Exploring generative atomic models in cryo-EM reconstruction

Ellen D. Zhong
MIT
zhonge@mit.edu

Adam Lerer
Facebook AI Research
alerer@fb.com

Joseph H. Davis
MIT
jhdavis@mit.edu

Bonnie Berger
MIT
bab@mit.edu

Abstract

Cryo-EM reconstruction algorithms seek to determine the 3D electron scattering density from a series of noisy, unlabeled 2D projection images captured with an electron microscope. Although 3D reconstruction algorithms typically model the 3D volume as a generic function parameterized as a voxel array or neural network, the underlying atomic model of the protein of interest places well-defined physical constraints on the reconstructed structure. In this work, we exploit prior information provided by an atomic model to reconstruct distributions of 3D structures from a cryo-EM dataset. We consider generative models for the 3D volume based on a coarse-grained model of the protein's atomic structure, with radial basis functions used to model atom locations and their physics-based constraints. Although the reconstruction objective is highly non-convex when formulated in terms of atomic coordinates (similar to the protein folding problem), we show that gradient descent-based methods can reconstruct a continuous distribution of atomic structures when initialized from a structure in its support. This approach is a promising direction for integrating biophysical simulation, learned neural models, and experimental data for 3D protein structure determination.

1 Introduction

Single particle cryo-electron microscopy (cryo-EM) is a powerful experimental technique for structure determination of proteins and macromolecular complexes at near-atomic resolution [1]. In this technique, an electron microscope is used to image a purified sample of the molecule of interest suspended in vitreous ice. Initial processing of the resulting micrograph produces a dataset of 10^{4-7} 2D projection images, where each image captures a unique molecule suspended in a random, unknown orientation. Reconstruction algorithms are then used to computationally infer the 3D protein structure from the dataset. Unlike structure determination through x-ray crystallography, cryo-EM is able to determine the structure of molecules that are not amenable to crystallization, including those that adopt an ensemble of heterogeneous conformational states. Thus, cryo-EM is in principle capable of modeling the entire ensemble of conformational states in a sample and probing the dynamic conformational landscape of proteins. To date, this promise has been limited by a lack of computational techniques for dealing with heterogeneous cryo-EM data, but promising new approaches have recently been developed to reconstruct heterogeneous structures [2, 3, 4].

Computational processing of cryo-EM datasets consists of several distinct stages (Figure 1). First, the noisy micrograph image is segmented, where bounding boxes containing individual molecules (i.e. particles) are identified and extracted (i.e. particle picking) [5]. Next, a 3D cryo-EM density volume, or distribution of 3D volumes that represent the molecule's electron scattering potential is reconstructed from the 2D projections[6]. Finally, an atomic model is built into the resulting 3D

Machine Learning for Structural Biology Workshop, NeurIPS 2020, Vancouver, Canada.

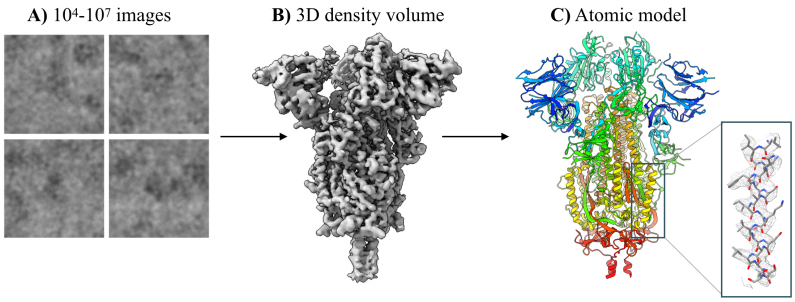

Figure 1: Structure determination via cryo-EM. Schematic of cryo-EM reconstruction (**A** $\rightarrow$ **B**) and atomic model fitting (**B** $\rightarrow$ **C**). In this work, we investigate the possibility of fitting the atomic protein structure directly during the reconstruction process. Dataset from Walls et al. [13].

volume either *de novo* or fit starting from a related model, typically with the help of automated tools [7, 8, 9, 10, 11, 12].

In this work, we investigate the possibility of fitting the atomic protein structure directly during the reconstruction process. We have two motivations for this. First, atomic fitting is a labor-intensive step in cryo-EM post-processing, and in particular it is not clear how to perform atomic fitting for models of distributions of protein conformations learned during reconstruction with new tools such as cryoDRGN [2]. Second, modeling the atomic structure of the 3D volume provides a strong prior over structures. In fact, in many cases the protein sequence and an approximate reference structure are known beforehand, strongly constraining the space of feasible 3D volumes. These structural priors are especially important for models of heterogeneous distributions of molecules, because they constrain the conformational dynamics to those that approximate realistic protein motions. Without such priors, it is common to observe artifacts in the motion of the protein, e.g. mass appearing and disappearing between two distinct conformations with rarely-sampled transition states.

To this end, we propose a reconstruction process based on a coarse-grained atomic model for the cryo-EM volume. The model fits parameters including atomic coordinates, to maximize the likelihood of the dataset images under a generative model that maps atomic structures to cryo-EM images. When initialized from an approximate reference structure for the protein of interest, this approach is able to learn both homogeneous structures and heterogeneous ensembles from synthetic cryo-EM images.

2 Background

A cryo-EM experiment produces a dataset of 10^{4-7} noisy 2D projection images, each containing a unique molecule captured in a random, unknown orientation. The goal of cryo-EM reconstruction is to infer the 3D density volume $V : \mathbb{R}^3 \rightarrow \mathbb{R}$ that gave rise to the imaging dataset $X_1, ..., X_N$. As cryo-EM images are integral projections of the molecule in this imaging modality, 2D images can be related to the 3D volume by the Fourier slice theorem [14], which states that the Fourier transform of a 2D projection is a central slice from the 3D Fourier transform of the volume. Traditional methods approximate the volume as a voxel array $\hat{V}(\mathbf{k})$ in Fourier space [6].

To recover the desired structure, cryo-EM reconstruction methods must jointly solve for the unknown volume $\hat{V}$ and image poses $\phi_i = (R_i, t_i)$, where $R_i \in SO(3)$ and $t_i \in \mathbb{R}^2$ are the 3D orientation of the molecule and in-plane image translation, respectively. Expectation maximization and simpler variants of coordinate ascent are typically employed to find a *maximum a posteriori* estimate of $\hat{V}$ marginalizing over the posterior distribution of ϕ_i's. The reader is referred to [15] for further details on traditional, homogeneous cryo-EM reconstruction method.

A unique advantage of cryo-EM is its ability to image heterogeneous molecules. *Heterogeneous reconstruction* algorithms aim to reconstruct a distribution of structures from the dataset. A standard approach involves extending the generative model to assume images are generated from a mixture model of K volumes $V_1, ..., V_K$ [16, 17]. More recently, neural models, such as cryoDRGN have been used to reconstruct heterogeneous ensembles of particles from cryo-EM data [2]. CryoDRGN represents a continuous n-dimensional distribution over volumes as a function $\hat{V} : \mathbb{R}^{3+n} \rightarrow \mathbb{R}$ approximated by a positionally-encoded MLP [2]. In cryoDRGN and in other advanced reconstruction

methods [18, 3, 19, 4], to simplify reconstruction, they find it sufficient to use poses ϕ computed using a traditional reconstruction method, and focus on the volume reconstruction.

3 Method

The central contribution of this work is a cryo-EM density model, $V(\mathbf{r}|\theta)$, that is parameterized in terms of coordinates for a coarse-grained atomic model given a known atomic reference structure (Figure 2A). In this coarse-grained model, each amino acid is represented by two Gaussian radial basis functions (RBFs), one representing the backbone and the second representing the sidechain. Each RBF is parameterized by (μ_i, σ_i, a_i), where μ_i is the position of the ith RBF; a_i is the amplitude of the ith RBF; and σ_i is the width of the ith RBF. We tie $a_i = a_0 Z_i$ where a_0 is a global learned amplitude constant, and Z_i is the total number of electrons in the fragment represented by RBF i. Furthermore, we tie all σ_i to the same value σ. Thus, the full RBF model has $3K + 2$ parameters, where K is the total number of amino acids in the protein complex. The cryo-EM density can be computed as a function of these parameters as:

$$V(\mathbf{r}) = \sum_i (2\pi\sigma_i)^{-3/2} a_i \exp\left(\frac{-||\mathbf{r} - \mu_\mathbf{i}||^2}{2\sigma_i^2}\right) \tag{1}$$

As described in Section 2, reconstruction is performed in the Fourier domain. This makes the choice of Gaussian RBFs convenient, as V can be computed efficiently in Fourier space[1]:

$$\hat{V}(\mathbf{k}) = \sum_i a_i e^{-2\pi i \mu_\mathbf{i} \cdot \mathbf{k}} \exp\left(\frac{-\pi^2 \mathbf{k}^2 \sigma_i^2}{2}\right) \tag{2}$$

To impose physical constraints on the RBF model, we add a set of harmonic bond terms $\mathcal{B}$ between consecutive backbone RBF centers. Each bond term $(i, j, k, l) \in \mathcal{B}$ is specified by a pair of RBF indices i, j and a bond strength k. The bond length l is set to 3.8 Å, the distance between protein C_α backbone carbons.

We constrain the side-chain RBFs to be located close to their backbone RBF using one-sided harmonic constraints $\mathcal{C}$. These are similar to the bond terms, but only induce a loss when the distance between RBF centers exceeds the max length l. We set l to the maximum distance between a backbone C-alpha carbon and it's side chain center of mass observed in the reference structure.

For homogeneous reconstruction, we fit $\{\mu_i\}, \sigma, a_0$ directly using stochastic gradient descent (SGD). The overall loss function, given a set of N images $\mathbf{X}$ with poses ϕ, bond terms $\mathcal{B}$ and side-chain constraints $\mathcal{C}$ is:

$$\begin{aligned}\mathcal{L}(\mu, \sigma, a_0|\mathbf{X}) = \frac{1}{N} \sum_{(x,\phi)\in X} ||\hat{X}(\phi|\mu, \sigma, a_0) - X||^2 + \sum_{(i,j,k,l)\in\mathcal{B}} k(||\mu_i - \mu_j|| - l)^2 \\ + \sum_{(i,j,k,l)\in\mathcal{C}} k \ \max\left(||\mu_i - \mu_j|| - l, 0\right)^2\end{aligned} \tag{3}$$

For heterogeneous reconstruction, we learn a continuous latent variable model of conformational heterogeneity expressed through motion of RBF centers (Figure 2B). Unlike standard heterogeneous cryo-EM reconstruction algorithms that use an unconstrained volume representation (e.g. voxel arrays or positionally encoded MLPs), the RBF model constrains the effect of the latent degrees of freedom to heterogeneity in the underlying atomic structure. An image encoder E with parameters θ_E predicts a latent $z \sim E(\hat{X}|\theta_E)$; A decoder network D with parameters θ_D predicts a z-dependent translation of the RBF centers $\mu_{het}(z) = \mu + D(z|\theta_D)$. We optimize $\mu, \sigma, a_0, \theta_D, \theta_E$ together end-to-end with SGD.

[1] Incidentally, projections of Gaussian kernels can also be computed analytically in real space, obviating the need for the Fourier slice theorem altogether. This real space formulation allows for algorithms that scale better with the number of RBFs since the RBFs are localized in real space, allowing for spatial decomposition via gridding, KD-trees, etc.

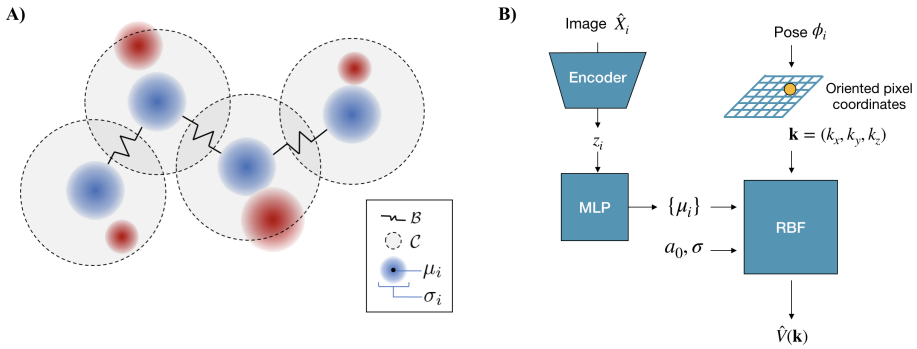

Figure 2: **A)** Generative model for reconstruction. The 3D cryo-EM density is modeled as a set of Gaussian RBFs, two for each AA in the reference structure. One RBF represents the backbone (blue) and one represents the sidechain (red). Backbone RBFs are connected with harmonic bond terms $\mathcal{B}$, and sidechain RBFs are connected to their backbone with max-distance harmonic constraints $\mathcal{C}$. **B)** Architecture for heterogeneous reconstruction. We use a VAE to learn z-dependent offsets of RBF centers μ_i, given an unlabelled imaging dataset of image, pose pairs $(\hat{X}_i, \phi_i)$.

A major shortcoming of this atomic reconstruction approach is the existence of many local minima of the loss function that do not approximate the true atomic coordinates. This is in contrast to voxel-based models, where SGD converges to the global minimum of the convex loss given the poses. We address the local minima problem primarily by initializing RBF centers μ_i from a reference structure that is a close approximation of the imaged structure (homogeneous) or some point on the distribution of structures (heterogeneous). Such reference structures are often available when studying a variant of a known structure, or heterogeneous protein dynamics.

4 Results

Here, we present results for our RBF model in reconstructing coarse-grained atomic structures from synthetic cryo-EM image data. We first explore the effect of reference structure initialization in homogeneous reconstruction, when initialized from either an approximate reference structure (fitting), the exact structure (cheating), or from randomly initialized locations (folding). We then turn to heterogeneous cryo-EM datasets, where we evaluate the ability of the RBF model to reconstruct continuous distributions of structures and their atomic coordinates. Finally, we assess choices in the design of the RBF model by ablating key components in the homogeneous setting.

Datasets. We generate homogeneous and heterogeneous cryo-EM datasets using a 141-residue atomic model (PDB 5NI1) as the ground truth structure. To generate homogeneous data, we simulate 50,000 noisy projection images based on the cryo-EM image formation model. To model heterogeneity, we introduce a bond rotation in the backbone of 5ni1 to create a continuous 1D motion, and generate cryo-EM images sampling along the ground truth reaction coordinate. Further details on dataset generation are given in the Appendix.

Architecture and training. For all experiments, we train for 10 epochs using the Adam optimizer in minibatches of 8 images and a learning rate of 1e-4. We initialize σ and a_0 to 3.71 and 0.1, and perform one epoch of 'warm-up' to refine these global parameters after which we reset the atomic coordinates to their initial values. For homogeneous reconstruction, we directly optimize all RBF parameters. For heterogeneous reconstruction, we use a VAE to predict the z-dependent offsets of the RBF centers from their reference values. Both the encoder and decoder are 3-layer MLPs of width 256 and residual connections, and a latent dimension $|z|$ of 1. We use ground-truth poses ϕ for training. In real applications, the poses would be inferred from traditional cryo-EM tools [20, 21, 22].

4.1 Homogeneous reconstruction

To explore the effect of reference structure initialization, we compare the reconstruction accuracy of the RBF model when initialized from the ground truth coordinates and from three alternate starting

configurations: 1) an approximate reference structure generated by evolving the system under a molecular dynamics simulation, 2) the ground truth structure with 6 Å uniform noise added to the ground truth coordinates of each C_α, and 3) a random initialization of each C_α RBF in the 64^3 region in the center of the box. Side-chain RBFs are initialized to their corresponding C_α RBF centers.

We report the root-mean-square-error (RMSE) of model backbone coordinates to the C_α of the true structure, the percent of C_α backbone atoms predicted within 3 Å of the true structure, and the normalized mean-square-error (NMSE) of the reconstructed volume to the true structure (Table 1). We find that our atomic model is quite sensitive to initialization. While the model performs adequately when initialized at nearby reference structures, it makes some mistakes due to local minima, and performs much better when initialized with the ground truth coordinates. When initialized from random coordinates, SGD is unable to recover the ground truth atomic coordinates whatsoever. Instances of atomic local minima include 'mismatched buttoning' of the amino acid backbone in the volume, as well as incorrect tracing of the protein backbone through the volume (Figure S2).

Initialization	**Initial RMSE (Å)**	**C_α RMSE (Å)**	**% within 3Å**	**Volume NMSE**
Exact 5NI1	0.00	0.843	99.29%	0.28
Approximate 5NI1	5.15	3.81	77.30%	0.32
5NI1 + Uniform[6 Å]	6.50	3.19	53.19%	0.35
Random	34.87	17.16	2.12%	0.43

Table 1: Comparison of different choices of initial atomic coordinates when training the model.

4.2 Heterogeneous reconstruction

As opposed to the previous section where we reconstruct a homogeneous structure from a nearby reference structure, here we attempt to reconstruct a continuous manifold of structures from a reference structure lying at one point in the distribution. We consider a dataset consisting of a 1D continuous motion of the 5ni1 protein. A bond in the center of the protein is rotated leading to a large-scale global conformational change (Figure S1). We initialize the RBF model to the structure at one end of the reaction coordinate, and train on the imaging dataset with ground truth poses.

The RBF model is able to correctly reconstruct the full distribution of this large conformational change. In both datasets, the latent encodings of the images is well correlated with the true reaction coordinate (Figure 3, left), and the RBF atomic coordinates from traversing the latent space nearly exactly reconstruct the underlying protein motion (Figure 3, right).

As a baseline, we perform heterogeneous reconstruction with cryoDRGN, which learns an unconstrained neural representation of cryo-EM volumes (Appendix A). Similarly, we provide the ground truth poses and train a 1-D latent variable model. While the latent space is well correlated with the ground truth motion, without the regularization provided by the structural RBF prior, the reconstructed volumes contain noise and blurring artifacts in the mobile region (Figure 3, bottom).

We also measure reconstruction accuracy across the reaction coordinate for four distributions of images across the reaction coordinate: a uniform distribution, two non-uniform distributions corresponding to an energy barrier; and two discrete clusters with no images in the middle (Figure 4). For each value of the reaction coordinate, we approximate the atomic coordinates using the median latent from the images at that coordinate, and measure its C_α RMSE with the ground truth. We see that when there is even a small probability mass across the reaction coordinate (top row), the atomic model learns the full distribution of conformations with high accuracy. If transition states are nearly or completely unobserved in the image distribution (bottom row), reconstruction accuracy is poor.

4.3 Model ablations

Using the homogeneous dataset and an initialization from the approximate 5NI1 structure, we explore various choices in design of the RBF model by ablating the side chain RBF, bond constraints, and the cryo-EM supervision (Table 2). We find that removing the sidechain RBFs and modeling each amino acid as a single RBF slightly degrades quality, while removing the internal bond terms leads to a dramatic degradation. Atomic accuracy is substantially worse but not completely degraded

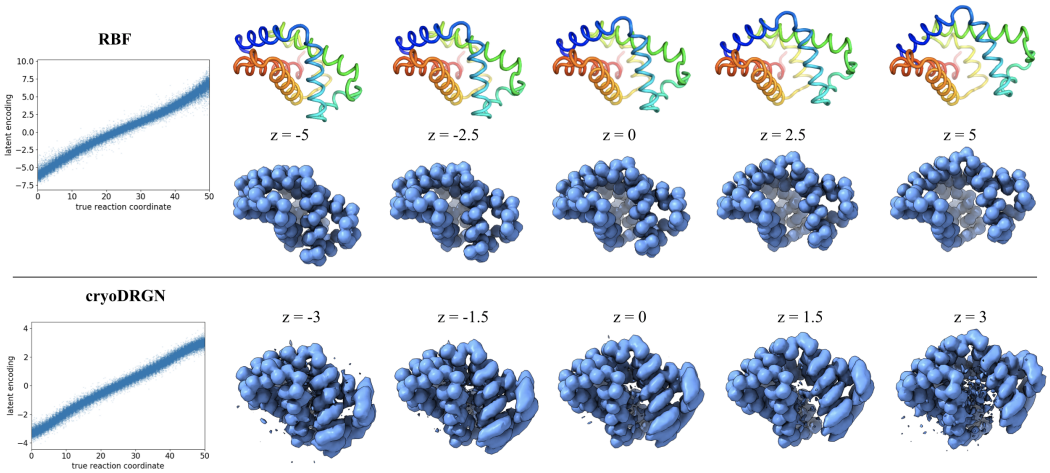

Figure 3: Heterogeneous reconstruction results of an unlabeled dataset containing a uniform distribution of images across the ground truth reaction coordinate. Predicted 1D latent encoding z plotted against the ground truth reaction coordinate (left), and reconstructed structures at the specified values of z. Our RBF model directly reconstructs atomic coordinates (top). The unconstrained neural volume representation (cryoDRGN) contains noise and blurring of moving atoms in the reconstructed volumes (bottom) and does not produce atomic coordinates.

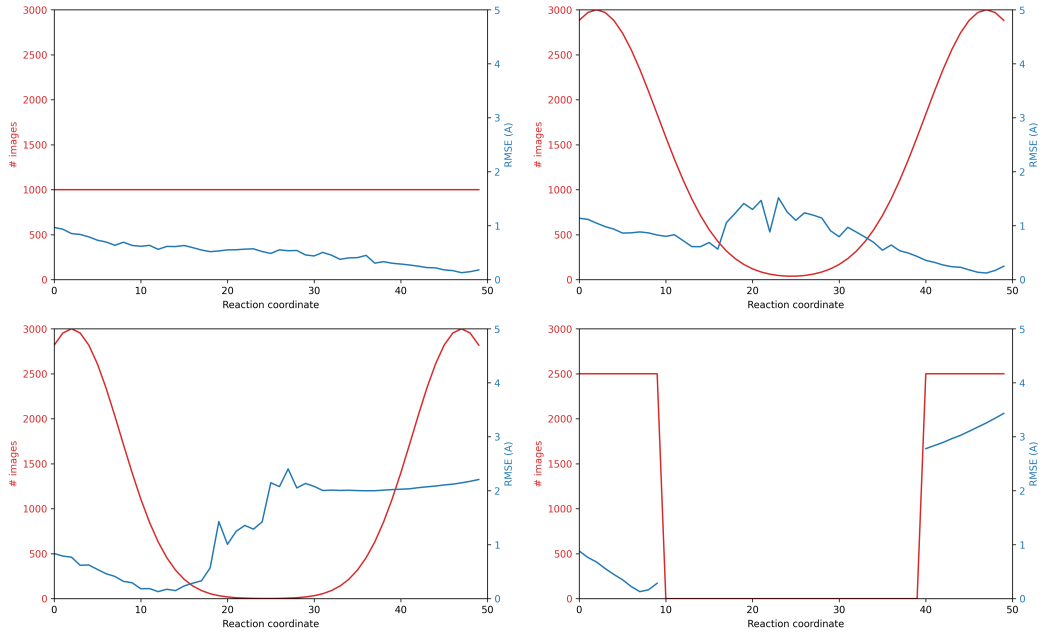

Figure 4: C_α RMSE (blue) for heterogeneous reconstruction of a large synthetic conformational change of 5ni1, with different distributions of images (red) across the reaction coordinate. The reference structure used for initialization is the ground truth atomic structure at reaction coordinate 0.

even when ignoring the cryo-EM images (due to the initialization); however, as expected there is low overlap with the true volume in this case. We expect the volume NMSE to be higher than unconstrained approaches (e.g. cryoDRGN), as the coarse-grained RBF model does not exactly match the underlying all-atom generative model even with correct coordinates (Figure S3).

Model	C_α RMSE (Å)	% within 3Å	Volume NMSE
Full Model	**3.81**	**77.30%**	**0.32**
No Sidechain RBFs	4.10	70.21%	0.49
No Bonds	10.42	37.59%	0.36
No Cryo-EM Loss	5.15	74.47%	1.82
CryoDRGN	N/A	N/A	0.08

Table 2: Ablations of model components. We remove the sidechain RBFs, the bond terms between RBFs, and the cryo-EM reconstruction loss; each degrades the quality of reconstruction.

5 Discussion

This work proposes a coarse-grained atomic generative model from cryo-EM reconstruction. Our experiments suggest that such models are a promising direction for incorporating structural priors into cryo-EM reconstruction, especially for heterogeneous structures. However the experimental validation is preliminary: we worked entirely with *synthetic* cryo-EM datasets of a *small* protein using *exact poses*. Follow-up work is required to understand how these techniques behave with real cryo-EM images and volumes, using realistic protein complexes of interest, and using approximate poses generated by existing tools. For larger protein complexes, follow-up work will investigate whether a coarser model granularity may be more appropriate, especially when modeling large-scale heterogeneous dynamics.

Results on homogeneous datasets suggest that local minima in optimization space are a major problem for this class of methods if coordinates are not initialized very close to the ground truth structure. These local minima problems could potentially be ameliorated with a combination of improved modeling of structural priors such as bond terms and steric interactions, and optimization methods such as Hamiltonian dynamics or Markov Chain Monte Carlo that can escape local minima. There is a rich literature on these topics in the domain of molecular dynamics simulation which could carry over to cryo-EM reconstruction.

However, results on heterogeneous datasets suggest that we can correctly learn distributions with large continuous conformational changes when initialized from some structure in the distribution, even for structural changes that are too large to be modeled correctly if only the endpoint structures are observed (as in homogeneous reconstructions). Follow up work is required to more fully characterize exactly when neural models of heterogeneous structures converge to the correct distribution.

References

[1] Eva Nogales. The development of cryo-EM into a mainstream structural biology technique. *Nature Methods*, 13(1):24–27, jan 2016.

[2] Ellen D Zhong, Tristan Bepler, Joseph H Davis, and Bonnie Berger. Reconstructing continuous distributions of 3D protein structure from cryo-EM images. *ICLR*, 2020.

[3] Ali Punjani and David J Fleet. 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *bioRxiv*, 11(2):2020.04.08.032466, apr 2020.

[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, June 2018.

[5] T Bepler, A Morin, M Rapp, J Brasch, and L Shapiro. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*.

[6] Amit Singer and Fred J. Sigworth. Computational Methods for Single-Particle Electron Cryomicroscopy. *Annual Review of Biomedical Data Science*, 3(1):163–190, jul 2020.

[7] P Emsley, B Lohkamp, W G Scott, and K Cowtan. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):486–501, April 2010.

[8] Dorothee Liebschner, Pavel V. Afonine, Matthew L. Baker, Gábor Bunkoczi, Vincent B. Chen, Tristan I. Croll, Bradley Hintze, Li Wei Hung, Swati Jain, Airlie J. McCoy, Nigel W. Moriarty, Robert D. Oeffner, Billy K. Poon, Michael G. Prisant, Randy J. Read, Jane S. Richardson, David C. Richardson, Massimo D. Sammito, Oleg V. Sobolev, Duncan H. Stockwell, Thomas C. Terwilliger, Alexandre G. Urzhumtsev, Lizbeth L. Videau, Christopher J. Williams, and Paul D. Adams. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, 75(10):861–877, oct 2019.

[9] Leonardo G Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank, and Klaus Schulten. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure (London, England : 1993)*, 16(5):673–683, May 2008.

[10] Daisuke Kihara, Genki Terashi, and Sai Raghavendra Maddhuri Venkata Subramaniya. De Novo Computational Protein Tertiary Structure Modeling Pipeline for Cryo-EM Maps of Intermediate Resolution. *Biophysical Journal*, 118(3):292a, feb 2020.

[11] Dong Si, Spencer A Moritz, Jonas Pfab, Jie Hou, Renzhi Cao, Liguo Wang, Tianqi Wu, and Jianlin Cheng. Deep Learning to Predict Protein Backbone Structure from High-Resolution Cryo-EM Density Maps. *Scientific Reports*, 10(1):1–22, mar 2020.

[12] Tristan Ian Croll. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology*, 2018.

[13] 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*, 2020.

[14] Ronald N Bracewell. Strip integration in radio astronomy. *Australian Journal of Physics*, 9(2):198–217, 1956.

[15] Sjors H W Scheres. A Bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, 415(2):406–418, January 2012.

[16] Sjors HW Scheres, Mikel Valle, Rafael Nuñez, Carlos OS Sorzano, Roberto Marabini, Gabor T Herman, and Jose-Maria Carazo. Maximum-likelihood multi-reference refinement for electron microscopy images. *Journal of Molecular Biology*, 348(1):139–149, 2005.

[17] Lyumkis, Dmitry, Brilot, Axel F, Theobald, Douglas L, and Grigorieff, Nikolaus. Likelihood-based classification of cryo-EM images using FREALIGN. *Journal of Structural Biology*, 183(3):377–388, September 2013.

[18] Ellen D Zhong, Tristan Bepler, Bonnie Berger, and Joseph H Davis. CryoDRGN: Reconstruction of heterogeneous structures from cryo-electron micrographs using neural networks. *bioRxiv*, 16:2020.03.27.003871, March 2020.

[19] Joachim Frank and Abbas Ourmazd. Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods*, 100:61–67, May 2016.

[20] 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, March 2017.

[21] Jasenko Zivanov, Takanori Nakane, Björn O. Forsberg, Dari Kimanius, Wim J.H. Hagen, Erik Lindahl, and Sjors H.W. Scheres. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, 7, nov 2018.

[22] Timothy Grant, Alexis Rohou, and Nikolaus Grigorieff. cisTEM, user-friendly software for single-particle image processing. *eLife*, 7:e14874, mar 2018.

[23]

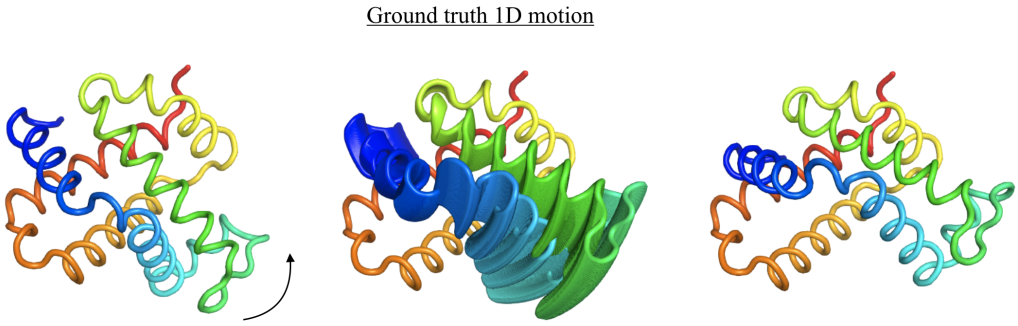

Figure S1: Ground truth structures of the heterogeneous datasets simulating a 1D continuous motion transitioning from left (5NI1) to right. All generated structures are shown in the center.

A Appendix

A.1 Dataset generation

Synthetic cryo-EM datasets were generated from an atomic model of PDB 5NI1 according to the image formation model as follows: starting from the deposited atomic model of 5NI1, the 141-residue A-chain subunit of the complex was extracted. A cryo-EM volume was generated with the 'molmap' command in Chimera [23] at 3 Å resolution and grid spacing of 1 Å. The volume was zero-padded to a cubic dimension of 128^3. 50,000 projection images were generated at random orientations of the volume uniformly from SO(3). Images were then translated in-plane by t uniformly sampled from $[-10, 10]^2$ pixels. We omit the application of the CTF for simplicity in this synthetic dataset. Gaussian noise was added leading to a signal to noise ratio (SNR) of 0.1, a typical value for cryo-EM image data. As real cryo-EM datasets have variable resolution and thus resolvability of the atomic structure, we investigate the effect of volume resolution and the included atoms in generating the dataset's ground truth volume/images (Table S1). We find that our atomic modeling is quite robust to the resolution and which atoms we model in the synthetic data, which suggests it may transfer well to real cryo-EM images.

To generate the heterogeneous datasets, a dihedral angle in the backbone of the atomic model was rotated through 0.25 radians and 50 structures were sampled along the motion (Figure S1). For simplicity of heterogeneous dataset generation, we use the 5NI1 atomic models with C_α backbone atoms only. We generate multiple datasets of 50k total images following the above image formation process, each with a different distribution of images along the reaction coordinate as shown in Figure 4.

Modeled Atoms	**Resolution**	C_α **RMSE (Å)**	**% within 3Å**
Cα	3Å	3.21	80.8%
Cα & Cβ	3Å	3.29	77.3%
All Atoms	3Å	3.75	78.0%
Cα	5Å	3.20	80.9%
Cα & Cβ	5Å	3.87	73.1%
All Atoms	5Å	3.80	74.5%
Cα	8Å	3.80	74.5%
Cα & Cβ	8Å	3.87	73.8%
All Atoms	8Å	4.20	78.7%

Table S1: Homogeneous reconstruction model accuracy for different ways of generating the dataset of cryo-EM images (resolution and which atoms are included). The model performs similarly across these choices. For other homogeneous experiments, we use all atoms at 3Å. For heterogeneous experiments, we use C_α at 3Å.

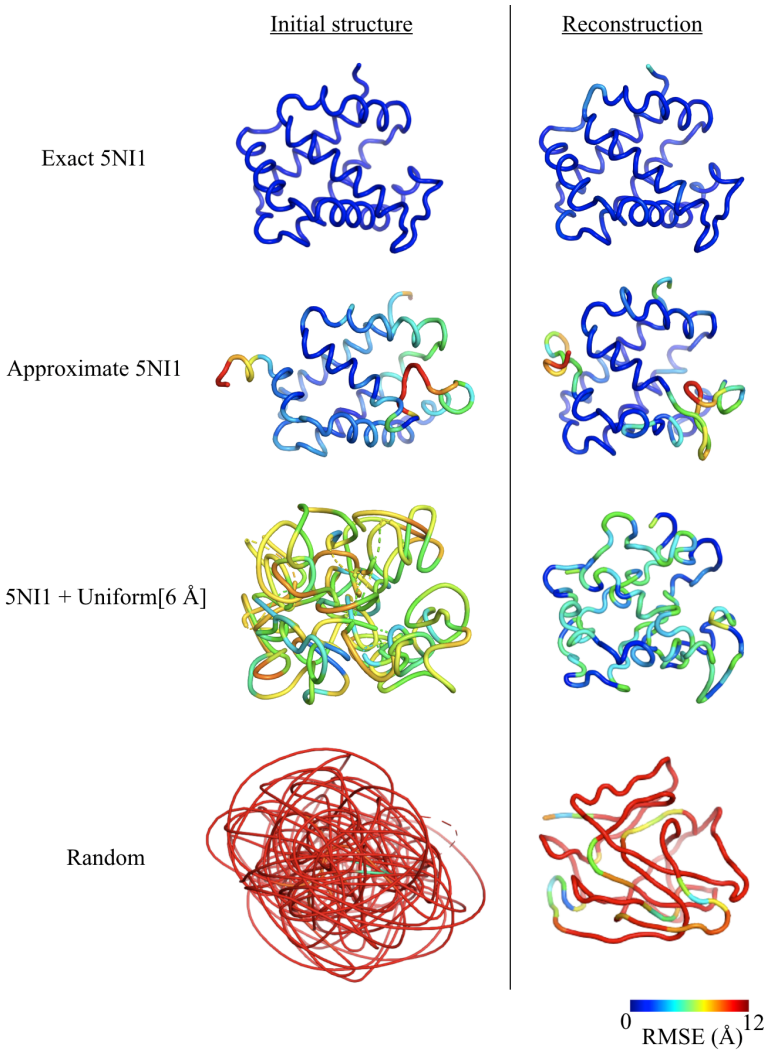

Figure S2: Reconstructed atomic structures with RBF centers initialized either 1) at the exact ground truth values of 5NI1, 2) at an approximate structure generated from evolution by molecular dynamics, 3) at 5NI1 coordinates randomly perturbed by Uniform[6 Å] noise, or 4) from random initial values. Structures are colored by C_α RMSE to the ground truth structure (top left).

A.2 cryoDRGN baseline

For homogeneous reconstruction, we train a cryoDRGN positionally-encoded 3-layer MLP of width 256 for 20 epochs. For heterogeneous reconstruction, both the encoder and decoder networks are 3-layer MLPs of width 256, and are trained for 20 epochs.

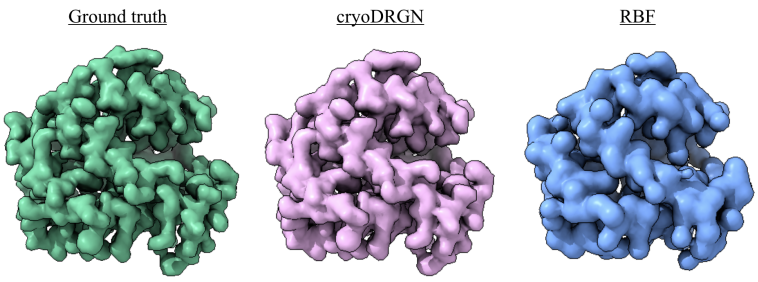

Figure S3: Ground truth and reconstructed volumes with a neural network representation of density (cyroDRGN) and with our RBF model initialized from exact coordinates. The RBF volume reconstruction is somewhat worse than that of cryoDRGN because even with exact coordinates, it cannot match the all-atom generative model.