data serving as the metric for pose estimation reconstruction resolution. The formula for FSC is given below:

$$FSC(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum_{r_i \in r} \|F_1(r_i)\|^2 \cdot \sum_{r_i \in r} \|F_2(r_i)\|^2}} \quad (5)$$

where $F_1$, $F_2$ are the Fourier transforms of the reconstructed and ground-truth volumes, respectively. $r$ represents all three-dimensional frequency components shown in a one-dimensional form. We use CryoSPARC's Align 3D Maps to automatically align the ground-truth volume with the other reconstructed volume.

## E. Details of Simulated Training Dataset Construction

**Data Curation.** The data curation pipeline for our atomic structure dataset is designed to obtain high-quality and biologically correct 3D structures from the Protein Data Bank(PDB) [1]. The process involves several key steps, including parsing, metadata extraction, filtering, and structural refinement.

**Parsing and Metadata Extraction** The initial step involves parsing the input files in the mmCIF format. The pipeline extracts essential metadata, including the release date, resolution, and experimental method. These metadata are crucial for ensuring the relevance and reliability of the structures included in the dataset.

**Filtering Criteria** The dataset is subjected to stringent filtering criteria to ensure the quality and suitability of the structures for downstream analysis. The filtering process includes:
- Release Date: Structures must have been released to the PDB before the cutoff date of 2021-09-30.
- Resolution: Only structures with a reported resolution of 9Å or less are retained.
- Hydrogen Removal: Hydrogen atoms are removed from the structures.
- Polymer Chain Integrity: Polymer chains with all unknown residues are removed.
- Clashing Chains: Chains with more than 30% of atoms within 1.7Å of an atom in another chain are identified as clashing. In cases where two chains are clashing, the chain with the greater percentage of clashing atoms is removed. If the same fraction of atoms are clashing, the chain with fewer total atoms is removed. If the chains have the same number of atoms, the chain with the larger chain ID is removed.

- Residue and Small Molecule Integrity: For residues or small molecules with CCD codes, atoms outside of the CCD code's defined set of atom names are removed. Protein chains with consecutive $C\alpha$ atoms larger than 10Å apart are filtered out.
- Bioassembly Selection: For bioassemblies with more than 20 chains, a random interface token is selected, ensuring that the center atom is within 15Å of the center atom of a token in another chain.

**Structural Refinement** To simplify subsequent analysis, the pipeline performs basic structural cleanup. This includes resolving alternative locations for atoms/residues by selecting the one with the largest occupancy and removing water and crystallization aids. Ligands, peptides, and nucleic acids are also removed to focus on the core protein structures.

**Simulated Particle Image Generation** We begin with the 113,600 curated 3D atomic structures and generate noisy projection images for training through the following steps: For each 3D structure, we convert it into a $128^3$ 3D volume density map using EMAN2 [11], the pixel size is set by the maximum length of the structure to make sure the whole structure is inside. Then we uniformly sample 100 projection images in SO(3) space, representing different views of the volume. Each clean projection image is then modified by randomly applying a Contrast Transfer Function (CTF) sampled from real distribution including 152,385 CTF parameters as described in [10], simulating imaging system imperfections such as aberrations or blurring. A random 2D shift, within the range of $[-10, 10]^2$, is applied to each image to account for potential translational errors during imaging. Finally, Gaussian noise is added to each image to ensure the signal-to-noise ratio (SNR) matches the desired level, simulating the noise characteristics in experimental data. This augmentation pipeline produces diverse, realistic projections for model training.

## References

[1] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. 2, 4

[2] Joseph H. Davis, Yong Zi Tan, Bridget Carragher, Clinton S. Potter, Dmitry Lyumkis, and James R. Williamson. Modular assembly of the bacterial large ribosomal subunit. *Cell*, 167: 1610–1622.e15, 2016. 2

[3] Andrii Iudin, Paul K. Korir, José Salavert-Torres, Gerard J. Kleywegt, and Ardan Patwardhan. Empiar: a public archive for raw electron microscopy image data. *Nature Methods*, 13 (5):387–388, 2016. 2

[4] Takanori Nakane, Dari Kimanius, Erik Lindahl, and Sjors HW Scheres. Characterisation of molecular motions

in cryo-em single-particle data by multi-body refinement in relion. *eLife*, 7:e36861, 2018. 2

[5] Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. Structure of a pre-catalytic spliceosome. *Nature*, 546, 2017. 2

[6] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature Methods*, 14 (3):290–296, 2017. 3

[7] Alexis Rohou and Nikolaus Grigorieff. Ctffind4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology*, 192(2):216–221, 2015. 2

[8] Heng Ru, Melissa Chambers, Tian-Min Fu, Alexander Tong, Maofu Liao, and Hao Wu. Molecular mechanism of v(d)j recombination from synaptic rag1-rag2 complex structures. *Cell*, 163, 2015. 2

[9] Shayan Shekarforoush, David Lindell, Marcus A Brubaker, and David J Fleet. Cryospin: Improving ab-initio cryo-em reconstruction with semi-amortized pose inference. *Advances in Neural Information Processing Systems*, 37: 55785–55809, 2025. 3

[10] Yingjun Shen, Haizhao Dai, Qihe Chen, Yan Zeng, Jiakai Zhang, Yuan Pei, and Jingyi Yu. Draco: A denoising-reconstruction autoencoder for cryo-em. *Advances in Neural Information Processing Systems*, 38, 2024. 4

[11] Guang Tang, Liwei Peng, Philip R. Baldwin, Deepinder S. Mann, Wen Jiang, Ian Rees, and Steven J. Ludtke. Eman2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*, 157(1):38–46, 2007. Software tools for macromolecular microscopy. 2, 4

[12] Alexandra C. Walls, Young-Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veesler. Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. *Cell*, 181(2):281–292.e6, 2020. 2

[13] Shengliu Wang, Renjing Wang, Christopher Peralta, Ayat Yaseen, and Nikola P. Pavletich. Structure of the fa core ubiquitin ligase closing the id clamp on dna. *Nature Structural & Molecular Biology*, 28:300 – 309, 2021. 2

[14] Ellen D Zhong, Tristan Bepler, Bonnie Berger, and Joseph H Davis. Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks. *Nature methods*, 18(2): 176–185, 2021. 2

[15] Ellen D. Zhong, Adam Lerer, Joseph H. Davis, and Bonnie Berger. Cryodrgn2: Ab initio neural reconstruction of 3d protein structures from real cryo-em images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4066–4075, 2021. 3