

---

# CryoSTAR: Cryo-EM Heterogeneous Reconstruction of Atomic Models with Structural Regularization

---

Yi Zhou\*, Yilai Li\*, Jing Yuan\*, Fei Ye, Quanquan Gu<sup>†</sup>  
Bytedance Research

{zhouyi.naive, yilai.li, yuanjing.eugene}@bytedance.com,  
{yefei.joyce, quanquan.gu}@bytedance.com

## Abstract

Atomic models, which directly represent molecular structural variations (i.e., conformation), have received increasing attention in the field of cryo-electron microscopy (cryo-EM) heterogeneity analysis. However, the nonconvex landscape of the structural space (the space of atomic coordinates) poses a significant challenge to finding a physical-plausible structure. In this paper, we address this challenge by proposing a novel approach, named cryoSTAR, with the aim of reconstructing atomic models from cryo-EM images. Our approach is motivated by the observation that weak regularization allows atomic models to be excessively flexible in the search space, resulting in a loss of local structural fidelity, while strong regularization tends to trap atomic models in the neighborhood of the initial structure, limiting their ability to explore the conformational landscape effectively. To strike a balance, we introduce adaptive structural regularization at the atomic level to modulate the reconstruction process. We relax the flexible region adaptively to allow for greater conformational changes. Our method achieves the lowest RMSD (up to a maximum decrease of 7.14Å) on a synthetic dataset, and uncovers reasonable dynamics on an experimental dataset, highlighting its generalizability across different protein systems. Our work sheds light on the potential of atomic models as an alternative to traditional volumetric density maps for cryo-EM heterogeneous reconstruction.

## 1 Introduction

Cryo-electron microscopy (cryo-EM) is a structural biology tool that can directly observe the conformational heterogeneity of biomolecules at the single particle level [1]. In cryo-EM, each dataset consists of numerous 2D projections of one or more 3D structures with potentially different conformations. Therefore, cryo-EM datasets can provide rich heterogeneous information at the conformational level.

Traditional approaches focus on solving cryo-EM reconstruction in *volumetric* space by voxelizing the electronic density map into discrete 3D grids [2, 3, 4]. When dealing with heterogeneity, many of them assume a discrete number of classes and assign each particle to a single class [5]. Recently, there have been several algorithms that are designed to tackle heterogeneity in a continuous manner. For example, 3DVA [6] describes the variability within the dataset as a linear combination of a few bases. To achieve more expressive power with nonlinearity, deep learning based methods have been developed to map such heterogeneity to nonlinear manifold embeddings. Notably, cryoDRGN [7, 8, 9] uses a variational autoencoder (VAE) [10] approach to map heterogeneity within the dataset

---

\*These authors contributed equally to this work

<sup>†</sup>Corresponding author

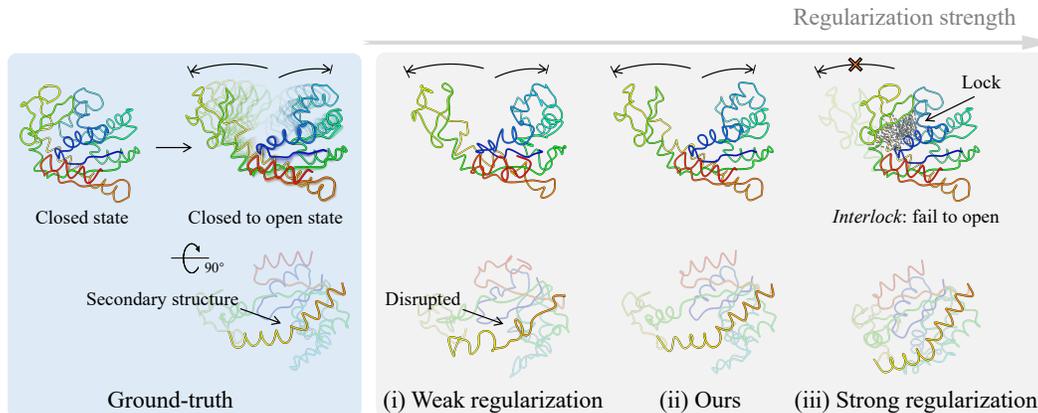


Figure 1: Reconstructed atomic structures *versus* the strength of regularization. Given the **closed** state as the initial model (pdb: *lake*), the goal is to recover the **open** state (pdb: *ake*). (i) A method with weak regularization (only enforcing backbone continuity [13, 14]) may destroy the local structure (e.g.,  $\alpha$ -helix, as highlighted) of a well-folded protein. (ii) We introduce a balance point which can reconstruct the open state while preserving its local structure. (iii) The *interlock* [18] problem prevents the NMA-based method [15, 17, 19] from finding the “closed-to-open” transition, despite its success in finding the “open-to-closed” transition.

to a latent space. A generative decoder is used to generate the 3D volume given a sampled point from the latent space. Most of these *volumetric* methods do not explicitly model the motion of flexible regions, which often leads to density vanishing and artifacts while sampling in the latent space. 3DFlex [11] learns an explicit 3D deformation field, while preserving local geometry and structural information.

Recent advancements in the field have sought to incorporate *atomic*-level information into the pipeline, characterizing conformational changes through the deformation of a reference atomic model. This approach offers the advantage of easily integrating structural information from the reference model. For example, e2gmm [12], cryoFold [13] and atomVAE [14] explored the explicit modeling of (pseudo-)atoms or (pseudo-)residues using Gaussian mixtures. Other methods have attempted to decompose heterogeneity into a few bases using normal mode analysis (NMA) [15, 16, 17]. This atomic level or residue level information often improves output interpretation by providing sensible models with motion. However, these methods either only found relatively small continuous motions, or were only verified on synthetic data.

The *non-convex* landscape of the *atomic* space poses a major challenge to finding a physical plausible structure in cryo-EM reconstruction [15, 13]. An intuitive approach to tackle this challenge is to characterize the search space by structural regularization from prior knowledge. Figure 1 illustrates the results with two popular regularization approaches. Weak regularization that only enforces backbone continuity is prone to disrupt local secondary structures (Figure 1 (i)) [13, 14]. On the other hand, strong regularization such as NMA may suffer from the “*interlock*” [18] problem, causing the results to be biased by the reference atomic model due to the large energy barrier between the conformational states (Figure 1 (iii)). We aim to pursue a solution that *preserves most local structures and prevents the undue influence from the initial structure*.

In this paper, we propose an adaptive **structural** regularization approach, named cryoSTAR, to resolve continuous heterogeneity from cryo-EM datasets. CryoSTAR models the conformational heterogeneity as the deformations of a reference atomic model. To preserve local structures such as helices and strands, cryoSTAR employs an elastic network model (ENM) [20] to measure the energy of the residues after deformation, which is similar to NMA by approximating the energy surface with second-order expansion. Since ENM may be susceptible to the “*interlock*” problem, cryoSTAR modifies the topology of ENM adaptively, based on the assumption that protein dynamics manifest in the change of inter-residue distance.

Our contributions are twofold:

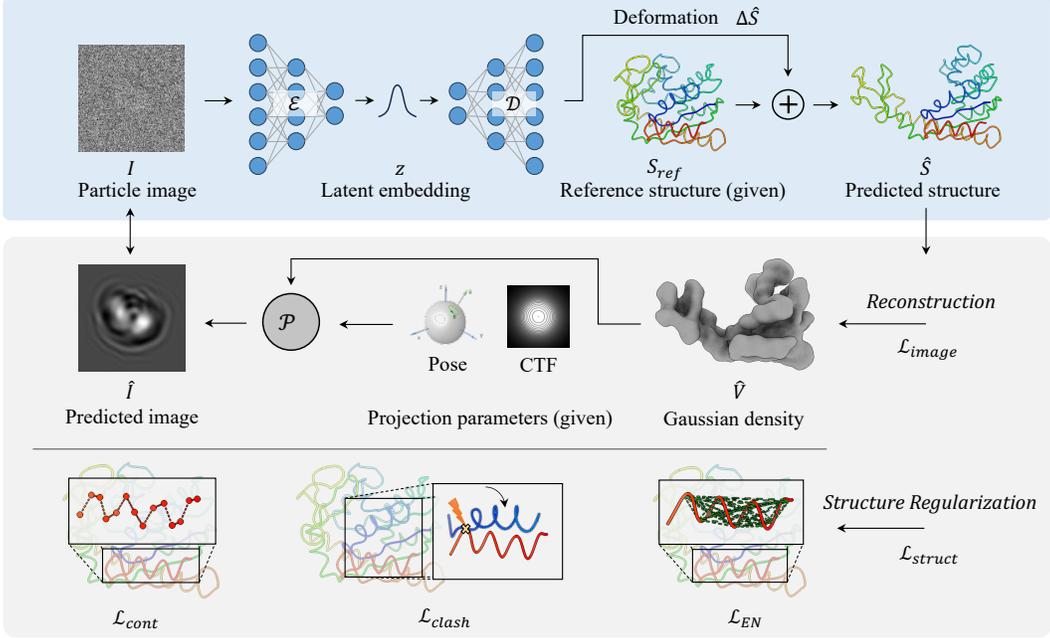


Figure 2: The overall architecture of cryoSTAR. The key component is a *trainable* deformation prediction network ( $\mathcal{E} + \mathcal{D}$ ) mapping an input image  $I$  to an atomic structure  $\hat{S}$ . The learning signal mainly comes from two sources: the reconstruction loss between  $I$  and  $\hat{I}$ , and the structural regularization loss of  $\hat{S}$ . The projection module  $\mathcal{P}$  which transforms the structure  $\hat{S}$  to a 2D image  $\hat{I}$  is pre-defined and *non-trainable*.

- We propose an atomic model based on adaptive structural regularization for cryo-EM heterogeneity analysis. The topology of the elastic network model is constructed adaptively with signals from the cryo-EM images. This enables the preservation of most local structures and prevents convergence into the local minima of the reference structure.
- We apply cryoSTAR to a synthetic dataset and a public experimental dataset. On the synthetic dataset, we achieve a maximum decrease of 7.14Å compared with NMA-based models. On the experimental pre-catalytic spliceosome dataset (EMPIAR-10180 [21]), we successfully capture motions that align with those observed in previous studies [21, 8].

## 2 Methods

The main goal of cryo-EM reconstruction is to recover a 3D volume  $V \in \mathbb{R}^{D \times D \times D}$  from its 2D projections  $\{I^{(i)} \in \mathbb{R}^{D \times D}\}_{i=1}^M$ , where  $D$  is the side length of the image or volume and  $M$  is the size of dataset. By associating an atomic structure with a volume, cryoSTAR outputs an atomic structure  $S \in \mathbb{R}^{N \times 3}$  where  $N$  is the number of residues. Figure 2 shows the pipeline of cryoSTAR. We will discuss each part in this section: (i) given an input image (particle)  $I^{(i)} \in \mathbb{R}^{D \times D}$  (the superscript  $i$  will be omitted for simplicity of notation), a VAE-based deformation prediction network first infers the atomic structure  $\hat{S} \in \mathbb{R}^{N \times 3}$  of a molecule, i.e., the coordinates of  $N$  residues (Section 2.1); (ii) we convert the atomic structure  $S$  to a volumetric representation  $\hat{V} \in \mathbb{R}^{D \times D \times D}$  and calculate its 2D projection  $\hat{I} \in \mathbb{R}^{D \times D}$  (Section 2.2); (iii) in Section 2.3, we propose the structure-aware regularization to constrain the atomic structure, including some basic constraints (Section 2.3.1) and elastic regularization (Section 2.3.2). In Section 2.3.3 we further introduce the adaptive strategy for relaxing the elastic regularization; and (iv) we combine it with the auto-encoding supervision between  $I$  and  $\hat{I}$  to form the overall loss function in Section 2.4.

## 2.1 Deformation prediction network

Following [13, 14, 15], cryoSTAR requires a reference atomic model  $S_{\text{ref}} \in \mathbb{R}^{N \times 3}$ , where  $N$  denotes the number of the residues. Given an image  $I \in \mathbb{R}^{D \times D}$  from the cryo-EM dataset, cryoSTAR uses a variational autoencoder to predict the corresponding deformation  $\Delta \hat{S}$  that transforms the reference structure to the deformed structure  $\hat{S} = S_{\text{ref}} + \Delta \hat{S}$ .

The encoder  $\mathcal{E} : \mathbb{R}^{D \times D} \rightarrow \mathbb{R}^{|z|}$  and decoder  $\mathcal{D} : \mathbb{R}^{|z|} \rightarrow \mathbb{R}^{N \times 3}$  are both MLPs. MLP is a global feature extractor and we do not observe any performance gain by using CNN in practice [8, 14]. The hidden dimensions of the encoder and decoder are set to (512, 256, 128, 64, 32) and (32, 64, 128, 256, 512) respectively. The dimension of the latent space is 8.

## 2.2 Projection module

We use  $\mathcal{P} : \mathbb{R}^{N \times 3} \rightarrow \mathbb{R}^{D \times D}$  to represent a physics-aware projection module which computes the projection from a given orientation. Specifically, the projection module  $\mathcal{P}$  first converts the deformed molecular structure  $\hat{S} \in \mathbb{R}^{N \times 3}$  into a volumetric representation  $\hat{V} \in \mathbb{R}^{D \times D \times D}$ , and then computes the projections  $\hat{I} \in \mathbb{R}^{D \times D}$ . Following cryoFold [13], cryoSTAR uses a Gaussian function to correlate the volumetric density with a coarse-grained atomic model. The volumetric density can be defined as the summation of  $N$  Gaussian blobs, each representing a residue’s electronic density:

$$\hat{V}(\mathbf{x}) = \sum_i^N A_i \exp\left(\frac{-\|\mathbf{x} - \boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right), \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^3$  is a point in the sampling space,  $\boldsymbol{\mu}_i$  is the coordinate of the  $i$ -th residue in the predicted structure  $\hat{S}$ ,  $A_i$  and  $\sigma_i^2$  are the amplitude and the variance of the Gaussian blob. Specifically, the center  $\boldsymbol{\mu}_i$  of each blob is determined by the coordinate of the  $C\alpha$  atom of the amino acid or the P atom of the nucleotide<sup>†</sup>. The amplitude  $A_i$  is determined based on the total electrons of the corresponding amino acid or nucleotide, and the width  $\sigma_i$  is set to 2 empirically [13]. Discussion on how well the Gaussian density resembles the real density can be found in Appendix A.

Given a volumetric density  $\hat{V}$ , the projector further projects it to a 2D image  $\hat{I}$ . The projection process requires an orientation and a contrast transfer function (CTF) determined by the microscope. We assume these information is known, thus the projector is *differentiable* but *non-trainable* (see Appendix B for details).

## 2.3 Structural regularization

For the sake of brevity, we only consider biomolecules with a single, continuous chain with length  $N$  in our formulation and use subscript  $i, j$  to index its residues. We use  $d_{ij}$  and  $\hat{d}_{ij}$  to denote the distance between the  $i$ -th and  $j$ -th residues in the reference structure  $S_{\text{ref}}$  and the predicted structure  $\hat{S}$ , respectively. We keep this convention for all structural regularization definitions. It is worth noting that the structural regularization can be applied to more complex cases such as structures with multiple chains, missing residues, etc.

### 2.3.1 Incorporating basic structure information

The sequence of the target molecule should be unchanged in any conformational dynamics. The continuity loss  $\mathcal{L}_{\text{cont}}$  ensures that the connection between two adjacent residues remains intact, which is defined as:

$$\mathcal{L}_{\text{cont}} = \frac{1}{N-1} \sum_i^{N-1} \|\hat{d}_{i,i+1} - d_{i,i+1}\|^2, \quad (2)$$

This term enforces the two adjacent residues to maintain their connectivity. To prevent two residues from clashing after predicting the deformation, cryoSTAR calculates a clash loss  $\mathcal{L}_{\text{clash}}$  on the pairs of residues that ensure the following condition during training:

$$P_{\text{clash}} = \{(i, j) | 1 \leq i, j \leq N; i \neq j; \hat{d}_{ij} < k_{\text{clash}}\}, \quad (3)$$

<sup>†</sup>If P does not exist, we use C1’ instead.

$$\mathcal{L}_{\text{clash}} = \frac{1}{|P_{\text{clash}}|} \sum_{(i,j) \in P_{\text{clash}}} \|\hat{d}_{ij} - k_{\text{clash}}\|^2, \quad (4)$$

where  $P_{\text{clash}}$  denotes the set of residue pairs that experience collision during training.  $(i, j)$  is an index pair numbering the residues in the structure. Results will be penalized when the predicted distance between two non-adjacent residues is less than the threshold  $k_{\text{clash}}$ , thereby preventing residue clashing from happening. In this study, we set  $k_{\text{clash}}$  to 4Å. For a protein complex with different chains, we consider residues from other chains for the calculation of  $\mathcal{L}_{\text{clash}}$ .

### 2.3.2 Preserving local rigidity with elastic network

CryoSTAR assumes that the local structure should remain rigid. This is a reasonable assumption because: (1) modification of the local structure (secondary structure) is rare in the conformational dynamics captured by cryo-EM, (2) when changes in secondary structure do occur, they are typically resolved through discrete 3D classification methods. CryoSTAR parameterizes the preservation of the local shape and geometry of the backbone model with an elastic network (EN). Specifically, given the initial structure, cryoSTAR builds an elastic network by connecting the residue pairs within a pre-defined distance. The elastic network loss is defined as:

$$P_{\text{EN}} = \{(i, j) | 1 \leq i, j \leq N; i \neq j; d_{ij} < k_{\text{EN}}\}, \quad (5)$$

$$\mathcal{L}_{\text{EN}} = \frac{1}{|P_{\text{EN}}|} \sum_{(i,j) \in P_{\text{EN}}} \|\hat{d}_{ij} - d_{ij}\|^2, \quad (6)$$

where  $P_{\text{EN}}$  is a set of edges for building the elastic network, and the constant value  $k_{\text{EN}}$  is a pre-determined cutoff to find all possible non-covalent interactive pairs of residues. We set  $k_{\text{EN}}$  to 12Å, which is a commonly chosen distance to model such interaction in a coarse-grained model [22]. For a protein complex, we do not build edges between different chains.

### 2.3.3 Adaptive relaxation

A static elastic network may be subject to the bias from the given reference atomic model, while conformational change often involves the forming and breaking of certain non-covalent interactions. For example, using the closed state of adenylate kinase (pdb: *Iake* [23]) as the reference atomic model, the elastic network will cause the “interlock” problem (Figure 1), making it impossible to transition to the open state (pdb: *Aake* [24]).

To identify and mitigate these undesirable interactions, cryoSTAR adaptively selects the edges present in the elastic network for regularization. To be specific, in each mini-batch of size  $b$ , cryoSTAR predicts  $b$  distance values,  $\hat{d}_{ij}$ ’s, for each edge in the elastic network defined in Section 2.3.2, where  $(i, j) \in P_{\text{EN}}$ . Subsequently, we compute the variance of the predicted distances for each edge, denoted by  $\text{Var}(\hat{d}_{ij})$ , over the set of these  $b$  values. We posit that the variance of the edge distance reveals the stability of edges during training, with higher variance indicating a greater likelihood of edge disruptions. All variances corresponding to the edges in  $P_{\text{EN}}$  can be grouped into a set  $\mathcal{V} \triangleq \{\text{Var}(\hat{d}_{ij}) | (i, j) \in P_{\text{EN}}\}$ . In the calculation of the loss, we only retain edges with variance below a certain percentile threshold  $p$ . Thus, only a subset of  $P_{\text{EN}}$  is used for the calculation of loss  $\mathcal{L}_{\text{EN}}$ :

$$P'_{\text{EN}} = \{(i, j) | (i, j) \in P_{\text{EN}}, \text{Var}(\hat{d}_{ij}) < \text{the } p\text{-th percentile of } \mathcal{V}\}. \quad (7)$$

Note that  $P'_{\text{EN}}$  changes for every iteration. The rationale behind this is that the interactions responsible for stabilizing secondary structures should be unchanged, while the interactions involved in the conformational changes are unstable, exhibiting greater variance during training. Eliminating these unwanted constraints encourages the model to explore a wider range of potential conformations supported by the data. This allows cryoSTAR to mitigate the potential bias stemming from the reference atomic model.

## 2.4 Loss function

The final loss function  $\mathcal{L}$  is composed of three parts: (i) the image reconstruction loss  $\mathcal{L}_{\text{image}}$ , (ii) the structural regularization loss  $\mathcal{L}_{\text{struct}}$  from Section 2.3, (iii) a posterior regularization term  $\mathcal{L}_{\text{KL}}$ ,

which encourages the latent code  $z$  to be normally distributed.

$$\mathcal{L} = \mathcal{L}_{\text{image}} + \underbrace{\mathcal{L}_{\text{cont}} + \mathcal{L}_{\text{clash}} + \mathcal{L}_{\text{EN}}}_{\mathcal{L}_{\text{struct}}} + w_{\text{KL}} \mathcal{L}_{\text{KL}}, \quad (8)$$

where  $w_{\text{KL}} = \frac{1}{D^2}$ .

For the image reconstruction loss, cryoSTAR uses cross correlation to measure the similarity between an input image (particle)  $I$  and the predicted projection  $\hat{I}$ :

$$\mathcal{L}_{\text{image}} = -\frac{I \cdot \hat{I}}{\|I\|_F \cdot \|\hat{I}\|_F}, \quad (9)$$

where  $\cdot$  denotes a pixel-wise multiplication, and  $\|I\|_F$  is the Frobenius norm of  $I$ , which is a constant and can be omitted.

### 3 Results

In this section, we first validate the effectiveness of our approach on a synthetic dataset, and then show the result on a real-world experimental dataset.

#### 3.1 Implementations

We implement two baselines for comparison. **CONT**: a method that only enforces backbone continuity like [13, 14]. This is a variant of cryoSTAR where  $\mathcal{L}_{\text{struct}}$  only holds the  $\mathcal{L}_{\text{cont}}$ . **NMA**: A method that predicts the normal mode coefficients of a molecule. Although the orientation (pose) is given, it may be slightly changed by the NMA vectors which only capture the conformational change [19, 15]. We predict an additional rotation matrix and a translation vector to correct the orientation [19]. For normal mode vector computation, we first use ProDy [25] to build the Hessian matrix with the cutoff set to 12Å, then run sparse eigen-decomposition to extract the top 16 eigenvectors on GPU with CuPy [26]. The eigen-decomposition of a  $10^4 \times 10^4$  matrix costs about 3 minutes on a Tesla-V100 GPU.

All these methods are implemented with PyTorch [27]. The batch size is set to 64. The total training step is set to 12,000 for the synthetic data and 96,000 for the experimental data. We set the variance threshold  $p$  (Section 2.3.3) to 80 for the synthetic dataset and 95 for the experimental dataset.

#### 3.2 Validation on a synthetic dataset

**Dataset** Following the setting of Nashed et al. [15], we generate a 50-frame trajectory between the atomic models of the closed (*lake* [23]) and the open (*4ake* [24]) states of adenylyate kinase (AK) using PyMOL [28]. We use the e2pdb2mrc program in the EMAN2 package [3] to compute the density maps from the full-atomic models, and then generate the synthetic particles. See Appendix C for statistics of the simulator (Table 1) and sample images (Figure 6). The closed state model (*lake*) is used as the reference structure.

**Discussion** In the synthetic dataset, the backbone model  $S^{(i)}$  of each particle  $I^{(i)}$  is already known. For each particle, we infer its latent code and generate the corresponding backbone model  $\hat{S}^{(i)}$ . We compute the root mean squared distance (RMSD) of  $C\alpha$ 's between  $S^{(i)}$  and  $\hat{S}^{(i)}$  to measure the accuracy of the deformation prediction network. Note that we do **not** align two structures before computing the

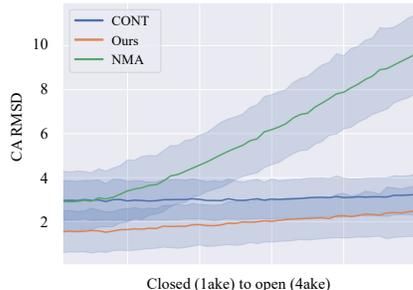


Figure 3: The  $C\alpha$ -RMSD at different conformational states (left side: *lake*, right side: *4ake*). Each state has several particle images, therefore for each state a distribution of RMSDs can be computed. We visualize the error band for cryoSTAR and the two baselines. The middle curve in every band denotes the mean and the light blue regions denote one sigma deviation. Our method achieves the lowest RMSD across all states.

RMSD, because the orientation is already given. Figure 3 shows that cryoSTAR has a consistently lower RMSD than the two baselines at different conformational states. NMA has a larger RMSD at the open state (7.14Å higher than cryoSTAR), suggesting the bias of using the closed state model as the reference structure.

### 3.3 Application on an experimental dataset

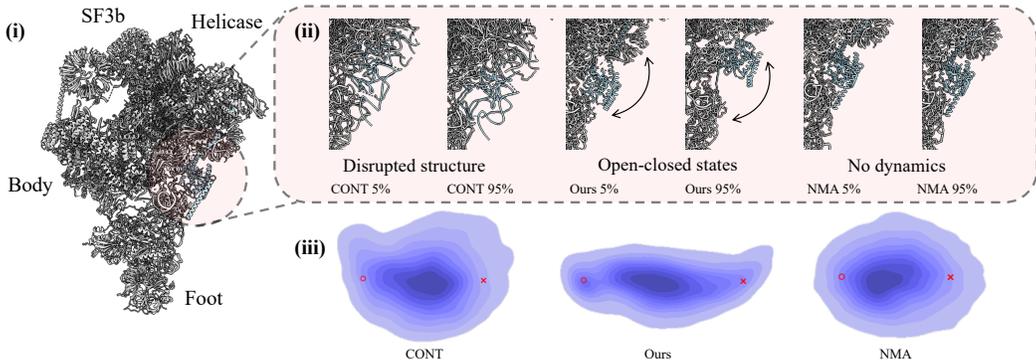


Figure 4: Results of cryoSTAR and the two baselines on the pre-catalytic spliceosome dataset (EMPIAR-10180). (i) The reference structure (pdb: *5nrl*) of the pre-catalytic spliceosome. The four domains (SF3b, helicase, the body and the foot) are labeled in the corresponding places. (ii) Reconstructed coarse-grained backbone models by the three methods, each sampled at the 5-th and the 95-th percentile of the first PCA dimension. CONT fails to preserve local structures (such as colored helices), and NMA fails to find the open state. CryoSTAR succeeds in finding dynamics and preserving local structures simultaneously. (iii) The latent space of the three methods, where color depth represents the population of each conformation state  $z$ , and red symbols  $\circ$  and  $\times$  denote the 5-th and 95-th percentile of the first PCA-dimension, respectively. Both CONT and NMA find a single mode, while cryoSTAR finds two modes in the latent space.

**Dataset** We applied cryoSTAR to the experimental pre-catalytic spliceosome dataset (EMPIAR-10180) [21]. This dataset contains preprocessed particles ready for 3D reconstruction, and has been widely used as a benchmark to solve the continuous heterogeneity problem in cryo-EM [8, 29, 30]. We resize the particles to  $128 \times 128$  with low-pass filtering. The atomic model built from the cryo-EM density map (pdb: *5nrl*) in the original paper [21] is used as the reference atomic model.

**Discussion** The pre-catalytic spliceosome can be divided into four domains: SF3b, helicase, the body and the foot (Figure 4 (i)). Although a ground truth of the dynamics is not available, previous method resolved an up-down movement of the SF3b and the helicase domains on the volumetric density [8], which was also supported by the hypothesis in the original paper [21]. We sampled two states from the two ends of the first PCA dimension of the latent code to compare the results of cryoSTAR with the two baselines (Figure 4 (ii, iii)). With only  $\mathcal{L}_{\text{cont}}$  as the regularization (CONT), the output backbone models for both states are chaotic, where the structured regions (e.g.,  $\alpha$ -helices) are totally destroyed. On the other hand, NMA preserves the local structures, but the difference between the two states is minimum, suggesting that it can only recover small movements that resemble fluctuations rather than capturing large conformational dynamics. In comparison, cryoSTAR finds large up-down motions of the SF3b and the helicase domains. Moreover, cryoSTAR preserves the local secondary structures without obvious distortions.

## 4 Conclusions and limitations

In this work, we study atomic level heterogeneity analysis through a structural regularization approach, and propose an adaptive approach to preserve local structures and avoid local minima simultaneously. Experiments on the two datasets validate the effectiveness of our method.

The study has one major limitation: the expressive power of isotropic Gaussian functions. Isotropic Gaussian functions are widely used in related exploration for modeling simplicity and computational

efficiency [12, 13, 14]. However, it is prone to spherical artifacts, which may impair the accuracy of likelihood estimation. Some recent work in computer graphics [31] replaces it with an anisotropic one, and it may unlock the potential of Gaussian functions. We will leave it for future work.

## 5 Acknowledgements

We thank Dr. Zaixiang Zheng, Dr. Yiqun Wang, and Dr. Dongyu Xue for their insightful discussions on the project. We would like to thank Dr. Hang Li for the feedback on the project that helped shape this study.

## References

- [1] “Method of the Year 2015”. In: *Nature Methods* 13.1 (2016), pp. 1–1. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.3730.
- [2] Sjors H.W. Scheres. “RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination”. In: *Journal of Structural Biology* 180.3 (2012), pp. 519–530. ISSN: 10478477. DOI: 10.1016/j.jsb.2012.09.006.
- [3] Guang Tang et al. “EMAN2: An Extensible Image Processing Suite for Electron Microscopy”. In: *Journal of Structural Biology* 157.1 (2007), pp. 38–46. ISSN: 10478477. DOI: 10.1016/j.jsb.2006.05.009.
- [4] Ali Punjani et al. “cryoSPARC: Algorithms for Rapid Unsupervised Cryo-EM Structure Determination”. In: *Nature Methods* 14.3 (2017), pp. 290–296. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.4169.
- [5] Sjors H.W. Scheres. “A Bayesian View on Cryo-EM Structure Determination”. In: *Journal of Molecular Biology* 415.2 (2012), pp. 406–418. ISSN: 00222836. DOI: 10.1016/j.jmb.2011.11.010.
- [6] Ali Punjani and David J. Fleet. “3D Variability Analysis: Resolving Continuous Flexibility and Discrete Heterogeneity from Single Particle Cryo-EM”. In: *Journal of Structural Biology* 213.2 (2021), p. 107702. ISSN: 10478477. DOI: 10.1016/j.jsb.2021.107702.
- [7] Ellen D. Zhong et al. “Reconstructing Continuous Distributions of 3D Protein Structure from CryoEM Images”. In: *International Conference on Learning Representations*. 2020, p. 20.
- [8] Ellen D. Zhong et al. “CryoDRGN: Reconstruction of Heterogeneous Cryo-EM Structures Using Neural Networks”. In: *Nature Methods* 18.2 (2021), pp. 176–185. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-020-01049-4.
- [9] Ellen D. Zhong et al. “CryoDRGN2: Ab Initio Neural Reconstruction of 3D Protein Structures from Real Cryo-EM Images”. In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, 2021, pp. 4046–4055. ISBN: 978-1-66542-812-5. DOI: 10.1109/ICCV48922.2021.00403.
- [10] Diederik P Kingma and Max Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013).
- [11] Ali Punjani and David J. Fleet. “3DFlex: Determining Structure and Motion of Flexible Proteins from Cryo-EM”. In: *Nature Methods* 20.6 (2023), pp. 860–870. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-023-01853-8.
- [12] Muyuan Chen and Steven J. Ludtke. “Deep Learning-Based Mixed-Dimensional Gaussian Mixture Model for Characterizing Variability in Cryo-EM”. In: *Nature Methods* 18.8 (2021), pp. 930–936. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-021-01220-5.
- [13] Ellen D. Zhong et al. “Exploring Generative Atomic Models in Cryo-EM Reconstruction”. In: *NeurIPS Workshop on Machine Learning for Structural Biology*. arXiv, 2021. eprint: 2107.01331 (q-bio).
- [14] Dan Rosenbaum et al. “Inferring a Continuous Distribution of Atom Coordinates from Cryo-EM Images Using VAEs”. In: *NeurIPS Workshop on Machine Learning for Structural Biology*. 2021. eprint: 2106.14108 (cs, eess).
- [15] Youssef Nashed et al. “Heterogeneous Reconstruction of Deformable Atomic Models in Cryo-EM”. In: *NeurIPS Workshop on Machine Learning for Structural Biology*. arXiv, 2022. eprint: 2209.15121 (physics, q-bio).

- [16] Ilyes Hamitouche and Slavica Jonic. “DeepHEMNMA: ResNet-based Hybrid Analysis of Continuous Conformational Heterogeneity in Cryo-EM Single Particle Images”. In: *Frontiers in Molecular Biosciences* 9 (2022), p. 965645. ISSN: 2296-889X. DOI: 10.3389/fmolb.2022.965645.
- [17] Geoffrey Woollard et al. “Physics Aware Inference for the Cryo-EM Inverse Problem: Anisotropic Network Model Heterogeneity, Global Pose and Microscope Defocus”. In: *NeurIPS Workshop on Machine Learning for Structural Biology*. 2022.
- [18] Jianpeng Ma. “Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes”. In: *Structure* 13.3 (2005), pp. 373–380. ISSN: 09692126. DOI: 10.1016/j.str.2005.02.002.
- [19] Bongjin Koo et al. “CryoChains: Heterogeneous Reconstruction of Molecular Assembly of Semi-flexible Chains from Cryo-EM Images”. In: *ICML Workshop on Computational Biology*. arXiv, 2023. eprint: 2306.07274 (cs, q-bio).
- [20] Monique M. Tirion. “Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis”. In: *Physical Review Letters* 77.9 (1996), pp. 1905–1908. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.77.1905.
- [21] Clemens Plaschka, Pei-Chun Lin, and Kiyoshi Nagai. “Structure of a Pre-Catalytic Spliceosome”. In: *Nature* 546.7660 (2017), pp. 617–621. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature22799.
- [22] Magnus Ekeberg et al. “Improved Contact Prediction in Proteins: Using Pseudolikelihoods to Infer Potts Models”. In: *Physical Review E* 87.1 (2013), pp. 1–19. ISSN: 15393755. DOI: 10.1103/PhysRevE.87.012707. eprint: 1211.1281.
- [23] Christoph W. Müller and Georg E. Schulz. “Structure of the Complex between Adenylate Kinase from Escherichia Coli and the Inhibitor Ap5A Refined at 1.9 Å Resolution”. In: *Journal of Molecular Biology* 224.1 (1992), pp. 159–177. ISSN: 00222836. DOI: 10.1016/0022-2836(92)90582-5.
- [24] CW Müller et al. “Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding”. In: *Structure* 4.2 (Feb. 1996), pp. 147–156. DOI: 10.1016/S0969-2126(96)00018-4. URL: [https://doi.org/10.1016/S0969-2126\(96\)00018-4](https://doi.org/10.1016/S0969-2126(96)00018-4).
- [25] Ahmet Bakan, Lidio M. Meireles, and Ivet Bahar. “ProDy : Protein Dynamics Inferred from Theory and Experiments”. In: *Bioinformatics* 27.11 (2011), pp. 1575–1577. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btr168.
- [26] Ryosuke Okuta et al. “CuPy: A NumPy-Compatible Library for NVIDIA GPU Calculations”. In: *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*. 2017.
- [27] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. In: *Advances in Neural Information Processing Systems* 32. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 8024–8035.
- [28] Schrödinger, LLC. “The PyMOL Molecular Graphics System, Version 1.8”. Nov. 2015.
- [29] Takanori Nakane et al. “Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION”. In: *elife* 7 (2018), e36861.
- [30] D. Herreros et al. “Estimating Conformational Landscapes from Cryo-EM Particles by 3D Zernike Polynomials”. In: *Nature Communications* 14.1 (2023), p. 154. ISSN: 2041-1723. DOI: 10.1038/s41467-023-35791-y.
- [31] Bernhard Kerbl et al. “3D Gaussian Splatting for Real-Time Radiance Field Rendering”. In: *ACM Transactions on Graphics* 42.4 (2023), pp. 1–14. ISSN: 0730-0301, 1557-7368. DOI: 10.1145/3592433.
- [32] Edward H. Egelman. “The Current Revolution in Cryo-EM”. In: *Biophysical Journal* 110.5 (2016), pp. 1008–1012. ISSN: 00063495. DOI: 10.1016/j.bpj.2016.02.001.

## Appendix

### A Quality of Gaussian density

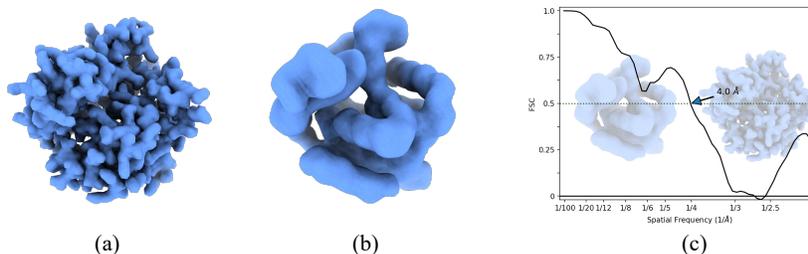


Figure 5: (a) The simulated *IakeA* full-atom density volume by EMAN2 [3]. (b) The CA atom based Gaussian density. (c) Determination of the low-pass cutoff frequency.

As illustrated in Figure 5, the Gaussian density is not equivalent to the density map reconstructed from the by particle images. To address this distortion, we calculate the Fourier shell correlation (FSC) between the Gaussian density and consensus density map, and low-pass filter the Gaussian density to the frequency where  $FSC = 0.5$ . In other words, cryoSTAR only uses low to intermediate resolution information to help resolve the continuous heterogeneity in cryo-EM data. This is also the reason that a more fine-grained model (e.g., a full-atomic model) is not used in this study.

### B Projection module’s image formation

Cryo-EM images are the 2D projections  $\{I^{(i)} \in \mathbb{R}^{D \times D} | i = 1, 2, \dots, M\}$  of the molecules’ density maps  $V \in \mathbb{R}^{D \times D \times D}$ . Numerous molecules of the same kind with different orientations are instantly frozen and imaged with electrons, and their 2D projections are recorded. Here we introduce the image formation of the projection module.

Briefly speaking, for a single molecule, its 3D density  $V$  is first rotated ( $R : \mathbb{R}^{D \times D \times D} \rightarrow \mathbb{R}^{D \times D \times D}$ ) and then projected ( $\Pi_{2D} : \mathbb{R}^{D \times D \times D} \rightarrow \mathbb{R}^{D \times D}$ ) on the 2D plane. An in-plane translation  $t : \mathbb{R}^{D \times D} \rightarrow \mathbb{R}^{D \times D}$  is applied to center the projected image. Finally the image is corrupted by a physics-aware noising operator CTF :  $\mathbb{R}^{D \times D} \rightarrow \mathbb{R}^{D \times D}$ .

$$\hat{I}^{(i)} = (\text{CTF}^{(i)} \circ t^{(i)} \circ \Pi_{2D} \circ R^{(i)})(\hat{V}^{(i)}), \quad (10)$$

where  $R$ ,  $t$  and CTF are **differentiable** and their parameters are **instance-wise** and **given**,  $\Pi_{2D}$  is **non-parametric**. Further details are listed as below:

- $R$  and  $t$  are the so-called “pose” in cryo-EM.  $R$  denotes the orientation of the molecule.  $t$  is required since the 2D projections are cropped from a large micrograph in standard cryo-EM pipeline, it is not well centered.
- $\Pi_{2D}$  is a projection operator that takes integral along the first dimension of a density.
- CTF is a noising operator named contrast transfer function that describes how aberrations in a transmission electron microscope (TEM) and the imaging defocus modify the image. Note that CTF is defined in the Fourier space, and the corresponding version in the real space is point spread function (PSF). We use CTF throughout this paper for simplicity.
- Additionally, the image is corrupted with noise stemming from the stochastic nature of electron detection events and sensor failures [32]. We omit this in our image formation model.

### C Synthetic dataset generation

In order to measure the accuracy of our methods, we generate a synthetic dataset for evaluation. A realistic simulator is equipped in the image formation process as described in Appendix B. We use Gaussian additive noise in the formation process. Each states from *Iake* to *4ake* generates 1,000

particle images. The parameter settings of synthetic dataset are shown in Table 1. An example of synthetic particle images are shown in Figure 6.

Parameters	Value
Number of particles	50,000
Image size	128 px $\times$ 128 px
Image pixel size	1.0 Å
Defocus	Lognormal(1.0, 0.3 <sup>2</sup> ) $\mu$ m
Accelerating voltage	300 keV
Spherical aberration (cs)	2.7 nm
Amplitude contrast ratio	0.1
SNR (with Gaussian noise)	0.0001
Pose/rotation	Uniform on $\mathbb{SO}(3)$
Pose/translation	0

Table 1: Statistics of the synthetic dataset.

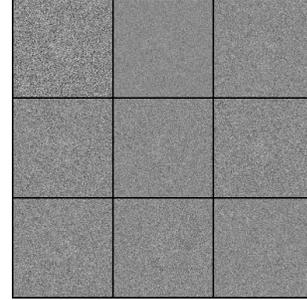


Figure 6: Examples of the particles from the synthetic dataset.