

---

# CryoChains: Heterogeneous Reconstruction of Molecular Assembly of Semi-flexible Chains from Cryo-EM Images

---

Bongjin Koo<sup>1</sup> Julien Martel<sup>2</sup> Ariana Peck<sup>3</sup> Axel Levy<sup>2,3</sup> Frédéric Poitevin<sup>3</sup> Nina Miolane<sup>1</sup>

## Abstract

Cryogenic electron microscopy (cryo-EM) has transformed structural biology by allowing to reconstruct 3D biomolecular structures up to near-atomic resolution. However, the 3D reconstruction process remains challenging, as the 3D structures may exhibit substantial shape variations, while the 2D image acquisition suffers from a low signal-to-noise ratio, requiring to acquire very large datasets that are time-consuming to process. Current reconstruction methods are precise but computationally expensive, or faster but lack a physically-plausible model of large molecular shape variations. To fill this gap, we propose CryoChains that encodes large deformations of biomolecules via rigid body transformation of their chains, while representing their finer shape variations with the normal mode analysis framework of biophysics. Our synthetic data experiments on the human GABA<sub>B</sub> and heat shock protein show that CryoChains gives a biophysically-grounded quantification of the heterogeneous conformations of biomolecules, while reconstructing their 3D molecular structures at an improved resolution compared to the current fastest, interpretable deep learning method.

## 1. Introduction

Cryogenic electron microscopy (cryo-EM) allows to reconstruct 3D biomolecular structures at near-atomic resolution. For this, estimation techniques process large datasets of 2D images obtained from many copies of the same biomolecule in different shapes, called *conformations*. The reconstruction task is challenging due to the intrinsic variability in a biomolecule conformations (*conformation heterogeneity*),

the high number of unknown, nuisance variables (e.g., the biomolecule’s position/orientation in 3D), and extremely low signal-to-noise ratios (SNRs). Thus, reconstruction algorithms require large amounts of data, posing major computational challenges (Kimanius et al., 2016).

**Related works** Recent efforts have turned to deep learning (DL) for cryo-EM reconstruction, reaching significant speed-ups through gradient methods and GPU acceleration (Donnat et al., 2022). However, these approaches have been limited in their parameterization of the biomolecular volumes’ variability. On the one hand, *homogeneous reconstruction* methods disregard the conformation heterogeneity entirely, as they estimate the average 3D biomolecular volume from the images (Nashed et al., 2021; Miolane et al., 2019; Gupta et al., 2021), which limits the accuracy of the 3D reconstruction as an average biomolecule appears “blurry”. On the other hand, *heterogeneous reconstruction* methods explicitly model structural variability (Zhong et al., 2021; 2020; Gupta et al., 2020; Ullrich et al., 2020; Chen & Ludtke, 2021; Rosenbaum et al., 2021; Punjani & Fleet, 2022) but rarely leverage domain knowledge to encode biophysically-plausible molecular deformations, which limits the interpretation of the conformations.

In biophysics, normal mode analysis (NMA) is a dimensionality reduction method, describing the vibrational modes of a biomolecular structure, i.e., the principal directions of its conformational variations (Levitt et al., 1985). Thus, NMA has naturally emerged as the first biophysically-grounded approach to efficiently parameterize the biomolecular volumes and their variability for cryo-EM reconstruction. HEM-NMA (Jin et al., 2014) uses normal modes to deform a reference volume  $\mathbf{V}^{(0)}$  that is fit to cryo-EM images by an iterative 3D-to-2D alignment. DeepHEMNMA (Hamitouche & Jonic, 2022) extends HEMNMA by adopting a residual network that predicts NMA weights (the contribution of each normal mode) and biomolecular poses given by HEMNMA in a supervised learning setting. By contrast, (Nashed et al., 2022) performs unsupervised heterogeneous reconstruction, inferring NMA weights as the latent variables of an autoencoder. However, NMA can only represent small conformation changes around a reference conformation. This poses a fundamental limit as biomolecular vari-

<sup>1</sup>Dept. of Electrical & Computer Engineering, UC Santa Barbara, CA, USA <sup>2</sup>Stanford University, CA, USA <sup>3</sup>LCLS, SLAC, Menlo Park, CA, USA. Correspondence to: Bongjin Koo <bongjinkoo@ucsb.edu>.

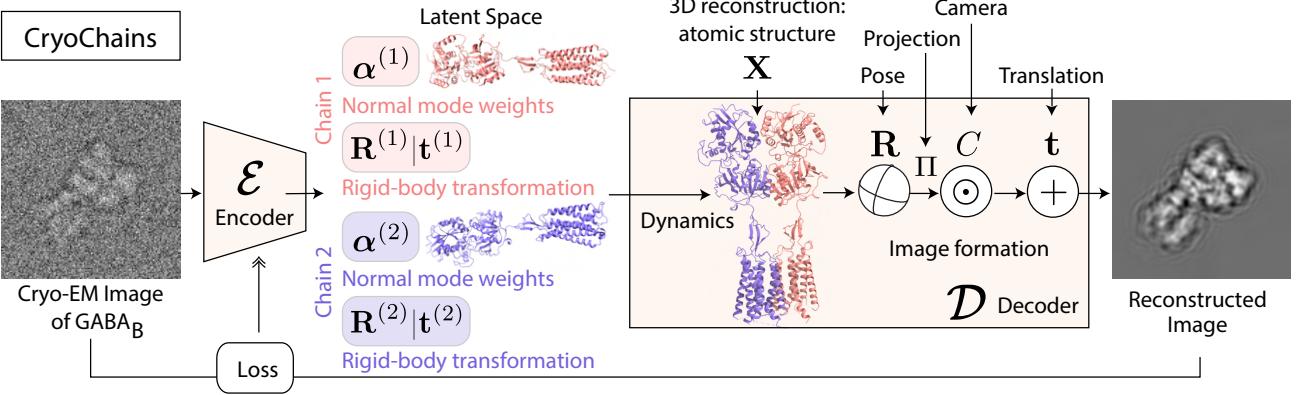


Figure 1: CryoChains autoencoder illustrated on the GABA<sub>B</sub> protein with two chains in red and blue. The encoder  $\mathcal{E}$  embeds the input image into a latent space with two sets of variables: (i)  $\alpha^{(c)}$ , the normal mode weights per chain, which deform each chain; and (ii)  $R^{(c)}|t^{(c)}$ , the rigid body transformation that poses each (deformed) chain within the global atomic structure  $\mathbf{X}$  of the protein. The dynamics model of the decoder  $\mathcal{D}$  applies the NMA deformations and rigid body transformations to each chain of a biomolecule, which is fed into the image formation model composed of a global rotation  $\mathbf{R}$  (“pose”), projection to 2D space ( $\Pi$ ), camera parameters ( $C$ ) and a global 2D translation  $\mathbf{t}$ , to reconstruct the input image. CryoChains is trained using a reconstruction loss, which is the pixel-wise mean squared error between the input and reconstructed images.

ability exhibits larger scale deformations. For example, the rotations of a biomolecule’s chains w.r.t. one another are typically larger deformations that are not captured by NMA.

**Contributions** We propose a new modular, biophysically-grounded method to parameterize the molecular volumes’ variability for accurate, interpretable heterogeneous reconstruction in cryo-EM. We decompose molecules into their *chains* and represent large scale molecular deformations by the rotation and translation of chains w.r.t. another. Then, we represent the smaller scale deformations per chain by NMA. Building on (Nashed et al., 2022), we propose a new unsupervised DL architecture *CryoChains*, which extracts the rotation and translation, with the normal mode weights, of each chain in the latent space of an autoencoder (Fig. 1). CryoChains achieves higher reconstruction accuracy on synthetic cryo-EM images with realistic noise levels, compared to the existing method using NMA only. We also show that CryoChains provides a biophysical interpretation of the conformation heterogeneity via the exploration of its latent space.

## 2. CryoChains Methods

CryoChains adopts an autoencoder architecture shown in Fig. 1. The encoder  $\mathcal{E}$  is a convolutional network that embeds a given image in a latent space composed of several conformation latent variables. The decoder  $\mathcal{D}$  implements the generative model of cryo-EM images composed of the dynamics and image formation. This section and Appendix C detail  $\mathcal{E}$  and  $\mathcal{D}$ . The pixel-wise mean squared

error (MSE) between the input and the reconstructed images is minimized during training. Only the encoder is trained, i.e., the decoder is pre-specified and fixed.

### 2.1. Encoder and Latent Space

The encoder  $\mathcal{E}$  first extracts the small, elastic conformational variations of the biomolecule via a latent variable  $\alpha^{(c)}$  for each chain  $c$ —see Fig. 1. The variable  $\alpha^{(c)} \in \mathbb{R}^{K^{(c)}}$  represents the chain’s deformation along its normal modes:

$$\mathbf{X}^{(c)}(\alpha) = \mathbf{X}^{(c,0)} + \sum_{k=1}^{K^{(c)}} \alpha_k^{(c)} \mathbf{U}_k^{(c)}. \quad (1)$$

Here,  $\mathbf{X}^{(c)}(\alpha) \in \mathbb{R}^{3n_c}$  is the atomic structure of chain  $c$  as the 3D coordinates of its  $n_c$  constitutive atoms. NMA defines this atomic structure as the deformation of a reference structure  $\mathbf{X}^{(c,0)} \in \mathbb{R}^{3n_c}$  through  $K^{(c)}$  normal modes  $\mathbf{U}_1^{(c)}, \dots, \mathbf{U}_{K^{(c)}}^{(c)}$  which are computed beforehand.

Importantly, the weights  $\alpha^{(c)}$  along the normal modes have a natural biophysical interpretation. As normal modes are the eigenvectors of the Hessian matrix of the atoms’ potential energy, obtained from a second order Taylor approximation around a conformation at equilibrium, the weights encode in which direction and how far each atom moves w.r.t. its position in the reference conformation—see (Levitt et al., 1985; Goldstein et al., 2008) and Appendix A for details. The normal modes are sorted by the eigenvalues of the Hessian matrix in an increasing order, where the square-root of each eigenvalue is the frequency that deforms the refer-

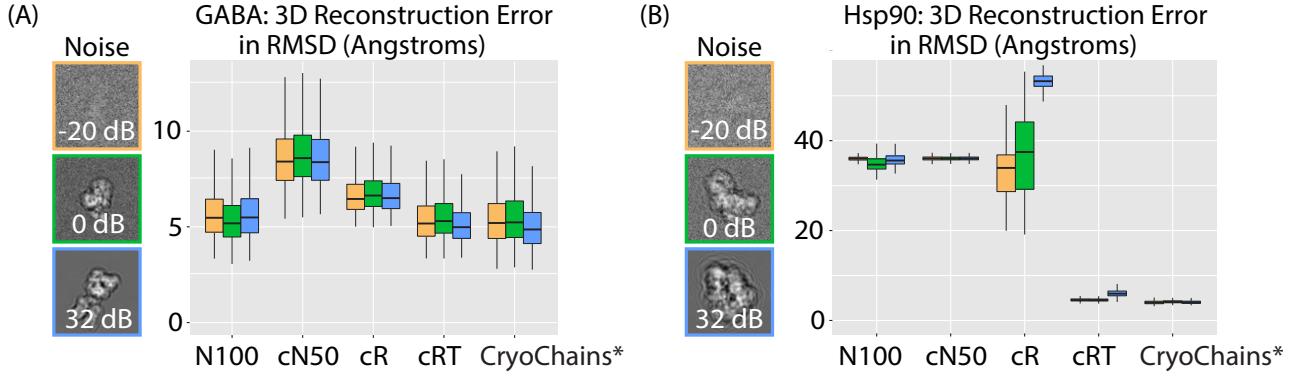


Figure 2: Comparison of 3D reconstruction accuracy between CryoChains, its ablation models ( $cN50$ ,  $cR$ ,  $cRT$ ), and the baseline  $N100$  (Nashed et al., 2022). Reconstruction error is measured as the root-mean-square deviation (RMSD) in Angstroms, between the ground truth (GT) and predicted atom coordinates on  $GABA_B$  and  $Hsp90$ , for 3 noise levels  $\{32, 0, -20\}$  in dB shown in 3 colors. The models differ in their representation of conformation heterogeneity in the latent space:  $N100$  uses whole protein NMA with 100 normal modes,  $cN50$ : per-chain NMA with 50 normal modes,  $cR$ : per-chain rotation,  $cRT$ : per-chain rigid body transformation, CryoChains: ours, i.e., per-chain rigid body transformation and per-chain NMA with 50 normal modes.

ence conformation  $\mathbf{X}^{(0)}$  along each normal mode  $\mathbf{U}_k$ . Thus, the first few normal modes describe low-frequency (large) conformation changes, in which we are mainly interested. The number of normal modes  $K^{(c)}$  per chain is chosen to be the minimal number of normal modes that could account for the majority of the deformations. We highlight that we use NMA on each chain independently, as opposed to NMA on the whole biomolecule as proposed in (Nashed et al., 2022).

$\mathcal{E}$  also extracts the larger, rigid body transformations that arise when chains of a biomolecule rotate and translate w.r.t. each other. We parameterize the orientation of each chain as a rotation around its center-of-mass (CoM) to which we add a translation.  $\mathcal{E}$  extracts two 3D vectors that are orthonormalized into  $\mathbf{w}_1$  and  $\mathbf{w}_2$ , from which the rotation matrix  $\mathbf{R}^{(c)}$  is built per chain. Translation  $\mathbf{t}^{(c)}$  is a 3D vector per chain. We highlight that these per-chain rotation  $\mathbf{R}^{(c)}$  and translation  $\mathbf{t}^{(c)}$  differ from the global rotation  $\mathbf{R}$  and translation  $\mathbf{t}$  of the whole biomolecule within the decoder.

## 2.2. Decoder

The dynamics component of the decoder  $\mathcal{D}$  takes the conformation latent variables  $\{\boldsymbol{\alpha}^{(c)}, \mathbf{R}^{(c)}, \mathbf{t}^{(c)}\}_{c=1}^{N_C}$  to output the 3D atomic structure  $\mathbf{X}$  of the biomolecule in the input 2D image—see Fig. 1. The structure  $\mathbf{X}^{(c)}$  of each chain  $c$  is given by Eq. (1) and  $\mathbf{X}$  is given by  $\mathbf{X} = [\mathbf{R}^{(c)}(\mathbf{X}^{(c)}(\boldsymbol{\alpha}^{(c)})) + \mathbf{t}^{(c)}]_{c=1}^{N_C}$ , where  $[ ]$  means concatenation.

The image formation component of the decoder  $\mathcal{D}$  generates cryo-EM images using the atomic model  $\mathbf{X}$  given by the dynamics. It is a feed-forward differentiable simulator without learnable components that successively creates a electrostatic potential  $\mathbf{V}$  from  $\mathbf{X}$ , orients it in 3D space us-

ing a global rotation  $\mathbf{R}$ , then acquires a 2D image through a camera model with parameter  $C$ , and finally adds a 2D translation  $\mathbf{t}$ . See Appendix B for details.

## 3. Experiments and Results

We show that CryoChains improves 3D reconstruction accuracy while providing an interpretable parameterization of the conformation changes in its latent space.

### 3.1. Reconstruction Accuracy

**Experiment** We consider a protein with two conformations  $\mathbf{X}^{(0,GT)}$  and  $\mathbf{X}^{(0)}$ , loaded from RCSB PDB (Berman et al., 2000).  $\mathbf{X}^{(0,GT)} \in \mathbb{R}^{3N}$  is a ground truth (GT) reference conformation with total number of atoms  $N$ .  $\mathbf{X}^{(0)} \in \mathbb{R}^{3N}$  is a source reference conformation used by CryoChains to be fit to the images generated with  $\mathbf{X}^{(0,GT)}$ . We generate a realistic distribution of heterogeneous conformations of this protein using the reference conformations  $\mathbf{X}^{(c,0,GT)}$  of its chains, Eq. (1) and the following approach simulating normal mode weights per chain. Each weight is computed as  $\boldsymbol{\alpha}_k^{(c)} = \sqrt{\frac{N}{K}}d$  where  $d$  is sampled from the Gaussian mixture model:  $d \sim \sum_{m=0}^1 \phi_m \mathcal{N}(\mu_m, \sigma_m^2)$  such that  $\sum_m \phi_m = 1$  and  $m \sim \text{Bernoulli}(0.5)$ . After each  $\mathbf{X}^{(c,GT)}$  is obtained via Eq. (1), each chain  $c$  is rotated randomly along each axis in 3D around the chain’s CoM. The final images are rendered using the image formation model of the decoder.

In practice, we use the human  $GABA_B$  receptor membrane protein and the  $Hsp90$  (heat shock protein of 90 kDa) PDB

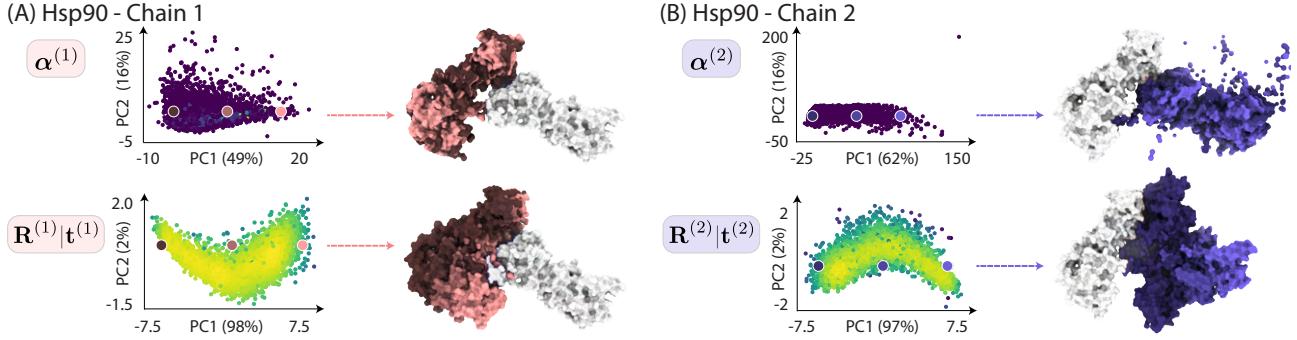


Figure 3: Interpretability of CryoChains’s biophysically-grounded latent space for chains 1 (A) and 2 (B) of the heat shock protein Hsp90. Left columns in (A) and (B) show the latent space of cryo-EM images projected on the first two principal components (PCs) of PCA, for normal mode weights ( $\alpha$ ) and rigid transformation ( $R|t$ ) per chain. Color represents density (yellow: high density) and explained variance per PC in %. Right columns show the reconstruction of 3 conformations in 3 shades of red in (A) and blue in (B), sampled along the first PC shown in the corresponding PCA plots. CryoChains’s latent space provides an interpretable representation of conformation heterogeneity, e.g., between open  $\leftrightarrow$  closed conformations through the rigid transformation.

files. Both proteins have two chains. Fig. 2 shows examples of images generated for each protein. The PDB files of each protein contains two conformations, closed and open, 6UO8/6UOA for GABA<sub>B</sub> and 2CG9/2CGE for Hsp90, which play the roles of  $\mathbf{X}^{(0,GT)}$  and  $\mathbf{X}^{(0)}$  respectively. Concretely,  $\mathbf{X}^{(0,GT)} = 6UO8/2CG9$  is used to generate synthetic images using 15 normal modes where  $d$  is sampled from a GMM with means  $\{0, 2.5\}$  and std 0.25. The rotation angles are 5° per axis for 6UO8 and 2.5° for 2CG9. During training, we deform  $\mathbf{X}^{(0)} = 6UOA/2CGE$  to fit 6UO8/2CG9 in the images, i.e., fit the open conformation to the closed one. This setup simulates a real case scenario where we obtain new cryo-EM images, without the GT conformation, and want to reconstruct the 3D structure using an existing PDB file. Finally, three different levels of noise (SNRs) are added to the images: {32 dB (no noise), 0 dB, -20 dB (close to experimental data)} following (Levy et al., 2022). For each noise level, 50,000 images are generated for the training set, and 5,000 images for validation and test sets.

**Results** We compare the reconstruction accuracy of CryoChains with (Nashed et al., 2022), the fast, yet biophysically-grounded DL-powered unsupervised heterogeneous reconstruction method in cryo-EM. We also compare CryoChains with its ablation models. We report the root-mean-square deviation (RMSD) between the predicted and the GT atom coordinates on GABA<sub>B</sub> and Hsp90 datasets for each model in Fig. 2. We highlight that N100 refers to (Nashed et al., 2022) using 100 normal modes computed on the whole protein. For both proteins, CryoChains achieves the lowest RMSD, which demonstrates better reconstruction capability. Our ablation study reveals that CryoChains outperforms N100, even when it only relies on per-chain rigid body transformation (cRT). This suggests that introducing

rigid transformation of chains is crucial to unlock higher reconstruction accuracy in cryo-EM—which is precisely the purpose of this paper. We also show the resulting high-quality 3D reconstructions of our model in Appendix D. We note that the RMSDs of N100 and CryoChains on GABA<sub>B</sub> in Fig. 2(A) are comparable because GABA<sub>B</sub>’s conformational changes existing in the dataset do not involve large rigid body transformation of the chains and thus can be explained well by whole protein NMA only. By contrast, CryoChains significantly improves the RMSDs on Hsp90, because the conformational changes of this protein involve large scale rigid body transformations that can only be captured by our proposed method. CryoChains will indeed provide researchers with maximum benefits in this specific experimental context.

### 3.2. Conformational Heterogeneity

**Experiment** We perform principal component analysis (PCA) on the latent space of CryoChains trained on images of Hsp90. We demonstrate that CryoChains yields interpretable representations of the protein’s conformational heterogeneity. We generate 50 Hsp90 conformations by morphing atom coordinates between Hsp90’s two conformations (open and closed) using ChimeraX (Pettersen et al., 2021). Then, 50,000 training and 5,000 test images are generated with the noise -20 dB. We train CryoChains using the open conformation as its source reference  $\mathbf{X}^{(0)}$ . We extract the latent variables corresponding to the test images. PCA is performed on the latent space of NMA weights ( $\alpha$ ) and rigid transformation ( $R|t$ ) per chain.

**Results** In Fig. 3, the left columns of (A) and (B) show the first two principal components of the latent space for

NMA weights (top) and rigid transformation (bottom). The right column depicts 3 conformations, reconstructed from 3 latent points sampled along the first principal component (PC1). As expected, we observe that the reconstructions from the rigid transformations' principal component vary between the open and closed conformation. This demonstrates CryoChains's ability to encode the protein's conformation changes into an interpretable latent space.

## 4. Discussions and Conclusion

We introduced CryoChains to address outstanding challenges in cryo-EM reconstruction. Our experiments on synthetic datasets show that CryoChains is a deep-learning powered approach that improves the 3D reconstruction accuracy of biomolecular structures, while being biophysically interpretable. We hope to open new avenues of research for high-resolution reconstruction of proteins in solution.

## References

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The protein data bank. *Nucleic Acids Research*, 28(1): 235–242, 2000.
- Chen, M. and Ludtke, S. J. Deep learning-based mixed-dimensional gaussian mixture model for characterizing variability in cryo-em. *Nature Methods*, 18(8):930–936, Jul 2021.
- Donnat, C., Levy, A., Poitevin, F., Zhong, E. D., and Miolane, N. Deep generative modeling for volume reconstruction in cryo-electron microscopy. *Journal of Structural Biology*, 214(4):107920, December 2022.
- Goldstein, H., Poole, C. P., and Safko, J. L. *Classical mechanics*. Addison Wesley, San Francisco Munich, 3. ed., [nachdr.] edition, 2008. ISBN 978-0-201-65702-9.
- Gupta, H., Phan, T. H., Yoo, J., and Unser, M. Multi-cryogan: Reconstruction of continuous conformations in cryo-em using generative adversarial networks. In *European Conference on Computer Vision*, pp. 429–444. Springer, 2020.
- Gupta, H., McCann, M. T., Donati, L., and Unser, M. Cryogan: a new reconstruction paradigm for single-particle cryo-em via deep adversarial learning. *IEEE Transactions on Computational Imaging*, 7:759–774, 2021.
- Hamitouche, I. and Jonic, S. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Frontiers in Molecular Biosciences*, 9:965645, September 2022.
- Jin, Q., S. Sorzano, C. O., de La Rosa-Trevín, J. M., Bilbao-Castro, J. R., Nunez-Ramirez, R., Llorca, O., Tama, F., and Jonic, S. Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. *Structure*, 22(3):496–506, 2014.
- Kimanis, D., Forsberg, B. O., Scheres, S. H., and Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2. *eLife*, 2016.
- Levitt, M., Sander, C., and Stern, P. S. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of molecular biology*, 181(3):423–447, 1985.
- Levy, A., Poitevin, F., Martel, J., Nashed, Y., Peck, A., Miolane, N., Ratner, D., Dunne, M., and Wetzstein, G. CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images. *ECCV*, 13681:540–557, October 2022.
- Miolane, N., Poitevin, F., Li, Y.-T., and Holmes, S. Estimation of orientation and camera parameters from cryo-electron microscopy images with variational autoencoders and generative adversarial networks. *CVPR Workshop on Computer Vision for Microscopy Image Analysis*, 2019.
- Nashed, Y., Peck, A., Martel, J., Levy, A., Koo, B., Wetzstein, G., Miolane, N., Ratner, D., and Poitevin, F. Heterogeneous reconstruction of deformable atomic models in Cryo-EM. *NeurIPS Machine Learning in Structural Biology Workshop*, September 2022.
- Nashed, Y. S. G., Poitevin, F., Gupta, H., Woppard, G., Kagan, M., Yoon, C. H., and Ratner, D. CryoPoseNet: End-to-End Simultaneous Learning of Single-particle Orientation and 3D Map Reconstruction from Cryo-electron Microscopy Data. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 4049–4059, October 2021.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science: A Publication of the Protein Society*, 30(1):70–82, January 2021.
- Punjani, A. and Fleet, D. 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM. *Microscopy and Microanalysis*, 28(S1):1218, August 2022.
- Rosenbaum, D., Garnelo, M., Zielinski, M., Beattie, C., Clancy, E., Huber, A., Kohli, P., Senior, A. W., Jumper, J., Doersch, C., Eslami, S. M. A., Ronneberger, O., and Adler, J. Inferring a continuous distribution of atom

coordinates from cryo-em images using vaes. *NeurIPS Machine Learning in Structural Biology Workshop*, 2021.

Ullrich, K., van den Berg, R., Brubaker, M., Fleet, D., and Welling, M. Differentiable probabilistic models of scientific imaging with the fourier slice theorem. In *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, pp. 399–411. PMLR, 22–25 Jul 2020.

Zhong, E. D., Lerer, A., Davis, J. H., and Berger, B. Exploring generative atomic models in cryo-em reconstruction. *NeurIPS Machine Learning in Structural Biology Workshop*, 2020.

Zhong, E. D., Bepler, T., Berger, B., and Davis, J. H. Cryo-DRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2):176–185, February 2021.

## A. Normal Mode Analysis

Normal mode analysis (NMA) assumes that a set of  $N$  atoms are an oscillating system and represents them with an elastic network model (ENM). Noting  $\mathbf{X} = \{\mathbf{r}_j\}_{j=1,\dots,N}$  the atomic Cartesian coordinates of the chain, the potential energy  $E$  of the ENM is approximated with a second-order Taylor approximation around a reference conformation  $\mathbf{X}^{(0)}$ :

$$E(\mathbf{X}) = \frac{1}{2} (\mathbf{X} - \mathbf{X}^{(0)})^\top \mathbf{H} (\mathbf{X} - \mathbf{X}^{(0)}), \quad (2)$$

with  $\mathbf{H}_{jk} \propto (\mathbf{r}_j^{(0)} - \mathbf{r}_k^{(0)}) (\mathbf{r}_j^{(0)} - \mathbf{r}_k^{(0)})^\top$ , where  $\mathbf{H}$  is the Hessian matrix of  $E$  in  $\mathbf{X}^{(0)}$ . Then, eigendecomposition of  $\mathbf{H}$  is performed to obtain the eigenvectors  $\mathbf{U} \in \mathbb{R}^{3N \times 3N}$  which are also referred to as the normal modes.

## B. Image Formation Model

The process of image formation in cryo-EM involves several physical phenomena, including pairwise interactions between atoms, interactions between the electron beam and the molecule’s electrostatic potential, or microscope effects. In CryoChains, the image formation model (Fig. 1) is placed in the decoder, taking in the deformed and rigidly transformed atom coordinates computed from the dynamics model. The image formation model is usually written in its “weak-phase approximation” where each image  $\mathbf{I}_i$  in a dataset of  $n$  images of  $n$  biomolecules is sampled according to:

$$\mathbf{I}_i = \text{PSF}_i * (\mathbf{t}_i + \Pi_{2D} \mathbf{R}_i)(\mathbf{V}_i) + \epsilon_i, \text{ with } i = \{1, \dots, n\}. \quad (3)$$

Here,  $\mathbf{R}_i \in \mathcal{SO}(3)$  is a 3D rotation matrix representing the 3D orientation of the volume  $\mathbf{V}_i$  w.r.t. the direction of the electron beam. Note that  $\mathbf{R}_i$  represents a “global” orientation of the whole biomolecule, as opposed to  $\mathbf{R}^{(c)}$  inferred in our model.  $\mathbf{R}_i$  is obtained using other methods and assumed that it is given. This applies to the global translation  $\mathbf{t}_i$  explained below as well. The oriented volume is subsequently “pierced through” by the electron beam and projected onto the detector, which is an operation represented in Eq. (3) by the 2D projection operator  $\Pi_{2D}$ . The variable  $\mathbf{t}_i \in \mathbb{R}^2$  represents the 2D translation of the projected volume w.r.t. the center of the image. The effect of the microscope’s lens is modeled through the convolution  $*$  of the 2D projection with an image-dependent operator  $\text{PSF}_i$  called the point spread function of the microscope whose parameters can depend on the image. Finally, additional noise  $\epsilon_i$  is introduced in the generated image, and typically assumed to be Gaussian with zero mean and variance  $\sigma_i^2$ .

## C. Autoencoder Architecture

The encoder  $\mathcal{E}_\theta$  comprises 4 DoubleConvBlocks where each of them has 2 blocks of {convolution - max\_pooling - ReLU} layers. The 4 DoubleConvBlocks have  $\{32, 128, 256, 512\}$  filters each and kernels of size  $3 \times 3$ .  $\mathcal{E}$  embeds the input images into the latent variables, which branch out to a group of three multilayer perceptrons (MLPs) per chain, inferring (i) per-chain NMA weights  $\alpha_k$ ; (ii) two 3D vectors for the per-chain rotation; and (iii) a 3D vector for the per-chain translation. Each of the MLPs consists of two fully connected layers of sizes  $\{512, 256\}$  with a ReLU as an activation function in the first layer. For the second layer, the per-chain rotation MLPs have a tanh and the per-chain translation MLPs do not have an activation function.

## D. Additional Experiments

### D.1. Qualitative Evaluation

Let us visually assess the reconstruction quality of CryoChains in Fig. 4. We compute the mean conformation of the GT’s atom coordinates as well as that of the predicted atom coordinates, over the test set. Then, the Euclidean distances between the vertices of the mean GT and predicted surface meshes are computed. We opted to do the above because the reconstruction from our model is in fact the atom coordinates, not voxels or surfaces. We can observe that the reconstruction of CryoChains is visually better than that of N100, and the RMSDs are mainly very small except for some extremities of the proteins for both cases (see also the histograms next to the colorbars). We can clearly see the merit of CryoChains in the reconstruction of Hsp90, compared to the existing approach.

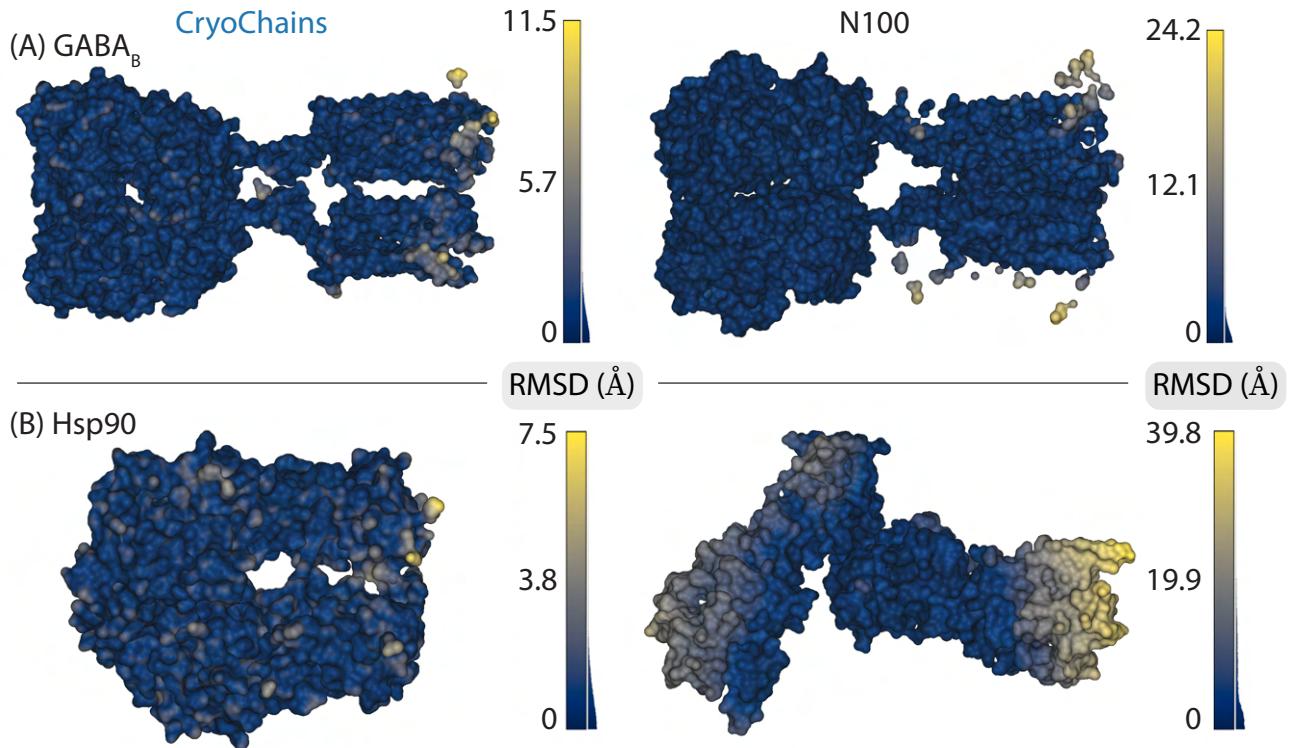


Figure 4: Reconstruction RMSD maps for (A) GABA<sub>B</sub> and (B) Hsp90, using CryoChains (ours) and N100 (Nashed et al., 2022) (whole protein NMA with 100 normal modes). RMSD is computed between the mean conformation of all GT conformations in the dataset and that of all reconstructed conformations. The reconstructed conformation is shown. Next to the colorbar is the histogram for the number of mesh vertices against the errors. The significant improvement in reconstruction for Hsp90 is clearly shown, i.e., the reconstruction using CryoChains correctly matches the closed conformation while that of N100 remains open.