PHYSICS-INFORMED GEOMETRIC REGULARIZATION OF HETEROGENEOUS RECONSTRUCTIONS IN CRYO-EM

Victor Prins
Amsterdam Machine Learning Lab
University of Amsterdam
victor.prins@outlook.com

Willem Diepeveen
Faculty of Mathematics
University of Cambridge
wd292@cam.ac.uk

Erik J. Bekkers
Amsterdam Machine Learning Lab
University of Amsterdam
e.j.bekkers@uva.nl

Ozan Öktem
Department of Mathematics
KTH Royal Institute of Technology
ozan@kth.se

ABSTRACT

Many proteins are flexible and occur in a continuum of 3D conformations. A protein's 3D conformation is the main determinant of its biological function, and is therefore of paramount importance to fields such as drug development and biomolecular engineering. Cryogenic electron microscopy (cryo-EM) enables the reconstruction of protein conformations. Because cryo-EM reconstruction is a fundamentally underdetermined problem, neural networks that incorporate prior knowledge such as physical constraints into the training process, can theoretically be more effective at reconstructing heterogeneous conformation distributions. We introduce a novel such prior, called the *geometry degradation loss*, grounded in the theory of normal mode analysis. The loss is generally applicable to all proteins and easily integrated into reconstruction algorithms that utilize geometric protein representations. We show on synthetic datasets of the flexible ADK and Nsp13 proteins that the loss greatly improves reconstruction quality and that our network is able to reconstruct both proteins accurately across the full conformation distribution. These results are further evidence that geometric cryo-EM reconstruction networks have large potential that can be tapped with the introduction of geometric priors. Code is published at https://github.com/VictorPrins/geometric-heterogeneous-cryoEM-reconstruction.

1 INTRODUCTION

Cryogenic electron microscopy (cryo-EM) is a technique that allows one to recover the conformations and dynamics of proteins, which are the prime determinants of protein function in a biological system. Conformationally heterogeneous protein structures, however, are notoriously hard to reconstruct, even from high-resolution cryo-EM datasets. This challenge originates from the inherent underdetermination of the cryo-EM reconstruction problem with finite data (Bendory et al., 2020). The solution is twofold: 1) reducing the degree of underdetermination by collecting more data and improving data quality and 2) developing reconstruction algorithms that can optimize in non-convex solution spaces. This work makes contributions to the latter area using a neural network regularized by a novel geometry degradation loss.

In an ideal cryo-EM experiment, N instances of the same protein, submerged in a thin slab of vitreous ice, are projected onto N grayscale 2D images. Each image is associated with three random variables: (i) the protein's pose $R_i \in SO(3)$, (ii) a displacement from the image centre $\boldsymbol{d}_i \in \mathbb{R}^2$, and (iii) a conformation parameterized by ϕ_i. All three variables are sampled from unknown distributions. The task is to inversely reconstruct the conformation distribution $\{\phi_i\}$. Variables R_i and $\boldsymbol{d}_i$ are nuisance variables.
Commonly used reconstruction packages, like Scheres (2012) and Punjani et al. (2017), are capable of producing high-resolution reconstructions for homogeneous proteins, which are rigid in structure.

They can also produce finite sets of conformations for heterogeneous proteins, thus approximating a discrete conformation distribution.
Neural networks are continuous estimators, and are known to be effective optimizers in high-dimensional non-convex solution spaces. This makes neural networks potentially powerful for the reconstruction of continuous conformation distributions. Zhong et al. (2021a) were the first to demonstrate a neural heterogeneous reconstruction algorithm on experimental (i.e. non-synthetic) data. They assumed $(R_i, \boldsymbol{d}_i)$ be known, and these assumptions were later relaxed by Zhong et al. (2021b) and Levy et al. (2022).

Geometric protein representation All methods discussed so far use a volumetric protein representation. We call a representation volumetric if it constitutes a 3D map of the electrostatic potential of the protein. The specifics of implementations vary; some methods use voxel grids whereas others use volumes defined over a continuous range. From the 3D map, one can build an atomic model of the protein, which is a geometric representation as a point cloud where the points encode the positions of atoms that make up the protein (Vulović et al., 2013). Reconstructing this geometric representation directly from cryo-EM data has multiple theoretical benefits. Firstly, it enables incorporation of physical and chemical priors into the solution search. The number and identity of protein residues, the length of covalent bonds, and the permissible dihedral angles (Ramachandran et al., 1963) are a few well-defined priors that any reconstructed conformation should adhere to. These constraints are easily enforced in a geometric protein representation, whilst they are ignored by volumetric methods. Secondly, it makes redundant the atomic model building step required after a volumetric reconstruction.
This insight has given rise to work on geometric representations for cryo-EM reconstruction, concurrently introduced by Rosenbaum et al. (2021) and Zhong et al. (2021c). Both papers represent the protein as a point cloud, initialized at a known conformation, which the network deforms into other conformations of the distribution. The only structural prior used by the papers is a loss on the distance between residues. The authors observe that geometric reconstruction networks are much harder to optimize than their volumetric counterparts, and call for the development of novel structural constraints to close the existing performance gap—a call that this work endeavors to address.

Contributions We introduce a *geometry degradation loss* function, which approximates the potential energy of a protein and therefore guides the solution search towards conformations that are energetically likely to occur. We further use a *learnable output normalization*, specifically helpful for reconstruction under extreme noise conditions. An ablation study shows that the geometry degradation loss greatly improves our neural network's reconstruction performance, reaching EMD-RMSD values in the range of 2-3Å under realistic noise conditions.

2 METHOD

Our neural reconstruction network (Figure 1) is an autoencoder architecture that operates by encoding a cryo-EM projection to a latent variable z, which represents the conformation, then generating a protein conformation $X(z)$, and finally producing a projection from this conformation by simulating the cryo-EM image formation process in a differentiable manner. This approach is based on the assumption that the full conformation distribution can be embedded into a low-dimensional latent space (Das et al., 2006) —we use[1] $z \in \mathbb{R}^8$. The network is optimized by minimizing the following loss: $\mathcal{L} = \mathcal{L}_{\text{img}} + \alpha\mathcal{L}_{\text{geom}} + \beta\mathcal{L}_{\text{bond}} + \gamma\mathcal{L}_{\text{centering}}$, where the primary loss component $\mathcal{L}_{\text{img}}$ is the pixelwise L2 loss between the input images and the reconstructions produced by the network, $\mathcal{L}_{\text{geom}}$ is our novel geometry degradation loss (see the paragraph below), $\mathcal{L}_{\text{bond}} = \frac{1}{N-1}\sum_{i=1}^{N-1}(\|\boldsymbol{p}_i - \boldsymbol{p}_{i+1}\| - 3.8)^2$ represents the constraint that neighbouring residues maintain a fixed distance of 3.8 Å (Chakraborty et al., 2013), and $\mathcal{L}_{\text{centering}} = (\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{p}_i)^2$ keeps the generated conformations centered around the origin and serves to remove globally shifted (and thus equivalent) conformations from the solution space (Rosenbaum et al., 2021). The weighting factors[2] are set to $\alpha = 0.05$, $\beta = 0.005$, and $\gamma = 0.1$.

[1] Levy et al. (2022) also use a conformation latent of dimensionality 8.

[2] The values of β and γ are inspired by Rosenbaum et al. (2021). The exact values of (α, β, γ) were obtained through a limited hyperparameter search.

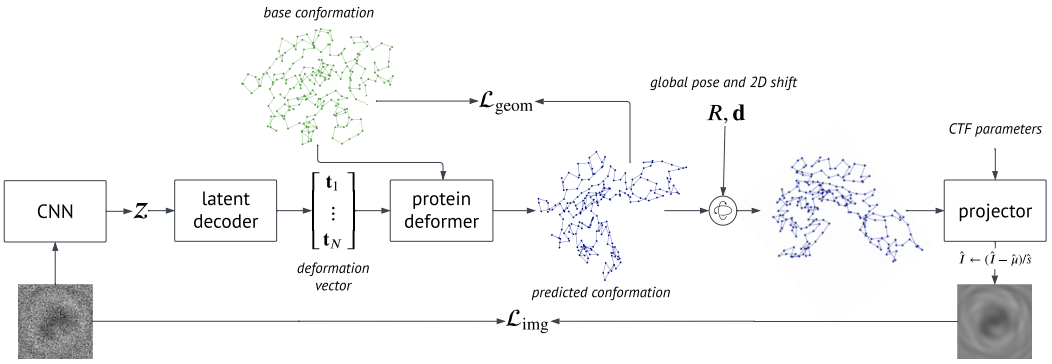


Figure 1: The network architecture: an autoencoder that learns the conformation distribution by producing conformations conditional on each input image. Details are in appendix A (CNN), appendix E (latent decoder), section 2 (protein deformer), and appendix B (projector).

Protein deformer The network produces conformations by deforming a *base conformation*, initialized at a known conformation, through the prediction of translations for each point in the point cloud. Let the base conformation be defined as $X^0 = [\boldsymbol{p}_1, ..., \boldsymbol{p}_N] \in \mathbb{R}^{3\times N}$, where $\boldsymbol{p}_i \in \mathbb{R}^3$ is the position of residue i for a protein consisting of N residues. The generative latent decoder outputs a $3N$-dimensional vector that is reshaped into a deformation matrix $T(z) = [\boldsymbol{t}_1(z), ..., \boldsymbol{t}_N(z)] \in \mathbb{R}^{3\times N}$ where $\boldsymbol{t}_i(z) \in \mathbb{R}^3$ is the displacement vector for residue i. The predicted conformation, conditioned on the conformation latent z, can now be expressed as $X(z) = X^0 + T(z)$.

Geometry degradation loss The introduction of the *geometry degradation loss* is the main contribution of this paper. This function, which approximates the difference in potential energy between two conformations, was developed by Diepeveen et al. (2023) to address the locality and linearity limitations of normal mode analysis using Riemannian manifold theory. Observing that high-potential conformations are physically less likely to occur and should therefore be discouraged by the network, it becomes apparent that potentials can serve as loss functions. We present the core idea here but refer to the original paper for the mathematical derivation. It was shown by Tirion (1996); Doruker et al. (2000) that the potential of a protein conformation w.r.t. a base conformation can be approximated by $E = \sum_{(a,b)} \frac{C}{2}(\|\boldsymbol{r}_{a,b}\| - \|\boldsymbol{r}^0_{a,b}\|)^2$, where the summation is over all atom pairs within a specified cutoff distance, $\boldsymbol{r}_{a,b}$ is the displacement vector from atom a to b, superscript 0 refers to the base conformation, and C is a constant. This function thus only incorporates local interactions (due to the cutoff distance), and approximates potential energy to have a linear relation to distance, an assumption which does not hold in general. To address the locality and linearity issues of the equation, it is modified in two ways: (i) The loss becomes a function of the *relative* change in pairwise distance, thereby giving atoms which are further apart in the base conformation greater liberty of movement. (ii) The change in pairwise distance is wrapped into a logarithm, which a) allows the network to separate nearby atoms (due to logarithm's slow rate of increase) if this leads to a much better fit in terms of the other loss functions, and b) prevents the network from producing (physically impossible) conformations where the positions of two atoms coincide, since $\lim_{x\to 0} \log^2(x/c) = \infty$. This concludes the geometry degradation loss, which is applied across all pairs of atoms (a, b) in the protein[3][4]:

$$\mathcal{L}_{\text{geom}} = \frac{N(N-1)}{2} \sum_{(a,b)} \log^2 \left(\frac{\|\boldsymbol{r}_{a,b}\|^2}{\|\boldsymbol{r}^0_{a,b}\|^2} \right). \tag{1}$$

[3]Note that the vector norms are squared purely for computational efficiency reasons; the squares can be extracted as a constant and subsumed into the weighting of this loss term.

[4]The loss scales $\mathcal{O}(N^2)$ in space and computational complexity. To mitigate the memory issue that arises for larger complexes, the loss can be computed across a subset of all N atoms. This work coarsens the protein and computes $\mathcal{L}_{\text{geom}}$ across all pairs of Cα atoms. Another option is to restrict the computation to the k-nearest neighbours of each atom, reducing the complexity to $\mathcal{O}(kN)$.

In short, the geometry degradation loss guides the network optimization towards conformations that are energetically close to the base conformation. Note that this puts importance on the base conformation: the geometry degradation loss will theoretically only aid the reconstruction of likely conformations if the base conformation itself is a relatively likely (i.e. stable) conformation.

Learnable output normalization For L2 image loss computation, input and output must be normalized to distributions with equal statistics. In cryo-EM, this is nontrivial to accomplish due to the extreme noise of unknown strength. We estimate output normalization statistics with a neural network that produces $(\hat{\mu}, \hat{s})$ with which projections are normalized: $\hat{I} \leftarrow (\hat{I} - \hat{\mu})/\hat{s}$. Learnable output normalization improved network performance significantly, compared to fixed normalization statistics estimated pre-training. Full details are in appendix F.

3 RESULTS

Data The ground truth conformation trajectories are simulated using molecular dynamics and the cryo-EM projections are generated by Parkhurst et al. (2021), a physically realistic cryo-EM simulation software package. We consider two proteins that undergo conformational changes, namely the 214-residue ADK protein and the larger 590-residue SARS-CoV-2 non-structural protein 13 (Nsp13). The ADK protein undergoes a relatively simple hinge movement (Figure 6a) and we sample this conformation trajectory at 102 points (Seyler et al., 2015). In contrast, Nsp13 exhibits a more disordered motion (Figure 7) and this conformation trajectory is sampled at 200 points (Shaw, 2020). Each dataset contains a uniform distribution across all conformations, with approximately 600 images per conformation. The poses R_i are sampled uniformly from $SO(3)$ and assumed known, and all particles are centered ($\boldsymbol{d}_i = \mathbf{0}$). Images have a pixel size of 1 Å. CTF parameters are assumed known. See appendix H for full details.

Benchmarks The literature on heterogeneous cryo-EM reconstruction methods using geometric representations is relatively nascent, making direct comparison with other papers challenging[5]. We mention three relevant data points in order to facilitate interpretation of our results. First, the distance between neighbouring Cα atoms of 3.8 Å (Chakraborty et al., 2013) provides a natural baseline; an (EMD-)RMSD around this value indicates a high degree of backbone similarity. Second, Rosenbaum et al. (2021) achieves an EMD-RMSD of 3.8 Å on a 282-residue protein at an electron dose of $1000\,\mathrm{e}/\text{Å}^2$, using 63,000 images. Unlike us, however, they do also estimate poses, thus introducing additional complexity. Third, Nashed et al. (2022) report an RMSD of 1.3 Å on the ADK protein, using 50,000 images. However, they used a dataset generated by their own reconstruction model, significantly reducing the complexity of the reconstruction problem (Wirgin, 2004). In conclusion, reconstructions with errors below 3.8 Å are in line with the current state of research.

Experimental setup The network is optimized with Adam (Kingma & Ba, 2017) at a constant learning rate of 10^{-5}. The CTF is oversampled at 300×300 pixels to prevent CTF aliasing and is applied to images after zero-padding those to the same size. The base conformation of the protein is initialized at the first conformation of the trajectory. We use a batch size of 128, and a conformation latent z of dimension 8. Training is performed on a single A100 Nvidia GPU and takes 10 minutes on the ADK dataset and 4 - 24 hours on the Nsp13 dataset.

Evaluation EMD-RMSD and RMSD are used as the quantitative metrics to measure heterogeneous reconstruction accuracy. EMD-RMSD is the Earth Mover's Distance between the reconstructed and the ground truth conformation distributions, using the Root Mean Squared Deviation of Cα atoms as the metric to minimize. This generalizes the RMSD metric, commonly used to measure similarity between pairs of structures, to distributions of structures. Whilst our network achieves very similar values for RMSD and EMD-RMSD, the latter is conceptually a more truthful measurement of the objective of this work, which is to reconstruct a *distribution* of conformations. Rosenbaum et al. (2021) showed that a model may be able to accurately reconstruct a distribution of conformations, even if it can only poorly reconstruct individual conformations given a single input image.

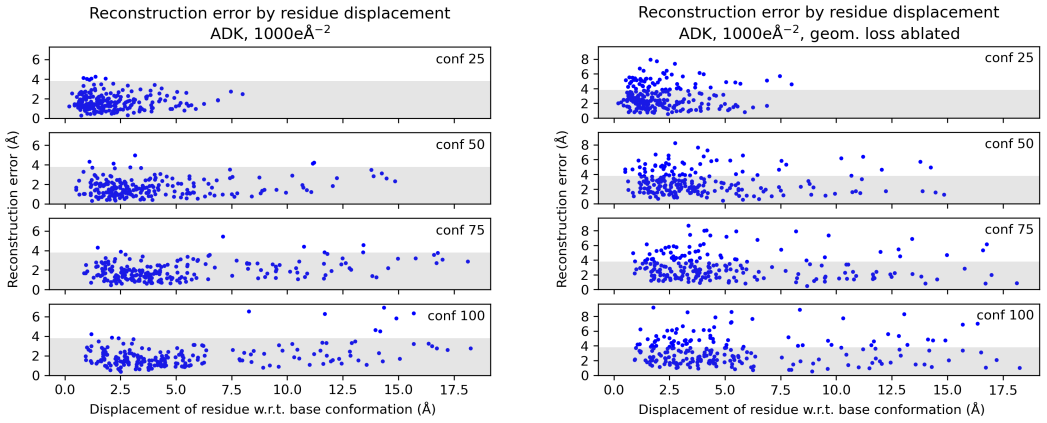
[5] Additionally, few papers in this literature have open sourced code or datasets.

Table 1: Heterogeneous reconstruction results in terms of EMD-RMSD (Å). The network is trained on synthetic datasets of the ADK and Nsp13 proteins with various electron doses ($50\,\mathrm{e}/\mathrm{Å}^2$ corresponds to a realistic noise level). The stark increase in the reconstruction error caused by ablating the geometry degradation loss (in all cases except for ADK at $50\,\mathrm{e}/\mathrm{Å}^2$) demonstrates the significant benefit of the geometry degradation loss.

electron dose ($\mathrm{eÅ}^{-2}$)	ADK		Nsp13	
	base	geom. loss ablated	base	geom. loss ablated
1000	1.9	2.5	2.5	4.0
100	2.1	2.5	2.7	4.5
50	2.5	2.5	2.8	4.4

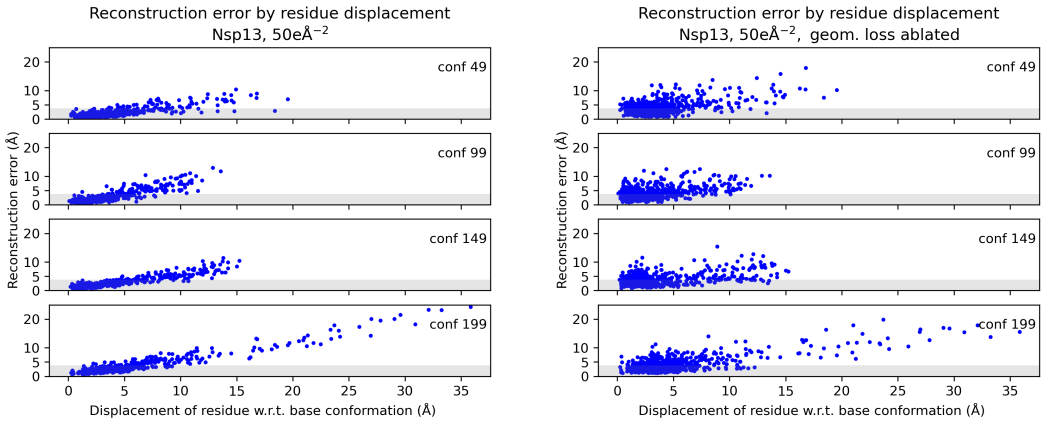
Given a validation set and a ground truth set of structures of equal size (we use 4000, uniformly distributed across all conformations), the EMD-RMSD is the minimum average RMSD achievable by finding the optimal one-to-one pairing between validation structures and ground truth structures. Full details are in appendix G.

Results The main results in terms of EMD-RMSD are displayed in Table 1. The corresponding results in terms of RMSD are in Table 2. When including the geometry degradation loss, our network achieves EMD-RMSD values between $1.9\,\text{Å}$ and $2.8\,\text{Å}$, compared to EMD-RMSD values of $4.4\,\text{Å}$ for ADK and $5.0\,\text{Å}$ for Nsp13 at random initialization (before training). Especially for the Nsp13 protein, the geometry degradation loss is essential to getting meaningful reconstruction performance; without it the EMD-RMSD improves no more than $1\,\text{Å}$ compared to random initialization. Figures 2 and 3 show the reconstruction performance for each residue separately, and compare these results between the easiest (ADK at $1000\,\mathrm{e}/\mathrm{Å}^2$) and the hardest (Nsp13 at $50\,\mathrm{e}/\mathrm{Å}^2$) dataset. Appendix I contains additional figures.



(a) (b)

Figure 2: Ablation experiment: the network in its base configuration (left) correctly positions both flexible and stable regions of the protein, whereas ablating the geometry degradation loss (right) leads to significantly poorer reconstructions. Each dot represents one of the 214 residues in the ADK protein. The y-axis is the distance (Å) between the ground truth position of the residue and the reconstructed position, averaged over the validation set. Residues within the bottom $3.8\,\text{Å}$ (marked grey) can be considered excellently fitted. The x-axis is the distance between the ground truth position of the residue and the position of the residue in the base conformation. Each plot corresponds to one conformational state of ADK. The absence of an increasing trend line in the plots shows that our network is able to fit stationary as well as mobile residues.



(a) (b)

Figure 3: Ablation experiment: without the geometry degradation loss (right figure), the reconstruction error is significantly greater for a large share of the residues. This plot is an analogous visualization to Figure 2, but for the Nsp13 protein at an electron dose of $50\,\mathrm{e}/\text{Å}^2$. At this noise level, a number of residues, particularly those with very large displacements, are not positioned accurately and keep a sizeable error. However, the benefit of the geometry degradation loss is very stark, as can be seen from the thick blob of dots in the right figure, which for a large part falls outside the $3.8\,\text{Å}$ band.

4 Discussion

We have introduced a novel *geometry degradation loss* and demonstrated how it enables the accurate reconstruction of a continuous conformation distribution. To our knowledge, this is the first demonstration of a significant reconstruction improvement attributable to a single physical and geometric prior. We have additionally introduced a new reconstruction network with a learnable output normalization that is specifically beneficial in the uniquely high-noise conditions of cryo-EM.
Our network is able to fit highly mobile as well as stable parts of the protein (Figure 2a) and learns to embed the conformation distribution manifold in a low-dimensional (8, in this paper) latent space (Figure 10). These results still hold, but are weaker for the more complex Nsp13 conformation distribution; the reconstruction errors of some sections remain high (see Figure 3a and appendix I). The Nsp13 network also has a less interpretable latent space; PCA doesn't quickly reveal a structure in this case. However, this is likely the result of the higher dimensional motion of the Nsp13 protein, which cannot be embedded in a 1-dimensional space or visualized with 1-dimensional PCA.

Like all geometric cryo-EM reconstruction papers known to the authors, the results are only demonstrated on a synthetic dataset. We further assume that poses are uniformly distributed and known, and assume a uniform conformation distribution over the MD-simulated states. The dataset also contains no junk particles. These are all assumptions that should be relaxed before geometric networks will be able to achieve convincing results on experimental datasets. However, the validation of our method on the complex Nsp13 protein at realistic noise levels is a promising result that suggests resilience of our method under even more challenging conditions. Another limitation is that it is unclear to what extent the geometry degradation loss remains beneficial for the reconstruction of extremely flexible complexes. Considering that the loss is an approximation of the transition energy between conformational states, it could degrade for particularly challenging energy landscapes.

Future work should continue to leverage the benefits of geometric models—the geometry degradation loss is only a first demonstration of the power of priors in geometric reconstruction methods.

Increasingly more attention will shift from homogeneous reconstruction to heterogeneous reconstruction, which offers the more complete view and understanding of the function of biology's most vital particles. Numerous influential fields, including drug development and biomolecular engineering stand to benefit from advancements in this area. We hope this paper is a step towards the ultimate objective of comprehensively unraveling protein heterogeneity.

REFERENCES

Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-Particle Cryo-Electron Microscopy: Mathematical Theory, Computational Challenges, and Opportunities. *IEEE Signal Processing Magazine*, 37(2):58–76, March 2020. ISSN 1053-5888, 1558-0792. doi: 10.1109/MSP.2019.2957822.

Sandeep Chakraborty, Ravindra Venkatramani, Basuthkar J. Rao, Bjarni Asgeirsson, and Abhaya M. Dandekar. Protein structure quality assessment based on the distance profiles of consecutive backbone Cα atoms. *F1000Research*, 2:211, December 2013. ISSN 2046-1402. doi: 10.12688/f1000research.2-211.v3.

Payel Das, Mark Moll, Hernán Stamati, Lydia E. Kavraki, and Cecilia Clementi. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proceedings of the National Academy of Sciences*, 103(26):9885–9890, June 2006. doi: 10.1073/pnas.0603553103.

Willem Diepeveen, Carlos Esteve-Yagüe, Jan Lellmann, Ozan Öktem, and Carola-Bibiane Schönlieb. Riemannian geometry for efficient analysis of protein dynamics data, August 2023.

Pemra Doruker, Ali Rana Atilgan, and Ivet Bahar. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to α-amylase inhibitor. *Proteins: Structure, Function, and Bioinformatics*, 40(3):512–524, 2000. ISSN 1097-0134. doi: 10.1002/1097-0134(20000815)40:3⟨512::AID-PROT180⟩3.0.CO;2-M.

Dan Hendrycks and Kevin Gimpel. Gaussian Error Linear Units (GELUs), June 2023.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pp. 448–456. pmlr, 2015.

Roy Jonker and Ton Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. In *DGOR/NSOR: Papers of the 16th Annual Meeting of DGOR in Cooperation with NSOR/Vorträge Der 16. Jahrestagung Der DGOR Zusammen Mit Der NSOR*, pp. 622–622. Springer, 1988.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32 (5):922–923, 1976.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017.

Axel Levy, Gordon Wetzstein, Julien Martel, Frederic Poitevin, and Ellen D. Zhong. Amortized Inference for Heterogeneous Reconstruction in Cryo-EM, October 2022.

Youssef Nashed, Ariana Peck, Julien Martel, Axel Levy, Bongjin Koo, Gordon Wetzstein, Nina Miolane, Daniel Ratner, and Frédéric Poitevin. Heterogeneous reconstruction of deformable atomic models in Cryo-EM, September 2022.

James M Parkhurst, Maud Dumoux, Mark Basham, Daniel Clare, C Alistair Siebert, Trond Varslot, Angus Kirkland, James H Naismith, and Gwyndaf Evans. Parakeet: A digital twin software pipeline to assess the impact of experimental parameters on tomographic reconstructions for cryo-electron tomography. *Open Biology*, 11(10):210160, 2021.

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. ISSN 1548-7105. doi: 10.1038/nmeth.4169.

G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, July 1963. ISSN 0022-2836. doi: 10.1016/S0022-2836(63)80023-6.

Alexis Rohou and Nikolaus Grigorieff. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221, November 2015. ISSN 1047-8477. doi: 10.1016/j.jsb.2015.08.008.

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, June 2021.

Sjors H.W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, December 2012. ISSN 10478477. doi: 10.1016/j.jsb.2012.09.006.

Sean L. Seyler, Avishek Kumar, M. F. Thorpe, and Oliver Beckstein. Path Similarity Analysis: A Method for Quantifying Macromolecular Pathways. *PLOS Computational Biology*, 11(10): e1004568, October 2015. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004568.

DE Shaw. Molecular dynamics simulations related to Sars-Cov-2. *DE Shaw Research Technical Data*, 2020.

Monique M. Tirion. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*, 77(9):1905–1908, August 1996. doi: 10.1103/PhysRevLett.77.1905.

Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(04):376–380, 1991.

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, July 2013. ISSN 1047-8477. doi: 10.1016/j.jsb.2013.05.008.

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, January 2014. ISSN 0304-3991. doi: 10.1016/j.ultramic.2013.08.002.

Armand Wirgin. The inverse crime, January 2004.

Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, 18(2):176–185, February 2021a. ISSN 1548-7105. doi: 10.1038/s41592-020-01049-4.

Ellen D. Zhong, Adam Lerer, Joseph H. Davis, and Bonnie Berger. CryoDRGN2: Ab initio neural reconstruction of 3D protein structures from real cryo-EM images. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4046–4055, Montreal, QC, Canada, October 2021b. IEEE. ISBN 978-1-66542-812-5. doi: 10.1109/ICCV48922.2021.00403.

Ellen D. Zhong, Adam Lerer, Joseph H. Davis, and Bonnie Berger. Exploring generative atomic models in cryo-EM reconstruction, July 2021c.

A IMAGE ENCODER

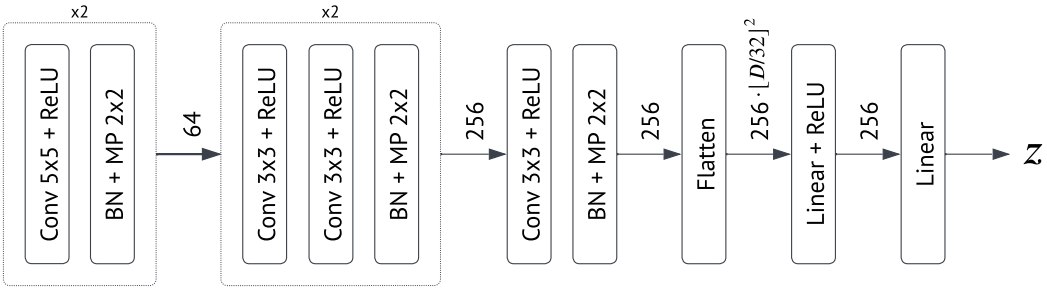


Figure 4: The image encoder used in our network. BN stands for batch normalization Ioffe & Szegedy (2015), and MP is max pooling. The numbers on the arrows after convolutional blocks indicate the number of channels. The numbers on the layers after flattening indicate vector dimensionality. The output of the image encoder is the conformation latent z, with dimensionality 8 in this research.

B IMAGE FORMATION MODEL AND DIFFERENTIABLE PROJECTOR

A transmission electron microscope (TEM) probes a specimen by incident high-energy electrons. These electrons scatter against the electrostatic potential $V(x, y, z)$ that is generated by the atoms in the specimen, which here are the atoms of the protein and the surrounding vitrified aqueous buffer. The first step in simulating the TEM image formation is to model this scattering. If the specimen is weakly scattering, as is the case for proteins in a thin aqueous buffer, then one can express the scattered electron wave in terms of the X-ray transform of the aforementioned electrostatic potential: $I(x, y) = \int_z V(x, y, z)\,\mathrm{d}z$, where the integral runs parallel to the TEM optical axis through the specimen. Next, this electron wave is propagated through the TEM optics, a procedure that is modelled by a convolution in the TEM detector plane with a kernel given by the microscope CTF (Vulović et al., 2014; 2013). The projector of our network relies on this image formation model to efficiently and differentiably produce output images based on the computed protein conformation.

Letting R represent the protein pose in the image, and $\boldsymbol{d}$ represent the global (2D) off-center shift of the protein in the image, our network projects conformations as follows:

$$\boldsymbol{p}_i \leftarrow R\boldsymbol{p}_i + \boldsymbol{d} \;\; i = 1, ..., N \tag{2}$$

$$\hat{I}_{\text{ideal}}(x, y) = \sum_{i=1}^{N} m_i (\sigma_i \sqrt{2\pi})^{-2} e^{-\frac{1}{2}\left(\frac{x - p_{i,0}}{\sigma_i}\right)^2} e^{-\frac{1}{2}\left(\frac{y - p_{i,1}}{\sigma_i}\right)^2} \tag{3}$$

$$\hat{I}_F(k_x, k_y) = CTF(k_x, k_y) \cdot \mathcal{F}\{\hat{I}_{\text{ideal}}(x, y)\}(k_x, k_y) \tag{4}$$

$$\hat{I}_{\text{real}}(x, y) = \mathcal{F}^{-1}\{I_F(k_x, k_y)\}(x, y)\,, \tag{5}$$

where $\boldsymbol{p}_i$ is the residue's position, and σ_i its standard deviation estimated as explained in appendix C. The Gaussian mass $m_i = \sum_{\text{atom} \in i} Z_{\text{atom}}$ is equal to the total number of electrons of the residue (since the image contrast is proportional to the electron density as per the X-ray transform). The CTF is computed in the conventional way, explained in appendix D.

C MODELLING RESIDUES AS GAUSSIANS

We model residues as points, and represent the protein as a point cloud of residues. More fine-grained representations (e.g. as in Zhong et al. (2021c) and Rosenbaum et al. (2021)) allow for potentially more precise projections, but only if the smaller units are also positioned accurately.

This adds significant complexity to the problem[6]. Furthermore it is the backbone movement that is of primary interest to structural biologists.

To account for the varying sizes and shapes of different amino acids, each residue is projected as a Gaussian with a standard deviation unique to its amino acid identity. Let $I_{\nu,R}(x, y)$ be the projection of amino acid $\nu \in \{\text{alanine}, \text{arginine}, ...\}$ in pose $R \in SO(3)$. Let $I_{\nu,R}(x, y)$ additionally be normalized such that its integral is 1. We then estimate σ_ν as the standard deviation that minimizes the L2 loss between the Gaussian $\text{pdf}_\sigma(x, y)$ corresponding to $\mathcal{N}(\mathbf{0}, \text{diag}(\sigma, \sigma))$ and $I_{\nu,R}(x, y)$, integrated across the projection plane $\mathcal{S}$, assuming rotations R uniformly sampled from $SO(3)$:

$$\sigma_\nu = \arg\min_\sigma \int \int_{\mathcal{S}} (\text{pdf}_\sigma(x, y) - E_{R\sim\text{Unif}(SO(3))}[I_{\nu,R}](x, y))^2 \, \text{d}x \, \text{d}y \, . \tag{6}$$

We use a discretized version of this estimator. The projections $I_{\nu,R}(x, y)$ are generated by taking an atomic model of amino acid ν, centering its Cα atom, and projecting all its atoms as spherical Gaussians with unit standard deviation[7] and mass equal to the atomic number. This estimator finds the Gaussians that most closely fit the projections of a residue (using atom-level projections and a (theoretically) infinitesimal pixel size) with random poses. Although $E_{R\sim\text{Unif}(SO(3))}[I_{\nu,R}]$ is not shaped exactly like a Gaussian pdf, we found that it is sufficiently close.

D CTF COMPUTATION

The contrast transfer function (CTF) is the Fourier transform of the convolution kernel used for modelling the TEM optics. It is applied in the detector plane (in Fourier space) through pixel-wise multiplication (equation 4). The CTF is computed as follows Rohou & Grigorieff (2015):

$$CTF(k_x, k_y) = -\sin\left(\pi\lambda\|\boldsymbol{g}\|^2(f - \frac{1}{2}\lambda^2\|\boldsymbol{g}\|^2 C_s) + \arctan(\omega/\sqrt{1-\omega^2})\right) \tag{7}$$

$$f = \frac{1}{2}(f_1 + f_2 + (f_1 - f_2)\cos(2(\alpha_{\boldsymbol{g}} - \alpha_{\text{ast}}))) \tag{8}$$

$$\|\boldsymbol{g}\|^2 = k_x^2 + k_y^2 \tag{9}$$

$$\alpha_{\boldsymbol{g}} = (k_y, k_x) \tag{10}$$

$$\lambda = 12.2639/\sqrt{V + V^2 \cdot 0.97845 \cdot 10^{-6}} \, , \tag{11}$$

where

- (k_x, k_y) are coordinates in frequency space (in units of Å^{-1}) corresponding to pixel (x, y) in real space.
- λ is the wavelength of the electron beam used by the electron microscope, which is a function of the accelerating voltage V (given). Equation 11 is the relativistic version of the *de Broglie wavelength* with physical constants filled out.
- C_s is the spherical aberration (given).
- ω is the amplitude contrast percentage (given).
- f_1 and f_2 are the two defocus values along normal directions (given).

The parameters marked as (given) are either experimental settings, or estimated in postprocessing by the cryo-EM experimentalists, and must be given as part of the dataset.

[6]Potentially even more complexity than the network can solve for. Side chains are small and it is questionable whether realistic heterogeneous datasets contain strong enough a signal to reconstruct their exact positions. Zhong et al. (2021c) found a minor accuracy improvement by modelling side chains separately, but only demonstrated this under low-noise conditions.

[7]Which is a simple yet sufficiently accurate model of the electron density of an atom (Rosenbaum et al., 2021).

E LATENT DECODER

The latent decoder is implemented as a 7-layer MLP with a hidden dimension of 350 and SiLU activations (Hendrycks & Gimpel, 2023) after each linear layer except the last.

F LEARNABLE OUTPUT NORM

Learnable output normalization is implemented using a 3-layer MLP with output dimension 2, hidden dimension 350, and SiLU activations between layers. The last linear layer is interpreted to contain normalization scalars $[\hat{\mu}, \hat{s}]$. The MLP is not conditioned on input data, i.e. it is backpropagated directly.

To ensure stable optimization from the first batch, it helps to additionally use a fixed set of normalization statistics (μ^0, s^0) that has been estimated prior to training. Since the learnable output norm will correct during training, there is no need to use a sophisticated estimation approach for (μ^0, s^0). We estimated μ^0 as the mean across a sample of projections of the base conformation under random poses. We estimated s^0 using a visual comparison between a set of images from the dataset and a set of projections of the base conformation, by trying to scale their intensity levels to the same order of magnitude.

Output projections of the network are then normalized twice: first $\hat{I} \leftarrow (\hat{I} - \mu^0)/s^0$, followed by $\hat{I} \leftarrow (\hat{I} - \hat{\mu})/\hat{s}$. Hence, $\hat{s}$ should be initialized at 1.

G EMD-RMSD

Root-mean-square deviation (RMSD) is a standard similarity metric used in structural biology for two different reconstructions of the same protein $\mathrm{RMSD}(P_1, P_2) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{p}_{i,1} - \boldsymbol{p}_{i,2}\|^2}$ where $\boldsymbol{p}_{i,j}$ is the position vector of point i in protein P_j and the protein is represented using N points (Cα atoms in this work). RMSD enables straightforward comparison between any pair of proteins, but we need a metric that quantifies similarity of two *distributions* of conformations, namely the heterogeneous conformation distribution learnt by the network and the ground truth conformation distribution (which is known for a synthetic dataset). To that end, we use the earth mover's distance (EMD) measure of similarity between two probability distributions[8]. Let sets $C_1 = \{c_{i,1} | i = 1, ..., K\}$ and $C_2 = \{c_{i,2} | i = 1, ..., K\}$, each consisting of K conformations of the same protein. Then there are K^2 pairs $\mathrm{RMSD}(c_{i,1}, c_{j,2})$. Let $S = \{s_{i,j} | i, j = 1, ..., K \wedge s_{i,j} \in \{0, 1\}\}$ be the set of binary variables that select whether $(c_{i,1}, c_{i,2})$ are paired up. We then define EMD-RMSD as:

$$\text{EMD-RMSD}(C_1, C_2) \equiv \min_S \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{K} s_{i,j} \mathrm{RMSD}(c_{i,1}, c_{i,2}) \tag{12}$$

$$\text{subject to} \tag{13}$$

$$\sum_{j=1}^{K} s_{i,j} = 1 \; \forall i \quad \wedge \quad \sum_{i=1}^{K} s_{i,j} = 1 \; \forall j \, . \tag{14}$$

$$\tag{15}$$

This finds the optimal pairing of conformations from C_1 and C_2 such that every conformation is paired exactly once with a conformation from the other set, and that the total RMSD across all pairs is minimized. The optimal pairing can be found with any linear assignment problem solver; we use Jonker & Volgenant (1988). The Kabsch-Umeyama algorithm (Kabsch, 1976; Umeyama, 1991) is used to rigidly align all structures to achieve $SE(3)$-invariant RMSD computation.

EMD-RMSD encapsulates the RMSD between the ground truth conformation distribution and the learned conformation distribution in a single value with units in Ångstrom.

[8]This evaluation approach was introduced by Rosenbaum et al. (2021).

H DATA

Defocus is set at $20.000\,\text{Å}$, spherical aberration at $2.7\,\text{mm}$, accelerating voltage at $300\,\text{kV}$. The effective pixel size, after accounting for magnification, is set at $1\,\text{Å}$. The electron dose is set at $\{1000, 100, 50\}\ \text{e}/\text{Å}^2$. No effect of radiation damage was included in the simulations. All other parameters were left at the default values, as described in Parkhurst et al. (2021). Figure 5 shows a sample of the synthetic dataset. Figure 6a shows the conformational states at the extremes of the hinge-motion of the ADK protein.

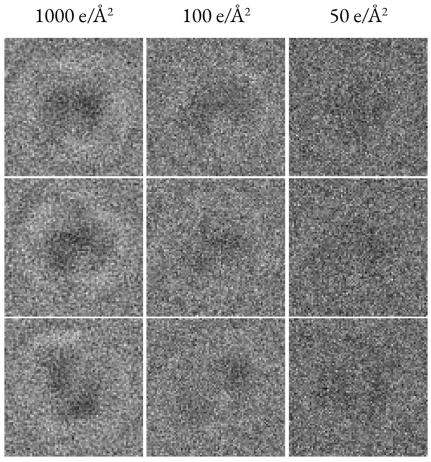
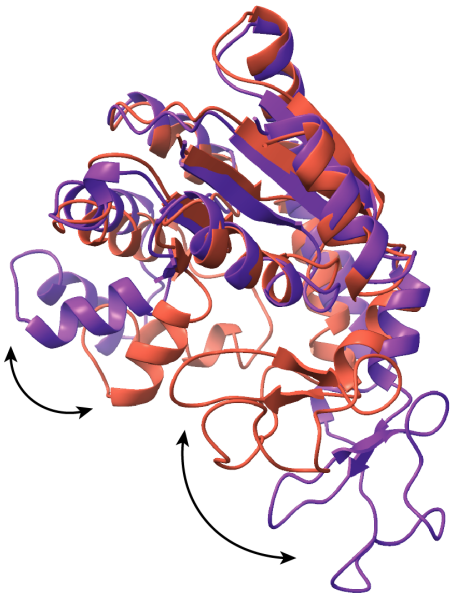


Figure 5: A sample of the synthetic datasets used for our experiments (the ADK protein is projected in these images). At 50 eÅ^{-2}, it is hard to even discern the presence of the particle.

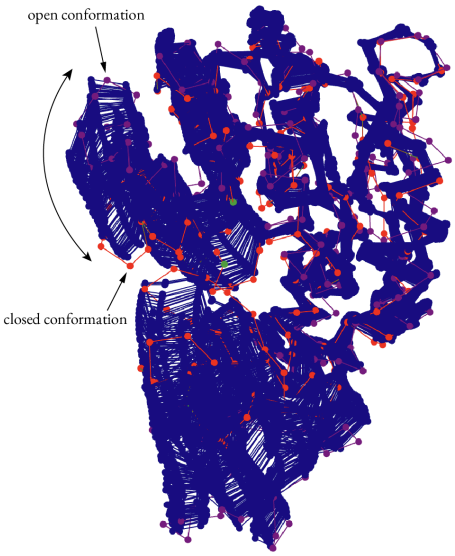
I ADDITIONAL RESULTS AND FIGURES

Table 2: Heterogeneous reconstruction results in terms of RMSD (Å). This table reports on the same experiments as Table 1, but displays the results in terms of a different metric (that is, RMSD instead of EMD-RMSD).

electron dose (eÅ^{-2})	ADK		Nsp13	
	base	geom. loss ablated	base	geom. loss ablated
1000	1.9	2.6	2.5	4.0
100	2.2	2.7	2.8	4.5
50	2.7	2.7	3.0	4.5



(a) Ribbon diagram of the ground truth open and closed conformations of the ADK protein. The two flexible regions that make a hinge-movement are marked with arrows.



(b) Reconstructed conformation distribution by the network. Note how the reconstructions form a uniformly distributed fan between the closed and open state.

Figure 6: The conformation trajectory of ADK, with the open and closed conformations respectively in purple and red.

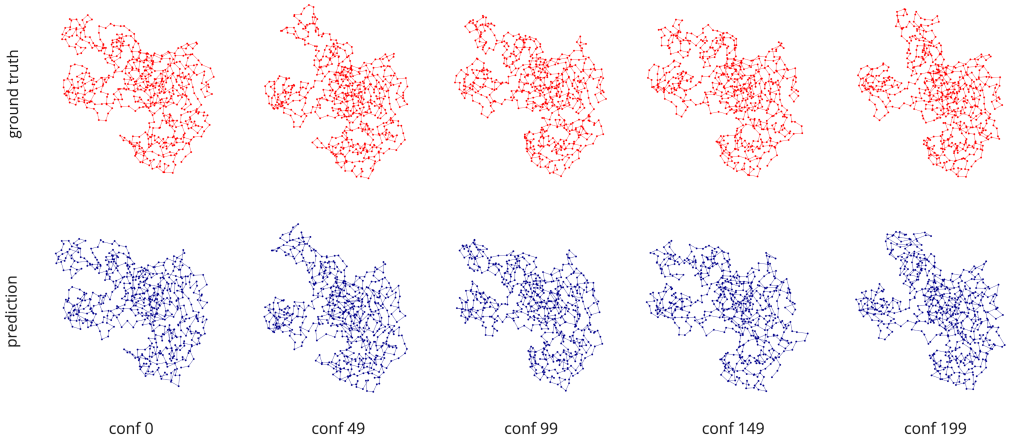


Figure 7: The conformation trajectory of Nsp13. The ground truth conformations are in red, and the corresponding predicted conformations in blue. A different visualization is chosen compared to Figure 6, to better be able to display the disordered motion of Nsp13 in the intermediate states.

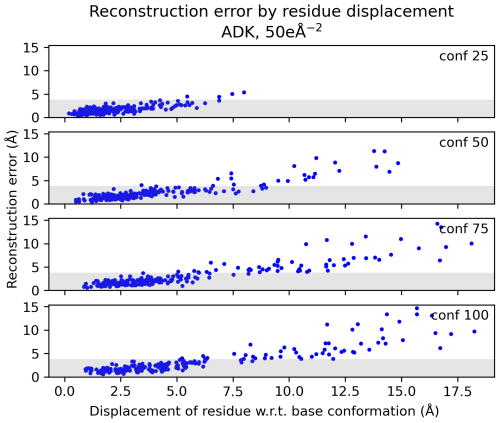


Figure 8: Analogous plot to Figure 2a, but for a different electron dose of $50\,\mathrm{e}/\mathrm{Å}^2$. Note how at this noise level the network no longer succeeds in accurately positioning all residues in the highly-flexible sections of the protein.

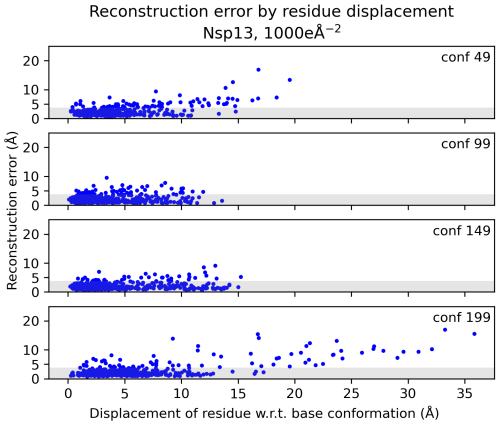


Figure 9: Analogous plot to Figure 3a, but for a different electron dose of $1000\,\mathrm{e}/\mathrm{Å}^2$.

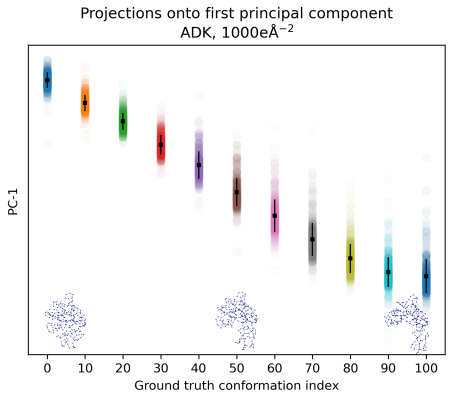


Figure 10: Latent space z visualized using principal component analysis. Each dot represents a latent embedding of an image from the dataset. The y-axis shows the norm of its projection on the first principal component. The x-axis shows the ground truth conformation corresponding to each dot, in increments of 10. Means $\pm$ standard deviation are shown. The clear structure shows that the network naturally learns to embed the low-dimensional conformational heterogeneity of ADK in an appropriate manifold, without the use of mechanisms to enforce structure on the latent space.