DOCTORAL RESEARCH PROJECT

(SUBMITTED TO AND GRANTED BY SOUND.AI)

Geometric deep learning for reconstructing conformation manifolds of biomolecules from cryo electron microscopy images

Thesis director: Slavica Jonic, IMPMC-UMR 7590, CNRS, Sorbonne Université, MNHN
Thesis co-director: Jean Feydy, HeKA team, PariSanté Campus

DESCRIPTION OF THE PROJECT

1. CONTEXT

The 3D structure of proteins and protein complexes is directly linked to their biological functions. In order to accomplish various biological functions, proteins undergo changes of their 3D structures known as conformational changes. There is significant interest in determining the full conformational landscapes of protein complexes, which should help us to understand their mechanisms of action in health and disease [1]. But unfortunately, fast, robust and user-friendly structural biology methods to extract this information are still lacking. If we could obtain conformational landscapes of proteins that are targeted for disease treatment (such as cancer, neurodegenerative diseases, etc.), drug designers could identify the conformations that should be targeted by inhibitors and thus increase the specificity of new treatments. Cryogenic electron microscopy (cryo-EM) single particle analysis (SPA) is currently the mainstream data collection technique for structural biology. This PhD thesis proposal aims at developing geometric deep learning methods for the extraction of conformational landscapes of proteins from their cryo-EM SPA images. The use of geometric deep learning models such as point neural networks is motivated by the fact that proteins and their conformational manifolds are not grid-like structures.

This is an interdisciplinary project that addresses new deep learning methods development for applications in structural biology and drug discovery. It involves a team of 4 supervisors, experts in cryo-EM image analysis methods development (Dr. Jonic, Sorbonne University Paris; see her CV attached here), geometric deep learning methods development (Dr. Feydy, Inria Paris, see his CV attached here), deep learning for cryo-EM (Dr. Sanchez Sorzano, CNB-CSIC, Madrid), and drug discovery (Dr. Mathieu, Sanofi, Vitry sur Seine). It involves short research stays at CNB-CSIC and Sanofi as detailed in the table above and in a dedicated section below.

2. STATE OF THE ART

Deep Mind's AlphaFold2 predicts 3D structures of proteins from their 1D amino-acid sequences. It produces 3D models of similar quality as those that can be obtained with experimental methods, but is limited to the prediction of small static structures [2]. The structure and dynamics of challenging, large (multi-subunit) and flexible complexes is still studied by experimental methods, such as cryo-EM SPA.

Cryo-EM SPA can be used to collect data of different coexisting conformational states of purified complexes. Without tilting the sample, a cryogenic electron microscope acquires 2D views of numerous copies of the same complex (referred to as particles) in random and unknown orientations. Each parallel electron-beam projection image is split into a dataset of thousands of individual particle images. In order to calculate 3D reconstructions of the different conformational states that coexist in the sample, advanced image processing algorithms and software are then needed to solve the heterogeneity of the particle orientations (three Euler angles), positions (shifts in x and y directions in the image plane) and conformations [3]. We note that microscopes use a low electron dose to minimize the radiation damage on the sample, which results in highly noisy images. This complicates the task of disentangling the conformational, orientational and translational heterogeneity.

In contrast to discrete conformational changes (such as two-state heterogeneity due to ligand binding and unbinding), continuous conformational changes of biomolecular complexes (gradual transitions with uncountable intermediate conformational states) yield a particularly challenging type of heterogeneity for image processing algorithms [3, 4]. In order to ease 3D reconstruction at a high resolution, current cryo-EM SPA research usually relies on two simplifications: (i) biochemical procedures are used to make samples as conformationally homogeneous as possible; (ii) images are clustered in conformationally homogeneous classes and each class of images is then averaged through 3D reconstruction [5, 6]. These methods usually result in a small number of classes related to different conformational states and are better suited to discrete conformational changes [7-10].

3. OBJECTIVE

We aim to go beyond these simplifying assumptions: the huge heterogeneity induced by continuous conformational changes provides a unique opportunity to describe multiple coexisting conformations at once. Unfolding the flexibility of complexes and obtaining a low-dimensional representation of the full conformational space that is visible in a sample is a prerequisite for understanding the mechanisms of action of the complexes in health and disease [1].

The last decade was marked by active research on methods that target the exploration of larger degrees of continuous conformational heterogeneity [1, 11-23]. These methods aim at determining the full conformational distribution (also called conformational space, landscape, or manifold), based on which the images with similar conformations could be assembled in 3D reconstructions and, optionally, a displacement of a 3D model can be animated [11, 19].

The problem of determining the particle conformation, orientation and shift from 2D images is an ill-posed inverse problem: the number of unknowns to be determined is larger than the number of input particle images. Combined with the low signal-to-noise ratio (SNR) of cryo-EM images, this makes the problem very challenging. In order to make it tractable, most authors consider a low-dimensional representation of the conformational distribution, such as a finite number of distinct conformations when assuming discrete conformational variability [9, 10, 24]) or a small number of flexible motions when assuming continuous conformational variability [1, 11, 12, 18]. This PhD project aims at developing new approaches, without making assumptions on the conformational variability.

4. METHODOLOGY

Deep learning (DL) approaches recently started to attract attention for the problem of continuous conformational variability [19-23]. But currently, few such methods have been validated with experimental cryo-EM images. To the best of our knowledge, those are: CryoDRGN [21], e2gmm [23] and the approach DeepHEMNMA developed in the lab of Dr. Jonic [19]. The former two use autoencoder architectures; they interpret the conformational heterogeneity in single particle images by assuming known Euler angles and shifts of the particles. These rigid-body parameters are determined prior to DL by classical cryo-EM classification and refinement methods. Unfortunately, the angles and shifts obtained by discretizing the continuous conformational heterogeneity into a small number of average conformational states are likely inaccurate and the mentioned DL methods do not include any refinement schemes to improve these initial angles and shifts. The most recent version of CryoDRGN (based on a variational autoencoder), CryoDRGN2 [25], makes use of a multi-scale exhaustive search of orientations and translations over a discretized 5D parameter space which is more efficient than the branch and bound algorithm that was used in an earlier version of CryoDRGN known as CryoDRGN-BNB [26]. The orientation and translation determination in CryoDRGN2 is done prior to the DL reconstruction of the volumetric protein structure or interleaved with it. Alternating between pose determination and updates on the reconstructed volume is expected to iteratively refine both the estimated poses and the volume. However, since the neural network training objective changes during the course of training because of this iterative process, the method suffers from the problem of vanishing gradients [25].

The third DL approach validated with experimental data is DeepHEMNMA [19], which learns and predicts all three types of unknown parameters: conformation, orientation and shift. Its performance was evaluated on both synthetic and experimental single particle cryo-EM data [19]. The first step of DeepHEMNMA is the creation of the training data set. For this, DeepHEMNMA uses HEMNMA [11] that determines conformational, orientational and shift parameters using a combination of image analysis and normal mode analysis - which is one of the two main approaches for simulation of conformational flexibility [27-29]. In DeepHEMNMA, HEMNMA is used to estimate these parameters from a subset of images. Then, a deep convolutional neural network is trained to learn the relationship between this subset of images and its HEMNMA-estimated parameters. The network is a ResNet-34 feature extractor [30] followed by a multilayer layer perceptron. The trained network is then used to predict the parameters from the remaining images (unseen during the training). Finally, the conformational landscape is obtained by mapping the inferred normal-mode amplitudes onto a lower-dimensional space, which enables 3D reconstruction using the inferred angles and shifts as well as animations of a model displacement and identification of possible hidden conformations. In the experiments of [19], DeepHEMNMA was shown to be more than 40 times faster than HEMNMA. However, DeepHEMNMA depends on the number of selected normal modes that are used in the image analysis step of HEMNMA. This number corresponds to the dimension of the space of simulated motion directions, which is usually chosen smaller than 6 to keep run times reasonable. We note that the selection of the most relevant normal modes is not a trivial task: going beyond the standard selection of modes with lowest frequencies and highest collectivities, it depends on the molecular complex under study and always involves some subjectivity [11]. This PhD proposal targets the development of new approaches that will be independent of any prior knowledge of the molecular complex, or of its conformational dynamics and rigid-body pose parameters (orientation and position).

Recently, the problem of learning continuous conformational variability has been addressed under a generative adversarial network (GAN) framework [20]. Multi-CryoGAN [20] is an extension of a single-conformation reconstruction approach CryoGAN [31] to the reconstruction of conformation manifolds. It is a likelihood-free method (contrary to CryoDRGN), meaning that it does not require pose estimation. It learns to reconstruct 3D structures whose randomly projected images match the acquired data in a distributional sense. Its capacity to reconstruct low-dimensional conformation manifolds has been shown with synthetic data, but not yet with experimental data. Besides, the resolution of the reconstructed structures is low due to the small size of the volumes used to generate the conformations (32 × 32 × 32 voxels, voxel size of 5 Å) and the small size of 2D projections (32 × 32 pixels). Even the single-conformation approach CryoGAN achieves a resolution of only around 11 Å on synthetic data. This is far from being enough to handle real cryo-EM images, which are commonly acquired with a pixel size of 1 Å (or lower) and thus represent particles as patches of 150 × 150 to 400 × 400 pixels. Considering that standard cryo-EM image analysis methods currently achieve near-atomic resolutions (~2-3 Å), even for samples with some (moderate) conformational heterogeneity, it is clear that further work is required to improve generative models such as GANs.

Deep learning models have been especially successful when dealing with signals supported on a grid (audio, images or video) and for which invariance to an underlying group of translations can easily be built into network architectures using convolutions. There is now growing interest in extending this methodology to other (non-Euclidean) geometric domains such as social networks, surface meshes in computational anatomy or atomic point clouds in biochemistry. This research on graphs and manifolds is known under the umbrella term of geometric deep learning [32].

Biological functions of proteins are defined by geometric and chemical properties of their 3D structures. Recent works contributed by Dr. Feydy have shown that geometric deep learning can be used on 3D protein surfaces to identify potential functional sites, such as binding targets for potential drugs in the context of virtual (in silico) drug screening [33, 34]. This question is closely related to the problem that will be addressed in this PhD project. The geometric deep learning methods that will be developed in this thesis will be useful for identifying 3D structures (from the entire reconstructed conformation manifold) that are potentially more favorable for drug screening.

We note that over the years, computing frameworks such as Fortran, Matlab, NumPy, TensorFlow or PyTorch have put a focus on linear operations (matrix products, convolutions, Fourier transforms...) and thus biased research in applied mathematics towards Euclidean models and geometries. A significant part of the work of Dr. Feydy is to break through this software bottleneck to enable efficient machine learning on geometric structures such as atomic point clouds. As part of the KeOps and GeomLoss packages, he develops extensions for Python and R that ease the development of geometric algorithms with cutting edge performance on parallel hardware (GPU), a transparent user interface and full support for automatic differentiation. The KeOps package optimizes a wide class of geometric computations (from nearest neighbor search to convolution layers for point neural networks) with x10 to x100 speed-ups compared with standard baselines in the deep learning community. It has been downloaded more than 250,000 times.

GeomLoss provides fast and scalable solvers for the optimal transport (OT) problem, which generalizes sorting to spaces of dimension D > 1. This mathematical tool has been used for shape analysis (instead of the nearest neighbor projection), data science (to re-order samples and match them with each other) and generative deep learning models (to compare probability distributions using the "Wasserstein" or "Earth Mover's" distance) [35-37]. The expertise of Dr. Feydy in high-performance computing and open-source software development is highly relevant to this PhD project. In cryo-EM, OT was used to address rigid-body alignment between two 3D density volumes (obtained by 3D reconstruction from 2D images) [38], interpolation between two different conformations encoded as density volumes [39] and shape analysis in a set of density volumes [40]. Using synthetic data, the authors of [40] have also shown the advantage of using the "Earth Mover's" distance induced by OT between 3D volumes to learn a conformational manifold. In this PhD thesis, we will use the tools available in KeOps and GeomLoss as building blocks to implement both OT-based layers and point neural networks that reconstruct the conformation manifold of 2D cryo-EM images. Upon validation, these new modules will be merged in these libraries to reach a wide public.

5. CONCEPT, WORK PLAN AND TIMELINE

In this PhD, the problem of reconstructing conformation manifolds from 2D cryo-EM images of biomolecules will be addressed by new unsupervised deep learning approaches, based on geometric deep learning. To this end, new graph- and OT-based learning approaches will be developed.

The following work plan and timeline is given for four sub-objectives of the project:

- **Objective 1 (Year 1, Supervisors: S. Jonic, J. Feydy):** Compare different metrics between 3D shapes for the reconstruction of conformational manifolds. We will notably study kernel metrics, the Wasserstein distance induced by OT and advanced shape metrics from the computer graphics and medical imaging literatures [41-42]. New geometric tools that will be developed will be integrated in KeOps and GeomLoss.
- **Objective 2 (Years 1-2, Supervisors: S. Jonic, J. Feydy)**: Develop an end-to-end generative model for conformation manifolds that will target higher resolutions of the conformational reconstructions and increased robustness to conformational heterogeneity compared with the approach [20]. This model will rely on a geometric backbone (point neural network) similar to that of the dMaSIF architecture [33-34] and may be trained using a variational auto-encoder, generative adversarial network or score-based loss function.
- **Objective 3 (Years 1-3, Supervisor S. Jonic, 1-month visit to Sanofi Vitry-sur-Seine at Year 2 - for more information on the visit, see below):** Validate and compare the methods that will be developed in this thesis with existing works, using both synthetic images and experimental data from the EMPIAR database. The speed and accuracy of the proposed methods will be compared with DeepHEMNMA, CryoDRGN2 and e2gmm, for which software implementations are available. We note that data synthesis to obtain close-to-real images is an important step in the development of cryo-EM methods: it enables testing in a controlled environment, before deployment on experimental data with unknown ground-truth solutions. As a first step, we will benchmark the proposed methods on the synthetic datasets that were used to test DeepHEMNMA and other methods of the team of Dr. Jonic. Then, we will perform extensive tests on experimental data that has already been analyzed with other methods and deposited in the standard public database EMPIAR.
- **Objective 4 (Years 1-3, Supervisor S. Jonic, 2-month visit to CNB-CSIC Madrid at Year 1 - for more information on the visit, see below):** Integrate the new methods into the existing open-source

software package ContinuousFlex developed in the lab of Dr. Jonic [43]. ContinuousFlex is a plugin of Scipion, which is one of the most widely used open-source software in the cryo-EM field.

6. RISK/FEASIBILITY ASSESSMENT

While access to large datasets is a major obstacle for deep learning applications in many fields, this will not be an issue for this project. Single cryo-EM SPA experiments already produce extremely large samples of a conformational manifold, with several millions of particle images being collected automatically in a short amount of time (1-2 days).

Part of the project will also be dedicated to the development of numerical routines and tools for cryo-EM that will be based on KeOps/GeomLoss and significantly faster than current mainstream implementations. This foundational numerical work will be a reliable contribution to the field that will allow the cryo-EM community to work at higher resolutions.

7. EXPECTED IMPACT

The methods that will be developed in this PhD thesis are expected to have a high impact on healthcare research, and in particular on drug discovery. Full conformational landscapes of biomolecular complexes are precious, but out of reach in a majority of cases due to the lack of suitable analysis methods for cryo-EM images. Such conformational landscapes potentially contain information on conformations that could be targeted by inhibitors to increase treatment specificity for numerous diseases.

The software developed during the thesis will be open-source and integrated into Scipion as a plugin (ContinuousFlex), meaning that it will be accessible to the worldwide structural-biology cryo-EM community. Intermediate sub-routines that are relevant outside of structural biology will also be packaged in the KeOps, GeomLoss and Python Optimal Transport libraries: these tools have a wide appeal in the communities of kernel methods, optimal transport and computational anatomy. Thus, other communities will also benefit from the software that will be developed during this project. All publications will be deposited on HAL in open-access.

8. POTENTIAL ETHICAL ISSUES: There are no ethical issues with this PhD thesis work. We will use public data from the EMPIAR database.

9. POSSIBILITIES FOR KNOWLEDGE TRANSFER: This project is a collaboration with Sanofi. The benefit for us from the interaction with Sanofi is that it will inspire new research directions potentially yielding additional funding. The benefit for Sanofi is that they will be trained to use our methods for their research based on cryo-EM, including those that will be developed during this PhD thesis.

10. OUTREACH ACTIVITIES: The PhD student will give presentations at international cryo-EM and AI flagship conferences and workshops as well as in the collaborators' labs. We will also organize an online and in-presence training of the researchers potentially interested in using the methods developed during the PhD thesis.

11. INSIGHTS ON TYPICAL CAREER ORIENTATIONS: There are already plenty of career opportunities after the successful completion of a PhD on this subject. Cryo-EM is under exponential development in France and worldwide, in both academic and industrial sectors. Regarding industrial opportunities, this PhD thesis could lead to software development positions for manufacturers of cryo-EM microscopes, or for pharmaceutical companies that use these imaging devices. The emerging development of AI for cryo-EM is expected to be a revolution for the field and to open up additional opportunities.

PROPOSED INDUSTRIAL AND/OR INTERNATIONAL SECONDMENTS

This interdisciplinary project involves collaboration with one foreign academic partner and one industrial partner.

The foreign academic partner is the Spanish National Center for Biotechnology (CNB-CSIC), hosting the EC Instruct Spanish Center (Madrid, Spain). Dr. Jonic has a long-term collaboration with the Biocomputing Unit lab of the CNB-CSIC (team leader: Dr. José Maria Carazo), which is developing Scipion, one of the most used image processing open-source software packages in the field of cryo-EM image analysis methods. Scipion was recently extended to include AI for several image processing tasks specific to 3D structure reconstruction from cryo-EM data [44-47]. The software that will be developed during this PhD thesis will be open-source and integrated into Scipion to allow its broad usage. A 2-month research stay in the Biocomputing Unit lab of the CNB-CSIC is planned for the 1st year of PhD and will be supervised by Dr. Carlos Oscar Sanchez Sorzano. During the stay at the CNB-CSIC, the PhD student will learn about the methods that are already available in Scipion in order to compare the new methods with the existing ones. Also, the PhD student will be trained by the Scipion developers team regarding the new software integration into Scipion. If necessary, an additional, 1-month stay at CNB-CSIC will be organized during the 2nd PhD year.

The industrial partner is Sanofi, Vitry-sur-Seine, France. Sanofi is interested in the development of new methods for analyzing conformational flexibility from cryo-EM images, for use in their research, considering the limitations of the available, standard methods. A 1-month research stay in the BioStructure and Biophysics Department of Sanofi is planned for the 2nd year of PhD and will be supervised by Dr. Magali Mathieu. During the stay at Sanofi, the PhD student will test the software developed during the thesis on the data provided by Sanofi and will interact with the team of Dr. Mathieu regarding the interpretation of the results of the data analysis.

REFERENCES OF THE TWO MAIN CO-SUPERVISORS RELATED TO THIS PHD PROJECT

1. Hamitouche, I., and **Jonic, S**. (**2022**). DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. Front Mol Biosci *9*, 965645. DOI: 10.3389/fmolb.2022.965645. Link
2. Harastani, M., Eltsov, M., Leforestier, A., and **Jonic, S**. (**2022**). TomoFlow: Analysis of Continuous Conformational Variability of Macromolecules in Cryogenic Subtomograms based on 3D Dense Optical Flow. J Mol Biol *434*, 167381. DOI: 10.1016/j.jmb.2021.167381. Link
3. Sverrisson, F., **Feydy, J**., Correia, B.E. and Bronstein, M.M. (**2021**). Fast end-to-end learning on protein surfaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15272-15281), CVPR 2021. DOI: 10.1109/CVPR46437.2021.01502 Link
4. **Feydy, J**., Glaunès, A., Charlier, B. and Bronstein, M., (**2020**). Fast geometric learning with symbolic matrices. Advances in Neural Information Processing Systems, 33, pp.14448-14462. NeurIPS 2020. Link
5. Shen, Z., **Feydy, J**., Liu, P., Curiale, A.H., San Jose Estepar, R., San Jose Estepar, R. and Niethammer, M. (**2021**). Accurate Point Cloud Registration with Robust Optimal Transport. Advances in Neural Information Processing Systems, 34, pp.5373-5389. NeurIPS 2021. Link

LITERATURE:

[1] Dashti A, Mashayekhi G, Shekhar M, Ben Hail D, Salah S, Schwander P, et al. Retrieving functional pathways of biomolecules from single-particle snapshots. Nat Commun. 2020;11:4734.

[2] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583-9.

[3] Jonić S. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. Current opinion in structural biology. 2017;43:114-21.

[4] Sorzano COS, Jiménez A, Mota J, Vilas JL, Maluenda D, Martínez M, et al. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. Acta Crystallogr F Struct Biol Commun. 2019;75:19-32.

[5] Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, et al. Single-particle cryo-EM at atomic resolution. Nature. 2020;587:152-6.

[6] Kato K, Miyazaki N, Hamaguchi T, Nakajima Y, Akita F, Yonekura K, et al. High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams. Communications Biology. 2021;4:382.

[7] Penczek PA, Frank J, Spahn CM. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. Journal of structural biology. 2006;154:184-94.

[8] Scheres SH, Gao H, Valle M, Herman GT, Eggermont PP, Frank J, et al. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat Methods. 2007;4:27-9.

[9] Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012;180:519-30.

[10] Lyumkis D, Brilot AF, Theobald DL, Grigorieff N. Likelihood-based classification of cryo-EM images using FREALIGN. J Struct Biol. 2013;183:377-88.

[11] Jin Q, Sorzano COS, De La Rosa-Trevín JM, Bilbao-Castro JR, Núñez-Ramírez R, Llorca O, et al. Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. Structure. 2014;22:496-506.

[12] Dashti A, Schwander P, Langlois R, Fung R, Li W, Hosseinizadeh A, et al. Trajectories of the ribosome as a Brownian nanomachine. Proc Natl Acad Sci U S A. 2014;111:17492-7.

[13] Tagare HD, Kucukelbir A, Sigworth FJ, Wang H, Rao M. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. Journal of structural biology. 2015;191:245-62.

[14] Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, et al. Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. Cell. 2018;172:454-64.e11.

[15] Lederman R, Andén J, Singer A. Hyper-Molecules: on the Representation and Recovery of Dynamical Structures for Applications in Flexible Macro-Molecules in Cryo-EM. Inverse Problems. 2020;36:044005.

[16] Moscovich A, Halevi A, Andén J, Singer A. Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. Inverse Problems. 2020;36:024003.

[17] Giraldo-Barreto J, Ortiz S, Thiede EH, Palacio-Rodriguez K, Carpenter B, Barnett AH, et al. A Bayesian approach to extracting free-energy profiles from cryo-electron microscopy experiments. Scientific Reports. 2021;11:13657.

[18] Punjani A, Fleet DJ. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. Journal of Structural Biology. 2021;213:107702.

[19] Hamitouche I, Jonic S. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. Front Mol Biosci. 2022;9:965645.

[20] Gupta H, Phan TH, Yoo J, Unser M. Multi-CryoGAN: Reconstruction of Continuous Conformations in Cryo-EM Using Generative Adversarial Networks. In: Bartoli A, Fusiello A, editors. Computer Vision – ECCV 2020 Workshops. Cham: Springer International Publishing; 2020. p. 429-44.

[21] Zhong ED, Bepler T, Berger B, Davis JH. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. Nature Methods. 2021;18:176-85.

[22] Rosenbaum D, Garnelo M, Zielinski M, Beattie C, Clancy E, Huber A, et al. Inferring a Continuous Distribution of Atom Coordinates from Cryo-EM Images using VAEs. arXiv preprint arXiv:210614108. 2021.

[23] Chen M, Ludtke SJ. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. Nature Methods. 2021;18:930-6.

[24] Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14:290-6.

[25] Zhong ED, Lerer A, Davis JH, Berger B. CryoDRGN2: Ab initio neural reconstruction of 3D protein structures from real cryo-EM images. 2021 IEEE/CVF International Conference on Computer Vision (ICCV)2021. p. 4046-55.

[26] Zhong ED, Bepler T, Davis JH, Berger B. Reconstructing continuous distributions of 3D protein structure from cryo-EM images. ICLR2020.

[27] López-Blanco JR, Chacón P. New generation of elastic network models. Current Opinion in Structural Biology. 2016;37:46-53.

[28] Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein Eng. 2001;14:1-6.

[29] Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Physical review letters. 1996;77:1905.

[30] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition2016. p. 770-8.

[31] Gupta H, McCann MT, Donati L, Unser M. CryoGAN: A New Reconstruction Paradigm for Single-Particle Cryo-EM Via Deep Adversarial Learning. IEEE Transactions on Computational Imaging. 2021;7:759-74.

[32] Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P. Geometric Deep Learning: Going beyond Euclidean data. IEEE Signal Processing Magazine. 2017;34:18-42.

[33] Sverrisson F, Feydy J, Correia BE, Bronstein MM. Fast end-to-end learning on protein surfaces. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)2021. p. 15267-76.

[34] Sverrisson F, Feydy J, Southern J, Bronstein MM, Correia BE. Physics-informed deep neural network for rigid-body protein docking. ICLR2022 Machine Learning for Drug Discovery2022.

[35] Peyré G, Cuturi M. Computational Optimal Transport: With Applications to Data Science: now; 2019.

[36] Feydy J, Charlier B, Vialard F-X, Peyré G. Optimal Transport for Diffeomorphic Registration. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, editors. Medical Image Computing and Computer Assisted Intervention − MICCAI 2017. Cham: Springer International Publishing; 2017. p. 291-9.

[37] Arjovsky M, Chintala S, Bottou L. Wasserstein Generative Adversarial Networks. International Conference on Machine Learning2017.

[38] Riahi AT, Woollard G, Poitevin F, Condon A, Duc KD. AlignOT: An optimal transport based algorithm for fast 3D alignment with applications to cryogenic electron microscopy density maps. arXiv:2210093612022.

[39] Ecoffet A, Poitevin F, Duc KD. MorphOT: Transport-based interpolation between EM maps with UCSF ChimeraX. Bioinformatics. 2020.

[40] Zelesko N, Moscovich A, Kileel J, Singer A. Earthmover-Based Manifold Learning for Analyzing Molecular Conformation Spaces. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)2020. p. 1715-9.

[41] Feydy J. Geometric data analysis, beyond convolutions. PhD thesis, 2020. Université Paris-Saclay.

[42] Shen, Z., Feydy, J., Liu, P., Curiale, A.H., San Jose Estepar, R., San Jose Estepar, R. and Niethammer, M., 2021. Accurate Point Cloud Registration with Robust Optimal Transport. NeurIPS 2021. *Advances in Neural Information Processing Systems*, *34*, pp.5373-5389.

[43] Harastani M, Vuillemot R, Hamitouche I, Moghadam NB, Jonic S. ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. Journal of Structural Biology. 2022;214:107906.

[44] Sanchez-Garcia R, Segura J, Maluenda D, Carazo JM, Sorzano COS. Deep Consensus, a deep learning-based approach for particle pruning in cryo-electron microscopy. IUCrJ. 2018;5:854-65.

[45] Ramírez-Aportela E, Mota J, Conesa P, Carazo JM, Sorzano COS. DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. IUCrJ. 2019;6:1054-63.

[46] Sanchez-Garcia R, Gomez-Blanco J, Cuervo A, Carazo JM, Sorzano COS, Vargas J. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Communications Biology. 2021;4:874.

[47] Sanchez-Garcia R, Segura J, Maluenda D, Sorzano COS, Carazo JM. MicrographCleaner: A python package for cryo-EM micrograph cleaning using deep learning. J Struct Biol. 2020;210:107498.