

# Mapping the motion and structure of flexible proteins from cryo-EM data

A deep learning algorithm maps out the continuous conformational changes of flexible protein molecules from single-particle cryo-electron microscopy images, allowing the visualization of the conformational landscape of a protein with improved resolution of its moving parts.

## This is a summary of:

Punjani, A. & Fleet, D. J. 3DFlex: determining structure and motion of flexible proteins from cryo-EM. *Nat. Methods* <https://doi.org/10.1038/s41592-023-01853-8> (2023).

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 12 May 2023

## The problem

Proteins form the molecular machinery of the cell and often exhibit continuous changes in their 3D structure. These motions are tightly linked to biochemical function. Cryogenic electron microscopy (cryo-EM) can enable structural studies of flexible proteins by collecting thousands of 2D images of single particles frozen in different conformational states, which together describe the target protein's heterogeneity<sup>1</sup>. Existing computational methods for reconstructing 3D structures from these images excel in resolving rigid structures; however, they yield blurred and uninterpretable results in the presence of flexible motion. The need for new algorithms that are capable of resolving continuous heterogeneity and, in doing so, determining motion and high-quality 3D structure in flexible regions is an exciting, long-standing open problem in cryo-EM research with the potential to illuminate fundamental questions in structural biology.

## The solution

We developed 3D flexible refinement (3DFlex), a deep neural network model for profiling continuously flexible protein molecules. It represents a flexible molecule in terms of a single, high-resolution canonical 3D density map plus a neural network that takes in a latent coordinate and generates a deformation field that models the flexible motion of the protein. The deformation field 'bends' the canonical density of the protein through convection to produce a version of the protein in a particular conformation. As the input latent coordinate is varied, the generated deformations span all conformations captured by the model. The parameters of the model are jointly learned from image data using a specialized training algorithm, with little prior knowledge about the flexibility of the molecule. The algorithm considers the fact that most conformational variability is a result of physical processes that tend to transport density over space. It preserves local geometry (for example, the relative positions and orientations of side chains). Because of the way 3DFlex models deformations, it naturally aggregates structural information across the conformational landscape of the target protein to improve the resolution of the density map in flexible regions. This is in contrast to other recent methods such as cryoDRGN<sup>2</sup>, 3DVA<sup>3</sup> and e2gmm<sup>4</sup>, which can model conformational landscapes but do not directly improve the resolution of moving parts.

Results from empirical datasets demonstrated the potential of 3DFlex to uncover the structure and motion of flexible proteins. On a dataset from tri-snRNP spliceosome particles (EMPIAR-10073), 3DFlex learned a wide range of non-rigid motions, including the bending of subunits across a span of more than 20 Å (Fig. 1). In doing so, the algorithm aggregated structural information from all conformations into a single, optimized density map that resolves high-resolution details in  $\alpha$ -helices and  $\beta$ -sheets even in the flexible domains. Further experiments on the TRPV1 ion channel (EMPIAR-10059), the SARS-CoV-2 spike protein (EMPIAR-10516),  $\alpha$ V $\beta$ 8 integrin (EMPIAR-10345) and a translocating ribosome (EMPIAR-10792) demonstrated that 3DFlex can reveal structures not resolved by conventional refinement and map out conformational variations for both small proteins and large protein complexes.

## Future directions

3DFlex represents a motion-based approach to cryo-EM reconstruction and helps to establish the potential for high-resolution reconstruction of continuous heterogeneity for molecules that were beyond the reach of previous methods. With such capabilities, 3DFlex opens up new avenues of inquiry into the study of biological mechanisms and functions. 3DFlex is now available as part of the cryoSPARC<sup>5</sup> software system, and we hope that its deployment in an accessible and user-friendly software package will facilitate its uptake within the community.

There are some limitations of the current instantiation of the 3DFlex algorithm. Notably, it is limited in its ability to resolve the intricate motion of small subunits, such as a single loop or side chain. Further, it cannot effectively model cases in which parts of a complex associate or dissociate (compositional heterogeneity).

Among the possible directions for future research, we highlight the need for rigorous methods for evaluation to help quantify the accuracy and resolution of the estimated motion fields. As research on continuous heterogeneity continues, this will become crucial both for methods development and for reliable interpretation of learned models. There is also a need to extend the formulation of the 3DFlex model to allow for compositional heterogeneity in the data.

## Ali Punjani<sup>1,2</sup> & David J. Fleet<sup>1</sup>

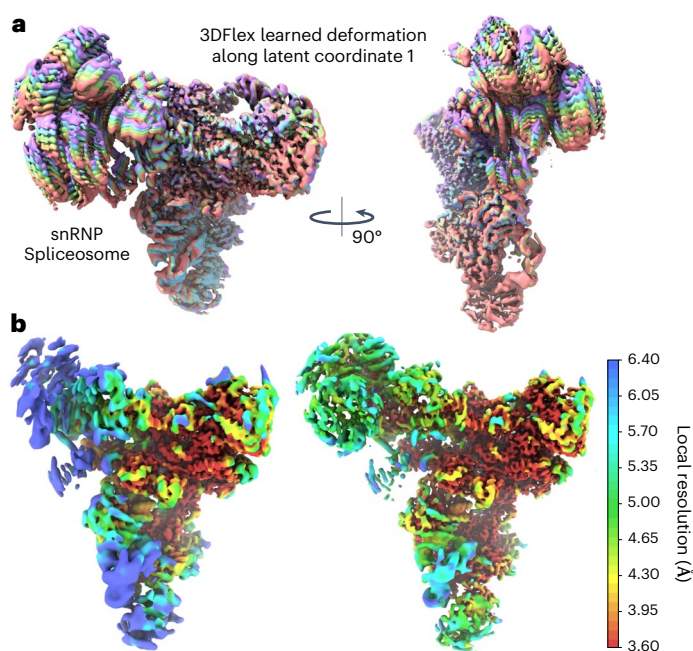
<sup>1</sup>University of Toronto, Toronto, Ontario, Canada. <sup>2</sup>Structura Biotechnology Inc., Toronto, Ontario, Canada.

## EXPERT OPINION

“Punjani and Fleet present an innovative method to reconstruct conformationally heterogeneous molecules from cryo-EM data. Unlike most other approaches, their method combines the information obtained from all conformations to gain an overall high-resolution representation of the molecule. They demonstrate the power

of their approach on several examples, including a synthetic example with known ground truth. This method will undoubtedly be of great value to the cryo-EM community.” **David Haselbach, Research Institute of Molecular Pathology, Vienna, Austria.**

## FIGURE



**Fig. 1 | Motion and structure of a spliceosome particle solved by 3DFlex.** **a**, Colored series of density maps at five positions along the first latent dimension of motion learned by 3DFlex, showing the non-rigid deformation of the head region (top left and top right of the first and second views, respectively). **b**, Density maps from conventional rigid (left) and 3DFlex (right) refinements, colored by local resolution. 3DFlex improves resolution in flexible parts such as the head region. © 2023, Punjani, A. & Fleet, D. J., CCBY 4.0.

## BEHIND THE PAPER

This is one of the most challenging but rewarding research projects either of us have undertaken. 3DFlex is a conceptual departure from previous work on continuous heterogeneity, but it might not be entirely unfamiliar to computer vision researchers. Initial progress on 3DFlex was slow, as we searched for an effective way to regularize the deformation field generator. Many early attempts were unsuccessful. But our first

encouraging results on a low-resolution ATPase dataset led to several months of concentrated work on the method and careful testing. We released a preliminary draft of this work in 2021, but continued to develop and stress test the method, resulting in the current version of 3DFlex. We're now extremely keen to see the creative ways in which other researchers will make use of it. **A.P. & D.F.**

## REFERENCES

1. Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis. *Biol. Chem.* **29**, 5181–5197 (2019).  
**This review article provides an overview of single-particle cryo-EM fundamentals and the importance of mapping out conformational heterogeneity.**
2. Punjani, A. & Fleet, D. J. 3D Variability analysis: directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *J. Struct. Biol.* **213**, 107702 (2021).  
**This paper reports a computational method for fitting a linear subspace model of cryo-EM density to single particle data and is an important baseline for methods that resolve continuous heterogeneity.**
3. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).  
**This article reports a deep generative model of cryo-EM density maps that can effectively map out non-linear changes across a conformational landscape.**
4. Chen, M. & Ludtke, S. J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat. Methods* **18**, 930–936 (2021).  
**This paper reports a deep generative model that can represent the motion of particle density using the displacement of Gaussian mixture components.**
5. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).  
**This article reports the development of the CryoSPARC software system, in which 3DFlex is implemented.**

## FROM THE EDITOR

“3DFlex models continuous molecular heterogeneity from cryo-EM data and resolves detailed 3D structure of flexible proteins. This is a challenging problem in structural biology, and we're excited to see how this approach impacts the field.”

**Arunima Singh, Senior Editor, Nature Methods.**