

---

# Heterogeneous reconstruction of deformable atomic models in Cryo-EM

---

**Youssef Nashed** 

SLAC National Accelerator Laboratory

**Ariana Peck** 

SLAC National Accelerator Laboratory

**Julien Martel** 

Stanford University

**Axel Levy** 

Stanford University

**Bongjin Koo** 

University of California Santa Barbara

**Gordon Wetzstein** 

Stanford University

**Nina Miolane** 

University of California Santa Barbara

**Daniel Ratner** 

SLAC National Accelerator Laboratory  
dratner@slac.stanford.edu

**Frédéric Poitevin** 

SLAC National Accelerator Laboratory  
frederic.poitevin@stanford.edu

## Abstract

Cryogenic electron microscopy (cryo-EM) provides a unique opportunity to study the structural heterogeneity of biomolecules. Being able to explain this heterogeneity with atomic models would help our understanding of their functional mechanisms but the size and ruggedness of the structural space (the space of atomic 3D cartesian coordinates) presents an immense challenge. Here, we describe a heterogeneous reconstruction method based on an atomistic representation whose deformation is reduced to a handful of collective motions through normal mode analysis. Our implementation uses an autoencoder. The encoder jointly estimates the amplitude of motion along the normal modes and the 2D shift between the center of the image and the center of the molecule. The physics-based decoder aggregates a representation of the heterogeneity readily interpretable at the atomic level. We illustrate our method on 3 synthetic datasets corresponding to different distributions along a simulated trajectory of adenylate kinase transitioning from its open to its closed structures. We show for each distribution that our approach is able to recapitulate the intermediate atomic models with atomic-level accuracy.

## 1 Introduction

The cellular machinery functions thanks to molecular populations constantly transitioning between metastable states — or *conformations*. As a result, understanding the structural heterogeneity of biomolecules is key to understanding these essential building blocks of life [1, 17] and to subsequently making progress in drug and molecular design [15, 41]. Cryogenic electron microscopy (cryo-EM) of molecules [9] is the leading technique for generating large imaging datasets of biomolecules at equilibrium. Each 2D image in the dataset is unlabelled and contains a unique molecule that adopts an unknown 3D *conformation* and an unknown *pose* (orientation and position).

While *homogeneous* reconstruction methods focus on estimating the *average volume* of the studied molecule from collected images [32, 29, 21, 11, 24, 20], *heterogeneous* methods take into account structural variability and introduce a latent variable (the *conformational state*) that characterizes the

*volume* associated with each observation. Given the image formation model, heterogeneous reconstruction can be formulated as an inverse problem where the volume representation, conformation and pose must be inferred from the images [8]. Most heterogeneous methods to date treat the variability in the dataset on purely statistical grounds, through expectation-maximization approaches [32] or inference in deep generative models [43]. However, emerging approaches also inject knowledge about the underlying physics of the molecule in the forward model of the solver - specifically, in the way deformations of the volume are modelled (the *dynamic* model). Punjani *et al.* [28] and Herreros *et al.* [13] model the deformation of the voxelized volume with convective fields. Zhong *et al.* [44] and Chen *et al.* [6] use an initial coarse-grained atomic model as a reference from which the pseudo-atoms can deviate with a penalty. Rosenbaum *et al.* [30] keep all the atoms and the molecule is segmented in rigid blocks (corresponding to the protein residues) that can also deviate from a reference with a penalty, followed by minimization of the molecule’s potential energy, parameterized with an all-atom pairwise interaction force-field. While promising, this approach suffers from the very large size of the structural space (scaling with the number of residues, e.g.  $10^2 - 10^4$ ) and expensive energy minimization steps at each iteration for every image. Performing a normal mode analysis (NMA) [19] of the molecule represented with an elastic network model (ENM) [37] is a popular approach to modelling the essential dynamics of a molecule at a reduced problem dimensionality. In this representation, the size of the conformational space could in principle be reduced to a handful of normal modes while preserving the ability to model relevant functional motions of the molecule [36]. This approach was implemented in [14], where the pose and normal mode amplitudes are iteratively estimated for each image of the dataset. However, as the authors note, this approach suffers from the same scalability issue as expectation-maximization approaches (e.g. [31]). To remedy this, two approaches could be taken. The first approach would consist of applying this method to a subset of the data (the training set) and using its output to train a neural network, which would then infer the latent variables for the remaining data — this supervised approach was implemented and tested on synthetic and experimental datasets in Hamitouche and Jonic [12].

Here we introduce an unsupervised alternative that uses direct gradient-based optimization to jointly estimate the translational component of the pose (the shift) and the normal mode amplitudes of the molecule. Our method uses an autoencoder pipeline [10]. The encoder associates each particle image with a shift and a normal mode amplitude. The decoder is a generative physics-based pipeline that uses the predicted shift and normal mode amplitude along with the reference atomic model to predict an image. By learning a mapping from images to shifts and normal mode amplitudes, this technique avoids the computationally expensive step that limits existing cryo-EM reconstruction methods. Our approach thus amortizes over the size of the dataset and provides a scalable approach to working with modern, large-scale cryo-EM datasets [20].

## 2 Approach

### 2.1 Image Formation Model

In single particle cryo-EM, probing electrons interact with the electrostatic potential created by the molecules embedded in a thin layer of vitreous ice. Each molecular volume  $V_i$  can be seen as a mapping from  $\mathbb{R}^3$  to  $\mathbb{R}$  and is indexed by  $i \in \{1, \dots, N\}$ . In the sample, each molecule is in an unknown orientation  $R_i \in SO(3) \subset \mathbb{R}^{3 \times 3}$ , and an unknown conformation  $z_i \in \mathbb{R}^d$ . We assume that the volumes  $\{V_i\}_{i=1, \dots, N}$  are drawn independently from a probability distribution  $\mathbb{P}_V$  supported on a low-dimensional subspace (the *conformational space*). More specifically, we assume there exist  $d \in \mathbb{N}$  and  $\mathcal{V} : \mathbb{R}^3 \times \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $\mathbb{P}_V$  is supported on the conformational space  $\{\mathcal{V}(\cdot, z), z \in \mathbb{R}^d\}$ .

**Elastic Network Models, Normal Mode Analysis and Coarse-Graining.** The structural space of the molecular structure is  $3M$  dimensional, noting  $M$  its number of constitutive atoms  $\sim 10^2 - 10^5$ . Efficient sampling of this very high-dimensional space to uncover the lower-dimensional subspace defined by a handful of collective variables (the *conformational space*  $\mathcal{V}$ ) has been an area of intense research for many decades [34, 18, 25]. A relatively cheap and efficient option consists in performing a harmonic approximation of the energy landscape around a chosen conformation [26, 19] by representing the molecule with an elastic network model (ENM) [37, 2]. Noting  $X = \{\mathbf{r}_j\}_{j=1, M}$  the atomic Cartesian coordinates of the molecule, the potential energy  $E$  of the ENM is further approximated with a second-order Taylor approximation around a reference conformation  $X^{(0)}$  and

using  $H$  the Hessian matrix of  $E$  in  $X^{(0)}$ :

$$E(X) = \frac{1}{2}(X - X^{(0)})^T H(X - X^{(0)}) \text{ with } H_{jk} \propto (\mathbf{r}_j^{(0)} - \mathbf{r}_k^{(0)})(\mathbf{r}_j^{(0)} - \mathbf{r}_k^{(0)})^T. \quad (1)$$

Solving the equations of motion in  $E$  leads to the eigenvalue problem  $HU = \Lambda U$  where each of the  $3M$  eigenpairs  $(\lambda_l, U_l)$  defines a *normal mode* of frequency  $\sqrt{\lambda_l}$  that deforms the reference conformation along  $U_l = \{\mathbf{u}_j^{(l)}\}_{j=1, M}$ . Pragmatically, the normal modes provide a linear basis of functions that spans the whole conformational space. Because the variance of each mode scales with the inverse of its frequency, it is customary to perform a low-rank approximation of  $H$  by only considering the  $d$  lowest frequency normal modes  $\tilde{U}_d = \{U_1, \dots, U_d\}$  - a choice empirically justified as they often encode relevant functional motions [36]. Formally, for rank  $d$ , the reference atomic model can be deformed as follows:

$$X(z) = X^{(0)} + \tilde{U}_d \cdot z^T. \quad (2)$$

In practice, the eigenvalue problem is challenging to solve for large  $M$  as the Hessian is a  $3M \times 3M$  matrix. Because the lowest-frequency modes usually exhibit strong collectivity [36],  $H$  is usually computed for a subset of atoms, the *coarse-grained (CG) model*. Its eigenvectors can then be interpolated on the remaining set of atoms and orthonormalized to yield  $\tilde{U}_d$ .

**Isolated Atom Superposition Approximation, Weak-phase Approximation and Projection Assumption.**

The electrostatic potential of the whole molecule can be approximated as the superposition of the electron form factors of its constitutive atoms [40]. Each form factor is defined as the sum of 5 Gaussians — their amplitude  $a$  and width  $b$  determined by the atomic type [7]. For typical cryo-EM experiments, the scattered wave can be linearized and simplifies to the integral of the rotated volume along the path of the probing electron [39], resulting in the following mapping from  $\mathbb{R}^2$  to  $\mathbb{R}$  where  $x_j^{(i)}$  and  $y_j^{(i)}$  are the planar coordinates of atom  $j$  in volume  $i$  derived from  $\mathbf{r}_j^{(i)} = R_i \cdot (\mathbf{r}_j^{(0)} + \sum_{l=1}^d z_l^{(i)} \mathbf{u}_j^{(l)})^T$ :

$$I_i(x, y) = \int_t V_i(x, y, t) dt = 4\pi \sum_{j=1}^M \sum_{k=1}^5 \frac{a_{jk}}{b_{jk}} \exp - \frac{4\pi^2}{b_{jk}} ((x - x_j^{(i)})^2 + (y - y_j^{(i)})^2). \quad (3)$$

The interaction between the beam and the microscope’s lens is modeled by the Point Spread Function (PSF)  $P_i$ . Imperfect centering of the molecule in the image is characterized by small translations  $\mathbf{t}_i \in \mathbb{R}^2$ . Finally, taking into account signal arising from the vitreous ice in which the molecules are embedded as well as the non-idealities of the lens and the detector, each image  $\hat{Y}_i$  can be modelled as [40, 32]

$$\hat{Y}_i = T_{\mathbf{t}_i} * P_i * I_i + Z_i \Leftrightarrow \hat{Y}_i = \mathcal{T}_{\mathbf{t}_i} \odot C_i \odot \mathcal{I}_i + \mathcal{Z}_i \quad (4)$$

where  $*$  is the convolution operator,  $\odot$  indicates element-wise multiplication,  $T_{\mathbf{t}}$  is the  $\mathbf{t}$ -translation kernel and  $Z_i$  white Gaussian noise on  $\mathbb{R}^2$ . In Fourier space, we note  $\hat{Y}_i$  the Fourier transform of  $\hat{Y}_i$ ,  $\mathcal{I}_i$  the complex Fourier transform of the ideal image  $I_i$ ,  $\mathcal{T}_{\mathbf{t}}$  the  $\mathbf{t}$ -translation operator or phase-shift in Fourier space,  $C_i$  is the Contrast Transfer Function (CTF), Fourier transform of the PSF, and  $\mathcal{Z}_i$  the Fourier transform of  $Z_i$ .

## 2.2 Overview of the Architecture

Fig. 1 summarizes the proposed encoder-decoder architecture. Images  $Y_i$  are fed into an encoder, parameterized by  $\psi$ , that predicts a shift  $\mathbf{t}_i$  and a conformational variable  $z_i$ . These estimated parameters are then applied to a reference structure as in Eq. 2. The resulting deformed atomic model is rotated using a known pose  $R_i$ , and Eq. 3 is evaluated at the desired pixel positions. Based on the estimated translation  $\mathbf{t}_i$  and given CTF parameters  $C_i$ , the rest of the image formation model described in Eq. 4 is simulated with the physics-aware decoder to obtain  $\hat{Y}_i$ , a noise-free estimation of  $Y_i$ . Pairs of measured and reconstructed images are compared using the L2 loss, and gradients are backpropagated throughout the differentiable model in order to optimize the encoder.

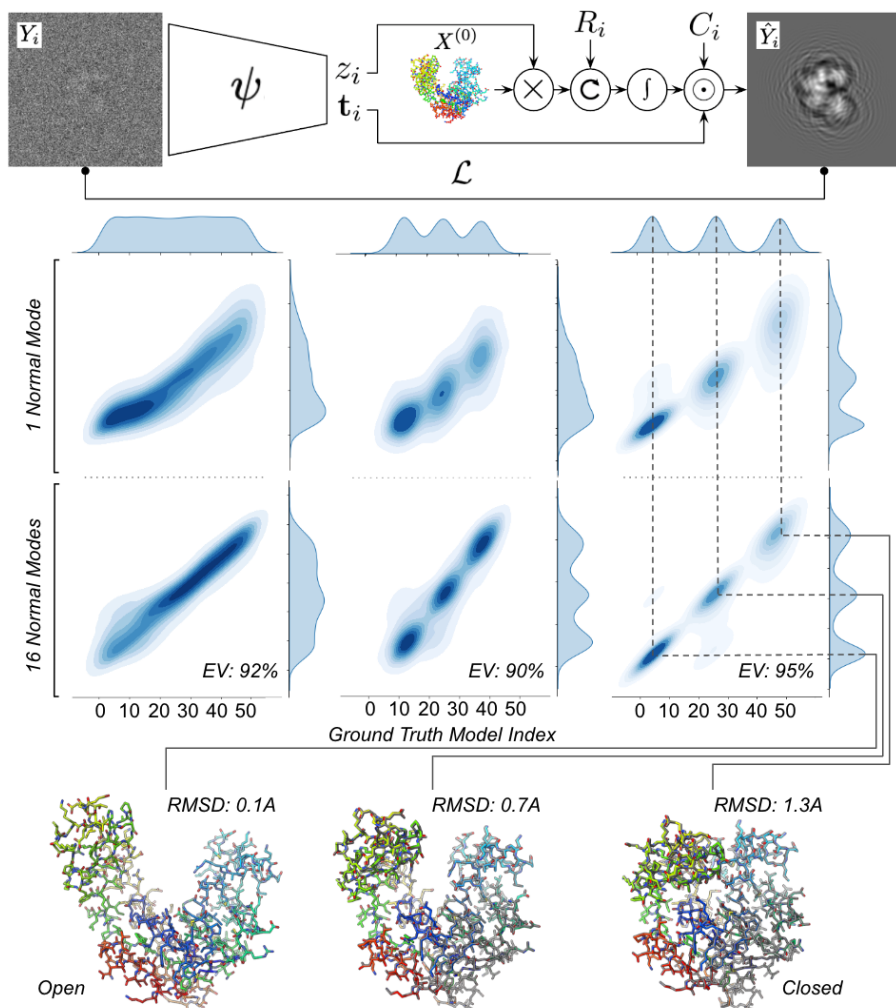


Figure 1: **(top) Autoencoder architecture.** The encoder, parameterized by  $\psi$ , maps an image  $Y_i$  to the shift  $t_i$  and conformational  $z_i$  variables. The decoder is a differentiable simulator that generates noise-free images  $\hat{Y}_i$  using the image formation model defined in Eq. 4 from a reference atomic model  $X^{(0)}$ , using the variables learned by the encoder. The input and output images are compared with the L2 loss, which is batch-minimized for  $\psi$ . **(bottom) Results.** GT and predicted distributions over the *uniform* (left), *continuous* (center) and *discontinuous* (right) datasets (see Section. 3) for  $d = 1$  (top row) and  $d = 16$  (bottom row). The explained variance (EV) of PC1 is indicated for  $d = 16$ . The GT (in gray) and  $d = 16$  predicted atomic models (rainbow-colored) corresponding to the local maxima in the *discontinuous* dataset are displayed together with their RMSD.

**Discriminative model.** The encoder acts as a discriminative model by mapping images to estimates of  $t_i$  and  $z_i$ . The encoder is structured sequentially with the following components. A Convolutional Neural Network (CNN) containing 5 blocks, each consisting of 2 *Conv3x3* layers, followed by batch normalization and average pooling by a factor of 2. The filter numbers for the convolutional layers in the blocks are set as [32, 64, 128, 256, 512]. The architecture of the CNN is inspired from the first layers of VGG16 [35], known to perform well on visual tasks. A *translation* Multi-Layer Perceptron (MLP) with 2 hidden layers of sizes [512, 256] that outputs a feature  $t_i$  of dimension 2. A *conformation* MLP with 2 hidden layers of sizes [512, 256] that maps  $y_i$  to  $z_i$  of dimension  $d$ , which is the number of normal modes used.

**Generative Model.** The generative model implemented in the decoder is a simulation of the image formation model (Section 2.1). *Pre-processing.* A reference atomic model  $X^{(0)}$  is fed to ProDy [3]

to pre-compute the Hessian matrix  $H_{CG}$  of the CG model.  $H_{CG}$  is diagonalized on the GPU [27] and its first  $d$  eigenvectors are interpolated on the remaining set of atoms with PyKeOps [5] to yield  $U_d$ . The electron form factors [7] are retrieved with Gemmi [42]. For each image  $i$ , the rotation matrix and CTF are provided either by the simulator or by an external software like RELION [32] or cryoSPARC [29]. *Training.* The forward pass is differentiable with respect to  $z_i$  and  $\mathbf{t}_i$ . For each image  $i$ ,  $X^{(0)}$ ,  $\tilde{U}_d$  and  $z_i$  are combined according to Eq.2. The resulting atomic coordinates are rotated by  $R_i$  via a matrix multiplication. A fixed grid of  $D^2$  coordinates on the x-y plane in real space is used to evaluate Eq.3. The resulting image is Fourier transformed and element-wise multiplied by  $\hat{T}_{\mathbf{t}_i}$  and  $C_i$ . A final inverse Fourier transform step is carried out.

**Training Procedure.** For all described datasets, the data was split into training and test set with a 90-10 ratio, resulting in 45,000 images in the training set and 5,000 images in the test set. Training was carried with the following parameters: we use the Adam optimizer [16] with a learning rate of  $1 \times 10^{-3}$ , minibatch size of 256 images. Each training run was stopped after 200 epochs, each taking approximately 10 minutes on 4 Nvidia Tesla A100 GPUs. Inference results below are reported on the test set.

### 3 Results

We generated a 50-frame trajectory between the atomic models of the open (4AKE [23]) and closed (1AKE [22]) forms of Adenylate Kinase (AK) using the *morph* tool in PyMOL [33]. For each frame, 10,000 images were generated using our simulator, each image is  $192 \times 192$  with a  $0.8\text{\AA}$  pixel size. Particle poses were randomly sampled from a uniform distribution on  $SO(3)$ , and CTF defocus values were randomly sampled from a Log-normal distribution with  $\mu = 1\text{ }\mu\text{m}$  and  $\sigma = 0.3\text{ }\mu\text{m}$ . Gaussian noise was added to each image with a Signal-to-Noise Ratio (SNR) set at  $-20$  dB. A variable number of images for each frame were then combined to generate 3 datasets, each consisting of 50,000 images, displaying different distributions (following the approach of [43]): (i) *uniform* — approximately the same number of images across frames, (ii) *continuous* — 3 overlapping Gaussians, and (iii) *discontinuous* — 3 non-overlapping Gaussians.

We tested our proposed architecture on each dataset, using AK in the open form a  $X^{(0)}$ . For each image  $i$ ,  $R_i$  and  $C_i$  are given from the simulation while  $\mathbf{t}_i$  and  $z_i$  are estimated by the encoder  $\psi$ . We ran three simulations in each case, with  $d = 0$  (no deformation),  $d = 1$  and  $d = 16$ . For  $d = 1$ , we compare in Fig.1-bottom the distribution of  $z_i$  to the ground-truth (GT) distribution. For  $d = 16$ , we first perform principal component analysis (PCA) on  $z_i$  and display the first component (PC1) against GT. In every case, PC1 explained more than 90% of the variance in  $z_i$ .

While the qualitative features of the GT distribution is recaptured for  $d = 1$  already, the predicted distribution deteriorates as the GT deviates from the reference model. This suggests that the morphing trajectory between the open and closed forms of AK cannot be fully captured with the first normal mode of open AK. On the other hand, we see a drastic improvement for  $d = 16$ , which suggests not only that the morphing trajectory is mostly captured with the first 16 modes but also that their amplitudes were correctly estimated by the encoder. To quantitatively measure the accuracy of the prediction, we selected 3 predicted atomic models corresponding to the 3 local maxima in the predicted *discontinuous* distribution for  $d = 16$  and compared them to their GT counterpart by measuring the root-mean-square deviation (RMSD) for each pair. As shown in Fig.1, the RMSD increases as the maxima recedes from the reference atomic model but stays below  $1.5\text{ \AA}$ . For reference, the RMSD between the open and closed form is approximately  $7\text{ \AA}$ .

### 4 Discussion

Generative modeling for heterogeneous reconstruction using atomistic representations is a nascent field in cryo-EM [8]. It is a challenging task due to the complex nature of the conformational landscape of biomolecules, and the difficulty in estimating both this landscape and the nuisance variables of the imaging process, such as the pose of the molecule in each image or the  $C_i$  parameters. In this work, we have proposed a framework that tackles the first challenge through a simplified ENM representation of the molecule which enables a drastic reduction in dimensionality of the problem through NMA and thus makes it tractable. We have illustrated the method on synthetic

datasets. The main drawbacks of this approach stem from its strength, namely that it relies on a reference atomic model and that the quality of the deformed models will deteriorate as they are pushed further away from the reference model—a known limitation of ENMs and NMA. These limitations are mitigated by the fact that the Protein Data Bank [4] and the AlphaFold Data Base [38] provide relevant starting models. While our proposed method does not estimate the orientation of the molecule, it does estimate the translational component of the pose in each image. Future work includes demonstrating its capability on experimental datasets and tackling the second challenge with joint estimation of the molecule’s conformation and pose.

## Acknowledgments and Disclosure of Funding

We thank Mike Dunne for supporting this project. This work was supported by the U.S. Department of Energy, under DOE Contract No. DE-AC02-76SF00515, the LCLS seed grant "From Atomic Models to Noisy Images and Back Again: End-to-end Differentiable SPI simulators for CryoEM and XFELs" (PI: GW), the SLAC LDRD project "AtomicSPI: Learning atomic scale biomolecular dynamics from single-particle imaging data" (PI: FP, YN). N.M. acknowledges support from the National Institutes of Health (NIH), grant No. 1R01GM144965-02. We acknowledge the use of the computational resources at the SLAC Shared Scientific Data Facility (SDF).

## References

- [1] Bruce Alberts. *Molecular biology of the cell*. WW Norton & Company, 2017.
- [2] Ali Rana Atilgan, SR Durell, Robert L Jernigan, Melik C Demirel, O Keskin, and Ivet Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical journal*, 80(1):505–515, 2001.
- [3] Ahmet Bakan, Lidio M Meireles, and Ivet Bahar. Prody: protein dynamics inferred from theory and experiments. *Bioinformatics*, 27(11):1575–1577, 2011.
- [4] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.
- [5] Benjamin Charlier, Jean Feydy, Joan Alexis Glaunès, François-David Collin, and Ghislain Durif. Kernel operations on the gpu, with autodiff, without memory overflows. *Journal of Machine Learning Research*, 22(74):1–6, 2021.
- [6] Muyuan Chen and Steven J. Ludtke. Deep learning-based mixed-dimensional gaussian mixture model for characterizing variability in cryo-em. *Nature Methods*, 18(8):930–936, Jul 2021.
- [7] C Colliex, JM Cowley, SL Dudarev, M Fink, J Gjønnes, R Hilderbrandt, A Howie, DF Lynch, LM Peng, G Ren, et al. Electron diffraction. 2006.
- [8] Claire Donnat, Axel Levy, Frederic Poitevin, Ellen Zhong, and Nina Miolane. Deep generative modeling for volume reconstruction in cryo-electron microscopy, 2022.
- [9] Robert M Glaeser, Eva Nogales, and Wah Chiu. *Single-particle Cryo-EM of biological macromolecules*. IOP Publishing, 2021.
- [10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [11] Harshit Gupta, Michael T McCann, Laurene Donati, and Michael Unser. Cryogan: a new reconstruction paradigm for single-particle cryo-em via deep adversarial learning. *IEEE Transactions on Computational Imaging*, 7:759–774, 2021.
- [12] Ilyes Hamitouche and Slavica Jonic. Deephemma: Resnet-based hybrid analysis of continuous conformational heterogeneity in cryo-em single particle images. *Frontiers in Molecular Biosciences*, 2022.
- [13] David Herreros, Roy R. Lederman, James Krieger, Amaya Jiménez-Moreno, Marta Martínez, David Myška, David Strelak, Jiri Filipovic, Ivet Bahar, Jose Maria Carazo, and Carlos Oscar S. Sanchez. Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3d zernike polynomials. *IUCrJ*, 8(6):992–1005, October 2021.

- [14] Qiyu Jin, Carlos Oscar S. Sorzano, José Miguel de La Rosa-Trevín, José Román Bilbao-Castro, Rafael Nunez-Ramirez, Oscar Llorca, Florence Tama, and Slavica Jonic. Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. *Structure*, 22(3):496–506, 2014.
- [15] Kristoffer E Johansson and Kresten Lindorff-Larsen. Structural heterogeneity and dynamics in protein evolution and design. *Current Opinion in Structural Biology*, 48:157–163, 2018.
- [16] Diederik P. Kingma and Jimmy Lei Ba. Adam: A Method for Stochastic Optimization. In *Proc of the 3rd International Conference for Learning Representations*, 2015.
- [17] Lev Davidovich Landau and Evgenii Mikhailovich Lifshitz. *Statistical Physics: Volume 5*, volume 5. Elsevier, 2013.
- [18] Michael Levitt. Birth and future of multiscale modeling for macromolecular systems (nobel lecture). *Angewandte Chemie International Edition*, 53(38):10006–10018, 2014.
- [19] Michael Levitt, Christian Sander, and Peter S Stern. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of molecular biology*, 181(3):423–447, 1985.
- [20] Axel Levy, Frédéric Poitevin, Julien Martel, Youssef Nashed, Ariana Peck, Nina Miolane, Daniel Ratner, Mike Dunne, and Gordon Wetzstein. Cryoai: Amortized inference of poses for ab initio reconstruction of 3d molecular volumes from real cryo-em images. *arXiv preprint arXiv:2203.08138*, 2022.
- [21] Nina Miolane, Frédéric Poitevin, Yee-Ting Li, and Susan Holmes. Estimation of orientation and camera parameters from cryo-electron microscopy images with variational autoencoders and generative adversarial networks. *CVPR conference, Workshop on Computer Vision for Microscopy Image Analysis*, 2019.
- [22] 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: A model for a catalytic transition state. *Journal of molecular biology*, 224(1):159–177, 1992.
- [23] CW Müller, GJ Schlauderer, Jochen Reinstein, and Georg E Schulz. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure*, 4(2):147–156, 1996.
- [24] Youssef S. G. Nashed, Frederic Poitevin, Harshit Gupta, Geoffrey Woollard, Michael Kagan, Chuck Yoon, and Daniel Ratner. End-to-end simultaneous learning of single-particle orientation and 3d map reconstruction from cryo-electron microscopy data, 2021.
- [25] Frank Noé and Cecilia Clementi. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Current opinion in structural biology*, 43:141–147, 2017.
- [26] Tosiyuki Noguti and Nobuhiro Gō. Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature*, 296(5859):776–778, 1982.
- [27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [28] Ali Punjani and David J. Fleet. 3d flexible refinement: Structure and motion of flexible proteins from cryo-em. *bioRxiv*, 2021.
- [29] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 2017.
- [30] Dan Rosenbaum, Marta Garnelo, Michal Zielinski, Charlie Beattie, Ellen Clancy, Andrea Huber, Pushmeet Kohli, Andrew W. Senior, John Jumper, Carl Doersch, S. M. Ali Eslami, Olaf Ronneberger, and Jonas Adler. Inferring a continuous distribution of atom coordinates from cryo-em images using vaes. *CoRR*, abs/2106.14108, 2021.
- [31] Sjors H.W. Scheres. A bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, 2012.
- [32] Sjors H.W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 2012.

- [33] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.
- [34] Ch Schütte, Alexander Fischer, Wilhelm Huisinga, and Peter Deuffhard. A direct approach to conformational dynamics based on hybrid monte carlo. *Journal of Computational Physics*, 151(1):146–168, 1999.
- [35] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [36] Florence Tama and Y-H Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein engineering*, 14(1):1–6, 2001.
- [37] Monique M Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters*, 77(9):1905, 1996.
- [38] Mihaly Varadi, Stephen Anyango, Mandar Deshpande, Sreenath Nair, Cindy Natassia, Galabina Yordanova, David Yuan, Oana Stroe, Gemma Wood, Agata Laydon, et al. Alphafold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research*, 50(D1):D439–D444, 2022.
- [39] Miloš Vulović, Lenard M Voortman, Lucas J van Vliet, and Bernd Rieger. When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-em. *Ultramicroscopy*, 136:61–66, 2014.
- [40] Miloš Vulović, Raimond B.G. Ravelli, Lucas J. van Vliet, Abraham J. Koster, Ivan Lazić, Uwe Lücken, Hans Rullgård, Ozan Öktem, and Bernd Rieger. Image formation modeling in cryo-electron microscopy. *Journal of Structural Biology*, 183(1):19–32, 2013.
- [41] Christoph Wigge, Aleksandar Stefanovic, and Mazdak Radjainia. The rapidly evolving role of cryo-em in drug design. *Drug Discovery Today: Technologies*, 38:91–102, 2020.
- [42] Marcin Wojdyr. Gemmi: A library for structural biology. *Journal of Open Source Software*, 7(73):4200, 2022.
- [43] Ellen D Zhong, Tristan Bepler, Joseph H Davis, and Bonnie Berger. Reconstructing continuously heterogeneous structures from single particle cryo-em with deep generative models. *arXiv preprint arXiv:1909.05215*, 2019.
- [44] Ellen D. Zhong, Adam Lerer, Joseph H. Davis, and Bonnie Berger. Exploring generative atomic models in cryo-em reconstruction, 2021.