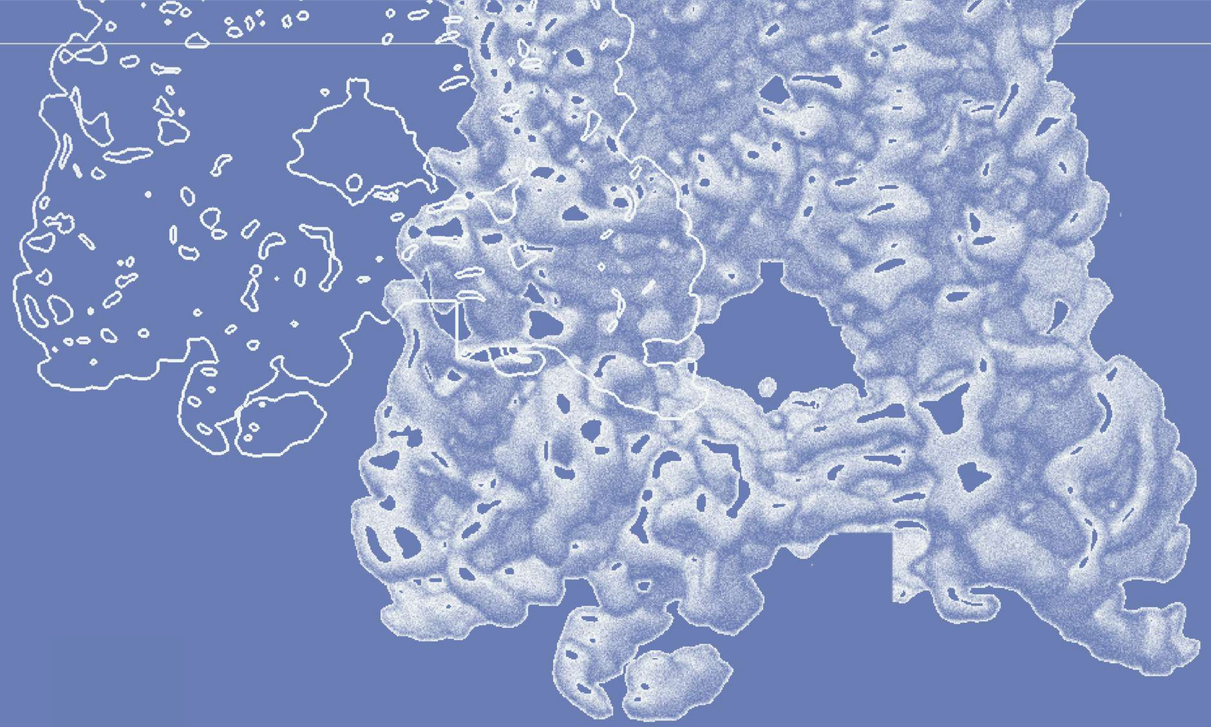
Sonya **Hanson**
shanson@flatironinstitute.org



ML methods for heterogeneity in SPA →

Sonya M. Hanson
Structural and Molecular Biophysics, CCB/CCM
Flatiron Institute, Simons Foundation, NYC

NYSBC-NCCAT SPA short course 2024

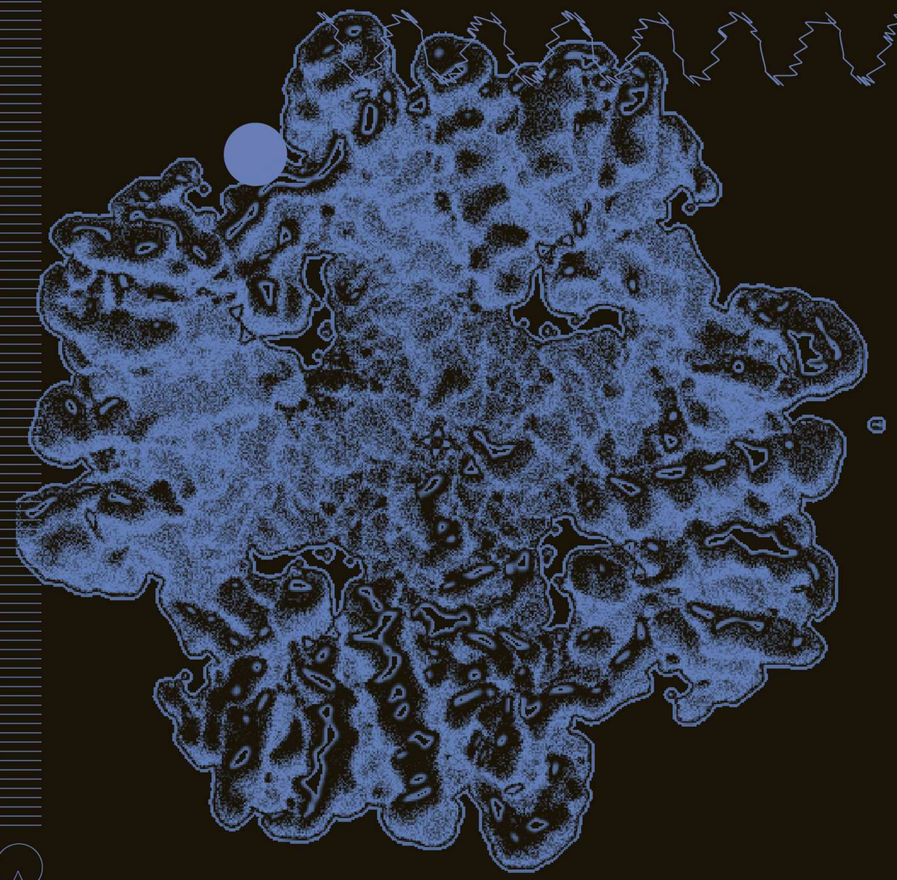
Thursday, March 14, 2024

FLATIRON INSTITUTE

2024

Sonya **Hanson**
shanson@flatironinstitute.org

Computational methods for heterogeneity in cryo-EM



ML methods for heterogeneity in cryo-EM

- PCA
- ManifoldEM
- cryoDRGN
- GMM in eman2
- cryoSPARC/RELION
- … and more!

Practical considerations when using continuous heterogeneity methods.

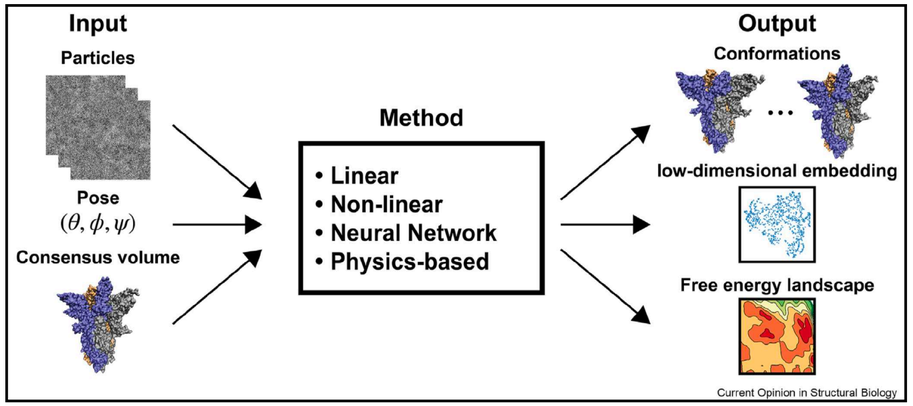
Combining cryo-EM and Molecular Dynamics

Assessing the performance of continuous heterogeneity methods.

A couple useful reviews to check out.

Conformational heterogeneity and probability distributions from single-particle cryo-electron microscopy

Wai Shing Tang[1], Ellen D. Zhong[3], Sonya M. Hanson[1,2], Erik H. Thiede[1] and Pilar Cossio[1,2]

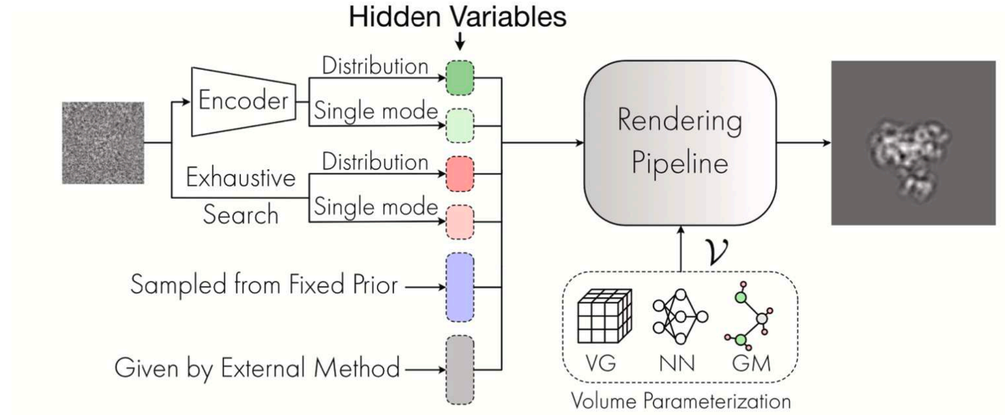


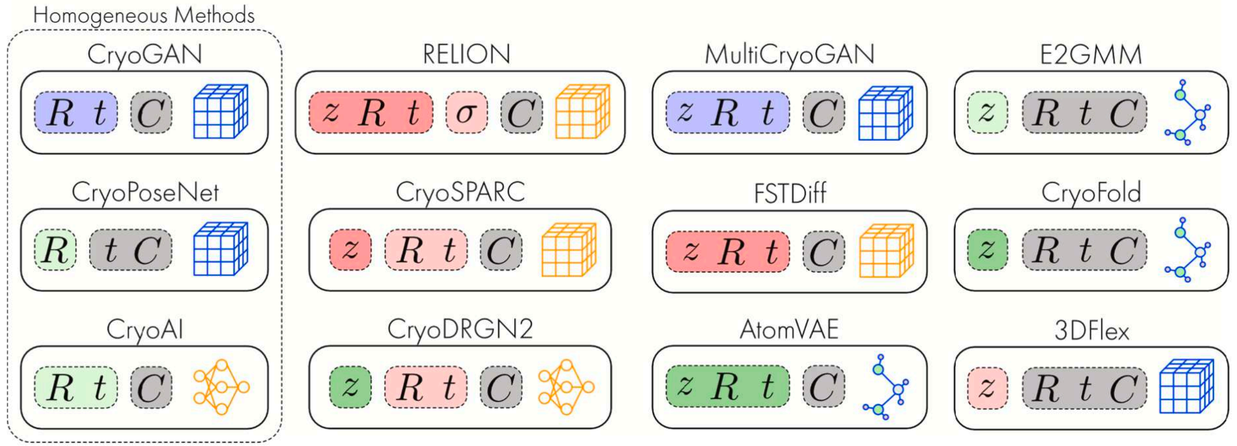
www.sciencedirect.com Current Opinion in Structural Biology 2023, 81:102626

Journal of Structural Biology 214 (2022) 107920

Deep generative modeling for volume reconstruction in cryo-electron microscopy

Claire Donnat[a,1], Axel Levy[b,c], Frédéric Poitevin[c], Ellen D. Zhong[d], Nina Miolane[e,1]

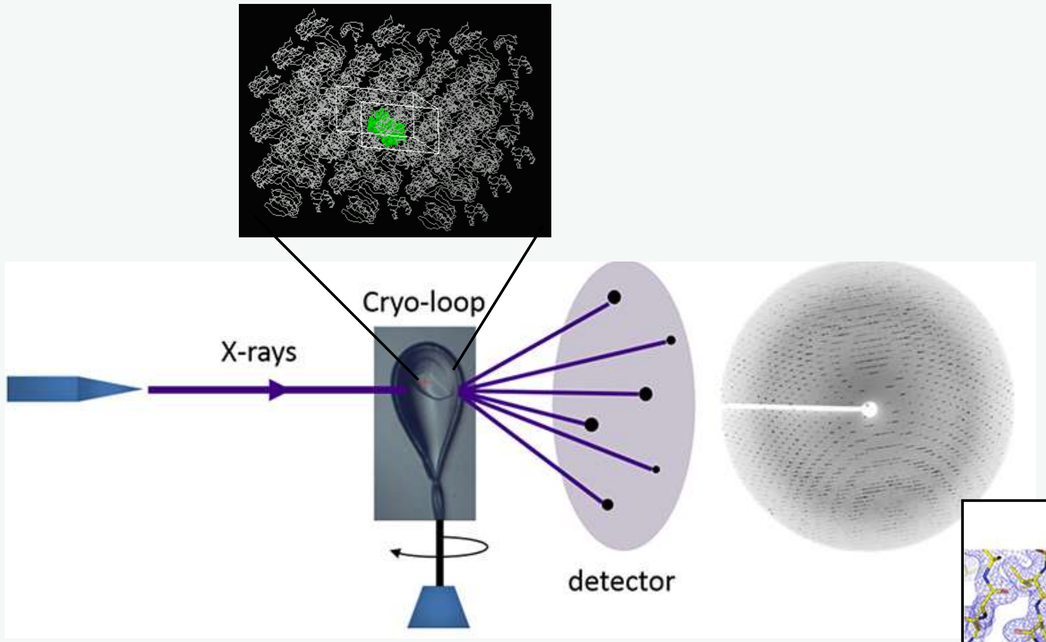


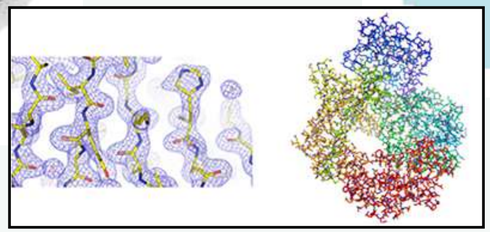


2024

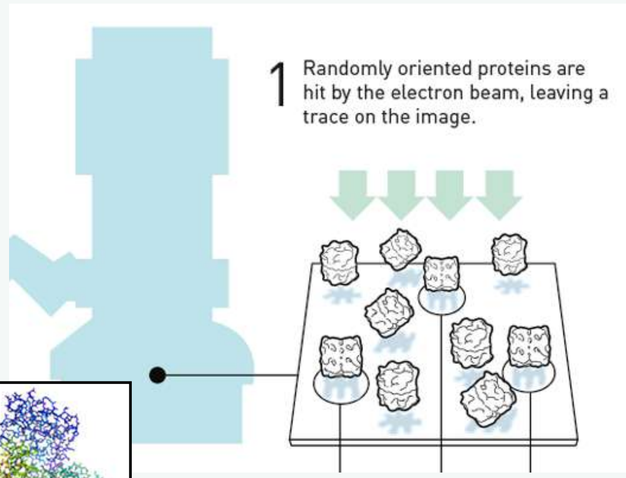
A key advantage of cryo-EM not yet fully exploited!

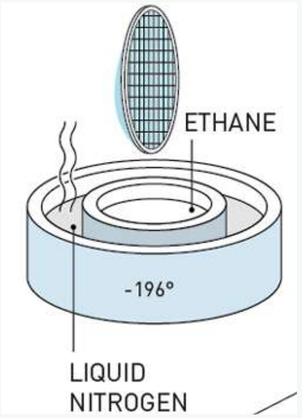
X-ray crystallography





cryo-EM

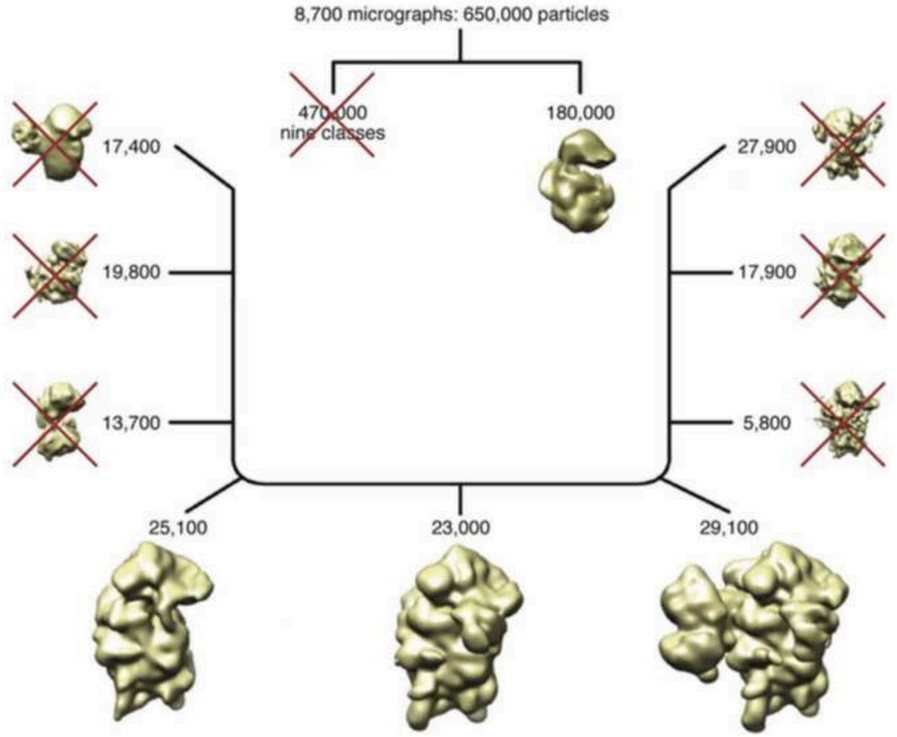




A key advantage of cryo-EM still to be fully exploited!

Traditional 3D classification methods are imperfect.

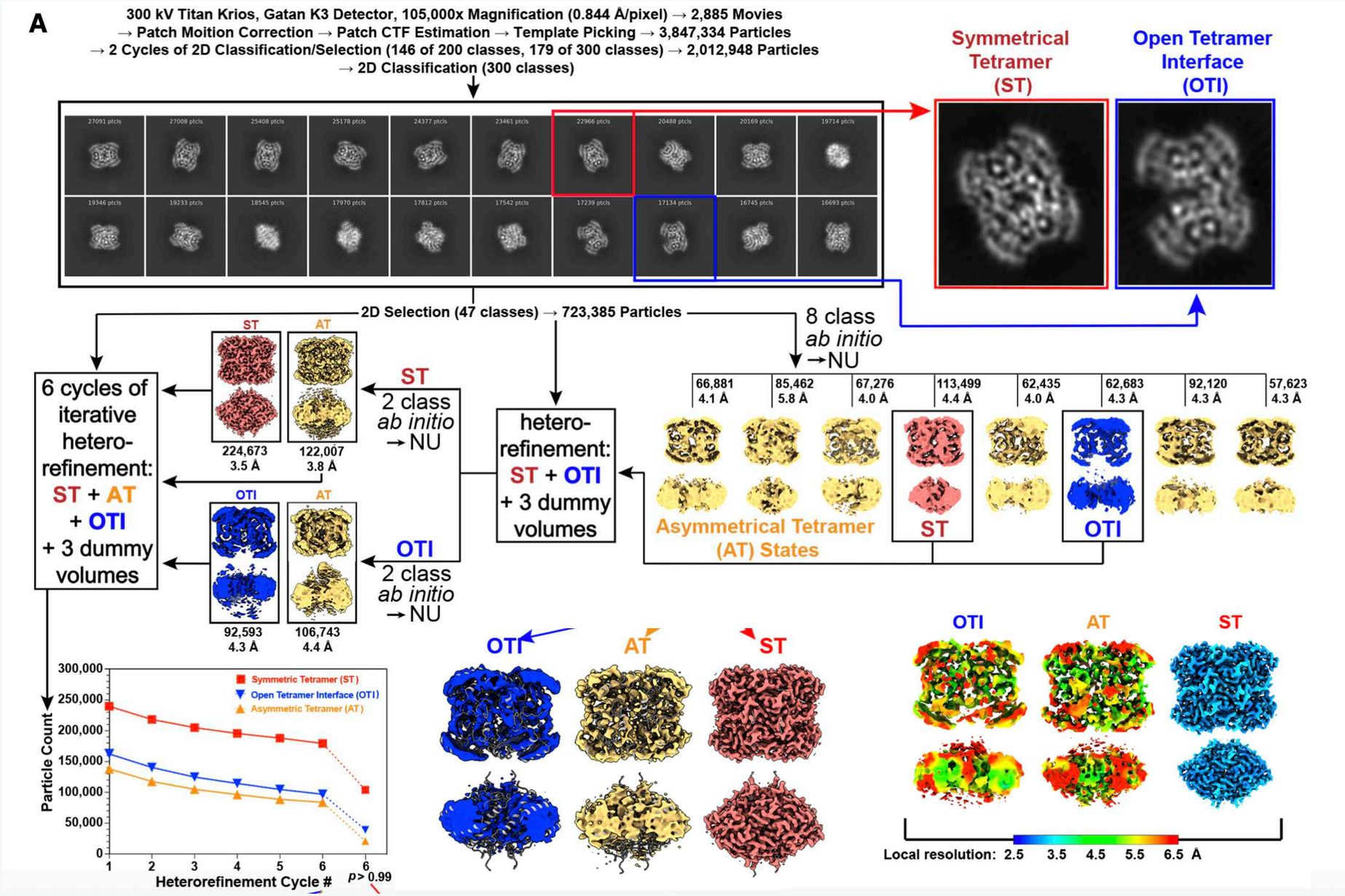
- Require a user-defined number of discrete states.
- Average out any other heterogeneity.



Two rounds of 3D classification:
10 classes, then 9 classes.

Frank & Ourmazd,
Methods (2016)

Discrete classes in cryo-EM lose information.



Li et al , *EMBO J.* (2022).

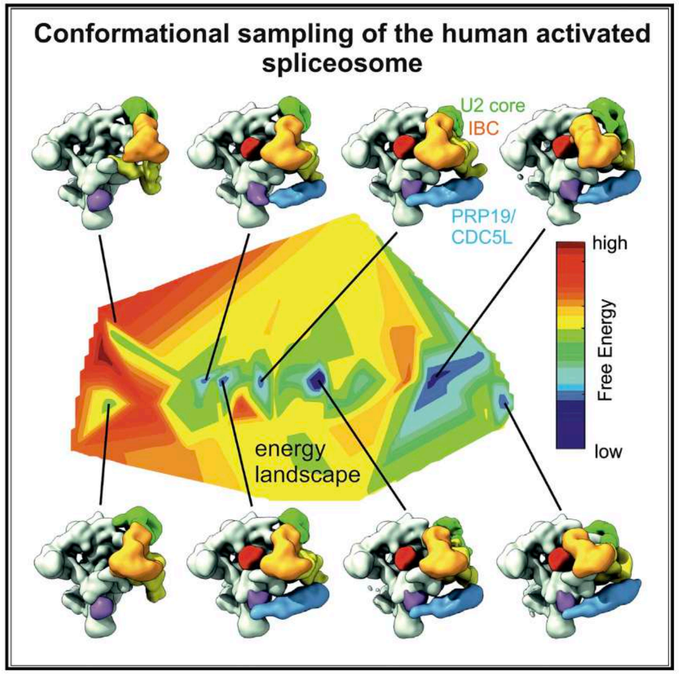
Principal component analysis (PCA) for continuous heterogeneity in SPA.

Article

Cell

Structure and Conformational Dynamics of the Human Spliceosomal B^{act} Complex

Graphical Abstract



Authors

David Haselbach, Ilya Komarov, Dmitry E. Agafonov, ..., Berthold Kastner, Reinhard Lührmann, Holger Stark

Correspondence

reinhard.luehrmann@mpi-bpc.mpg.de (R.L.),
hstark1@gwdg.de (H.S.)

In Brief

A new approach to analyzing cryo-EM data reports on conformational dynamics in the human spliceosome.

Haselbach et al., 2018, Cell *172*, 454–464
January 25, 2018 © 2018 Elsevier Inc.
https://doi.org/10.1016/j.cell.2018.01.010

Acta Cryst D STRUCTURAL BIOLOGY

ISSN 2059-7983
Acta Cryst. (2021). D**77**, 835–839

Principal component analysis is limited to low-resolution analysis in cryoEM

Carlos Oscar S. Sorzano* and Jose Maria Carazo

New Results

Posted November 01, 2023.

Follow this preprint

A Bayesian Framework for Cryo-EM Heterogeneity Analysis using Regularized Covariance Estimation

Marc Aurèle Gilles, Amit Singer

doi: https://doi.org/10.1101/2023.10.28.564422

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

ManifoldEM

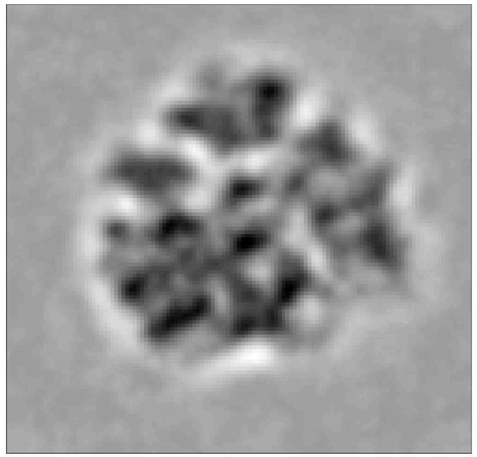
PNAS

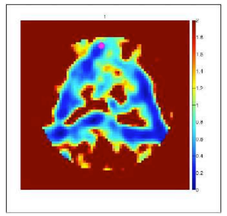
Trajectories of the ribosome as a Brownian nanomachine

Ali Dashti[a,1], Peter Schwander[a,1], Robert Langlois[b], Russell Fung[a], Wen Li[b], Ahmad Hosseinizadeh[a], Hstau Y. Liao[b], Jesper Pallesen[c,2], Gyanesh Sharma[b,3], Vera A. Stupina[d], Anne E. Simon[d], Jonathan D. Dinman[d], Joachim Frank[b,c,4], and Abbas Ourmazd[a,1,4]

[a]Department of Physics, University of Wisconsin, Milwaukee, WI 53211; [b]Department of Biochemistry and Molecular Biophysics, and [c]Howard Hughes Medical Institute, Columbia University, New York, NY 10032; and [d]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742

Contributed by Joachim Frank, October 8, 2014 (sent for review September 10, 2014)





Dashti et al, *PNAS* (2014).

ManifoldEM

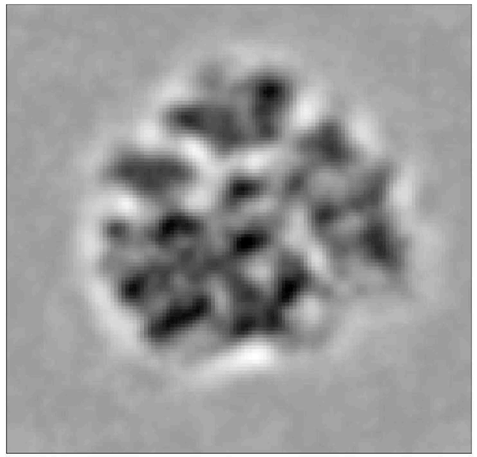
PNAS

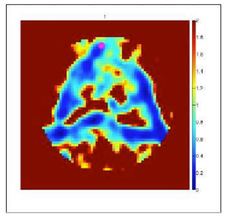
Trajectories of the ribosome as a Brownian nanomachine

Ali Dashti[a,1], Peter Schwander[a,1], Robert Langlois[b], Russell Fung[a], Wen Li[b], Ahmad Hosseinizadeh[a], Hstau Y. Liao[b], Jesper Pallesen[c,2], Gyanesh Sharma[b,3], Vera A. Stupina[d], Anne E. Simon[d], Jonathan D. Dinman[d], Joachim Frank[b,c,4], and Abbas Ourmazd[a,1,4]

[a]Department of Physics, University of Wisconsin, Milwaukee, WI 53211; [b]Department of Biochemistry and Molecular Biophysics, and [c]Howard Hughes Medical Institute, Columbia University, New York, NY 10032; and [d]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742

Contributed by Joachim Frank, October 8, 2014 (sent for review September 10, 2014)

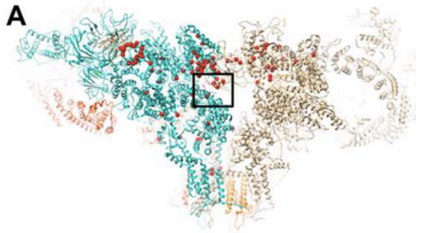


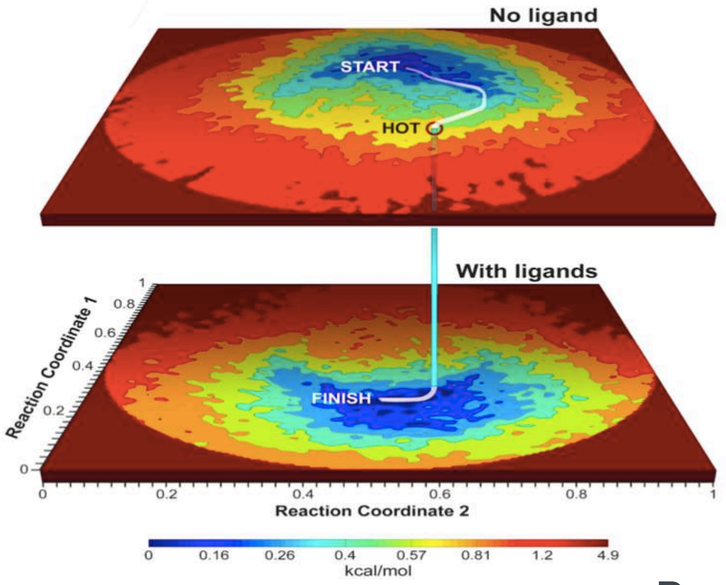


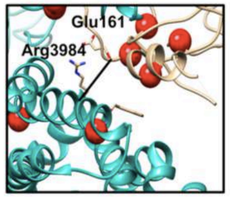
Dashti et al, *PNAS* (2014).

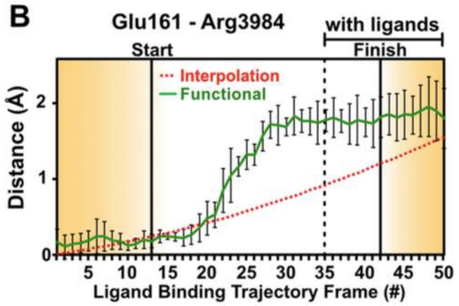
Retrieving functional pathways of biomolecules from single-particle snapshots

Ali Dashti[1,9], Ghoncheh Mashayekhi[1,9], Mrinal Shekhar[2,3], Danya Ben Hail[4], Salah Salah[4,5,6], Peter Schwander[1], Amedee des Georges[4,5,6✉], Abhishek Singharoy[3✉], Joachim Frank[7,8✉] & Abbas Ourmazd[1✉]







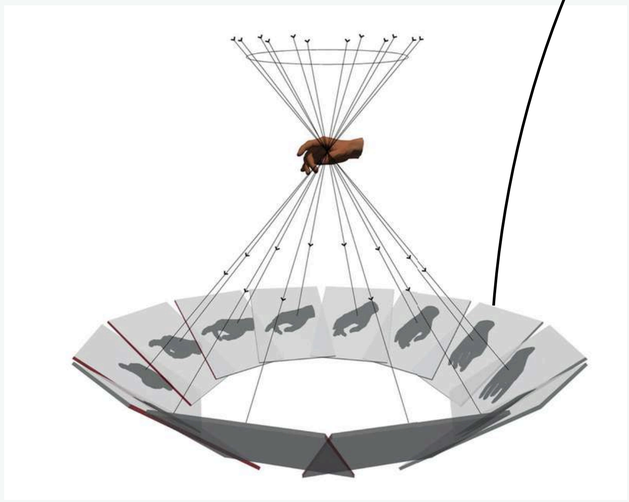


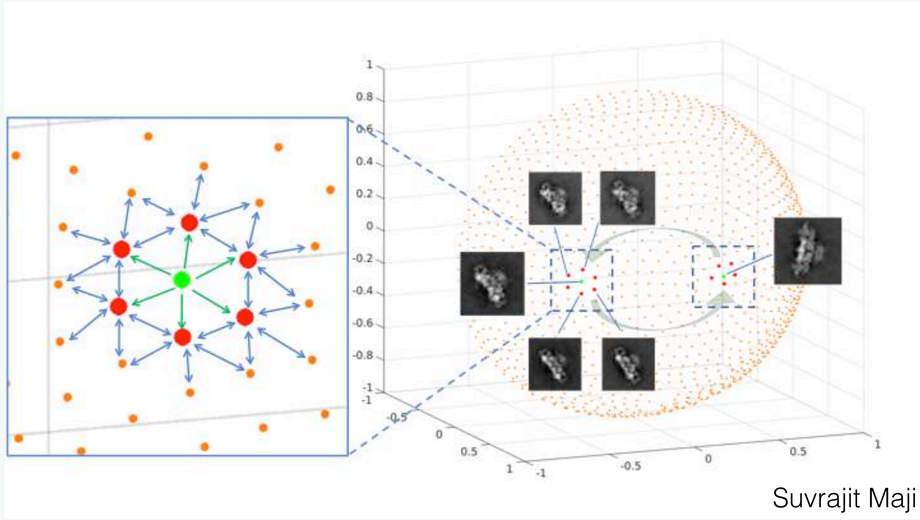
Dashti et al, *Nat Comm* (2020).

One proposed solution: **The Manifold Embedding Method in cryo-EM from Ourmazd/Frank labs**

The manifold embedding method uses geometric machine learning to use the information about molecular conformational heterogeneity already available in the particle images.

1) Particle images are clustered into ‘projection directions’
2) The Euclidean distance between each image is calculated
3) Spectral decomposition with a diffusion mapping[1] step
4) Nonlinear Laplacian spectral analysis (NLSA)[2] and ordering of images
5) Combining the results from individual projection directions into a united depiction of the energy landscape for the whole orientation sphere.[3]





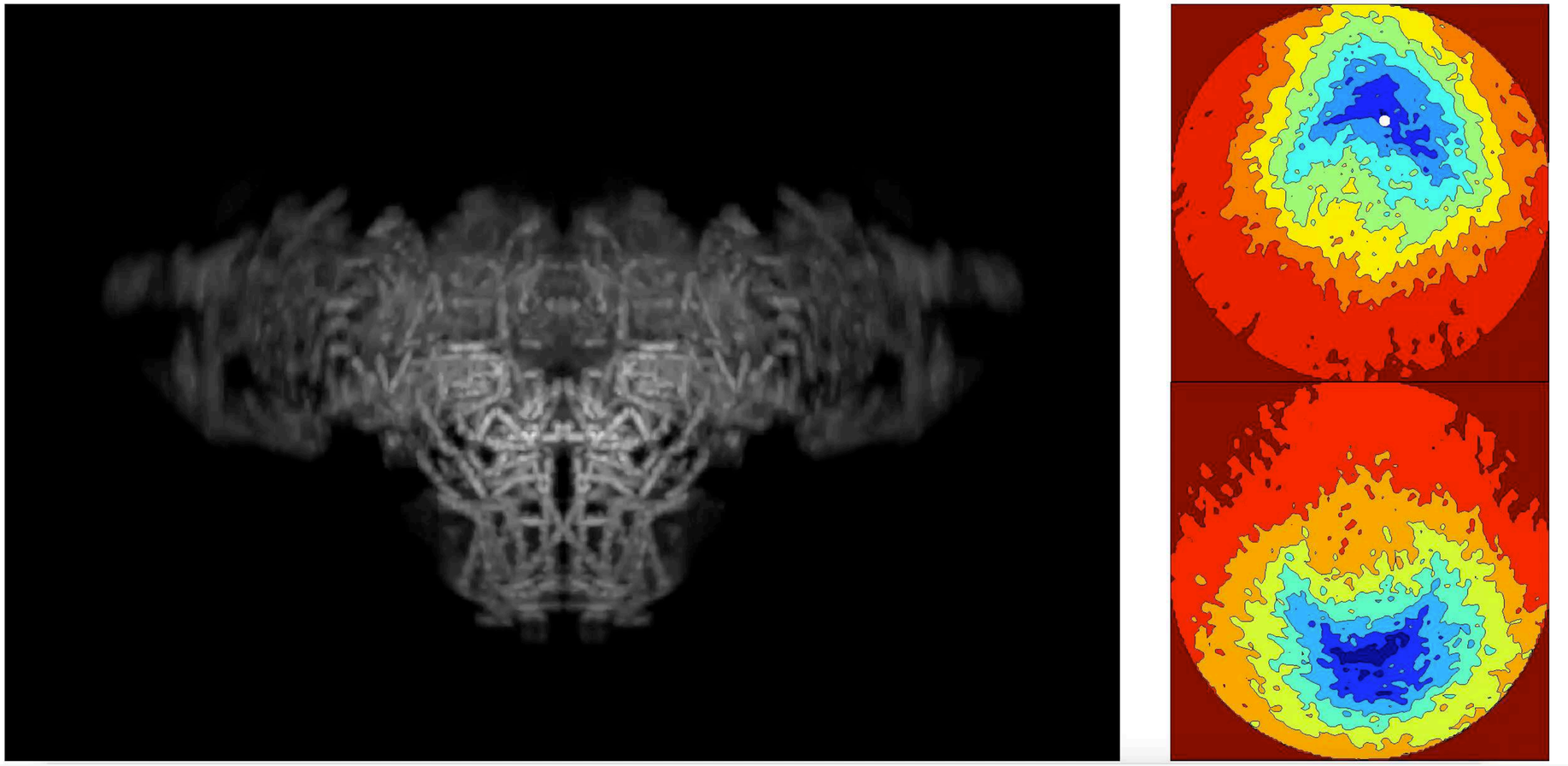
Suvrajit Maji

1. Coifman, R. R., Kevrekidis, I. G., Lafon, S., Maggioni, M. & Nadler, B. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Multiscale Model. Simul.* **7,** 842–864 (2008).

2. Giannakis, D. & Majda, A. J. Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci.* **109,** 2222–2227 (2012).

3. Maji, S., Liao, H., Dashti, A., Mashayekhi, G., Ourmazd, A., & J. Frank. Propagation of Conformational Coordinates Across Angular Space in Mapping the Continuum of States from Cryo-EM Data by Manifold Embedding. *J. Chem. Inf Model.* (2020).

Manifold Embedding: The Ryanodine Receptor



Dashti et al, *Nat Comm* (2020).

RELION: Multi-body Refinement

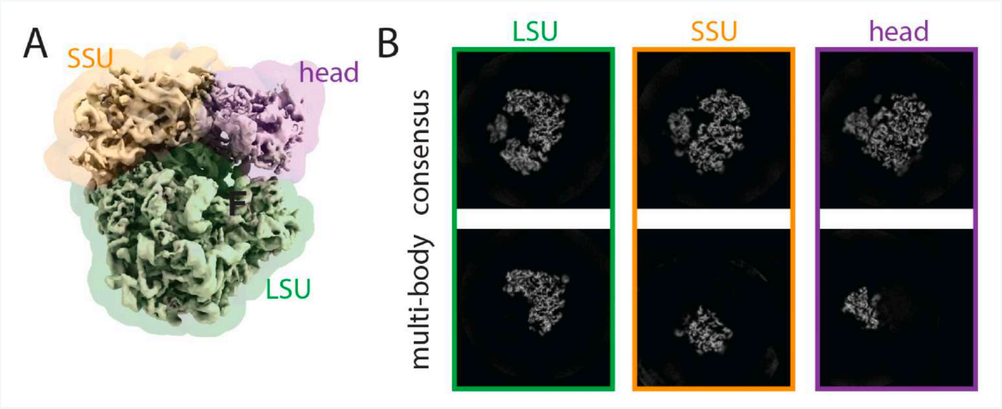
eLIFE elifesciences.org TOOLS AND RESOURCES

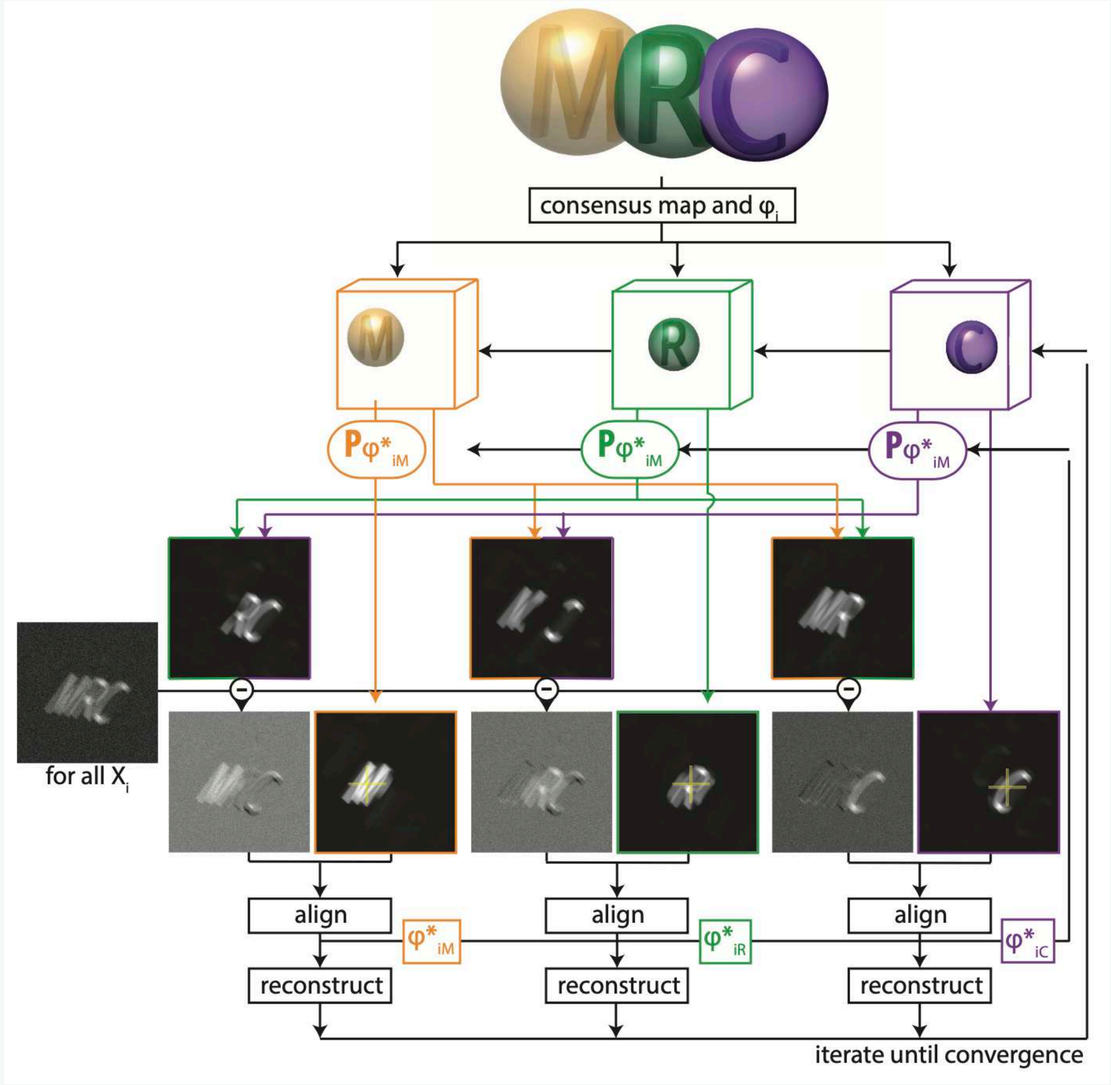
Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION

Takanori Nakane[1], Dari Kimanius[2], Erik Lindahl[2,3], Sjors HW Scheres[1]*

[1]MRC Laboratory of Molecular Biology, Cambridge, United Kingdom; [2]Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, Stockholm, Sweden; [3]Swedish e-Science Research Center, KTH Royal Institute of Technology, Stockholm, Sweden

Nakane *et al.* eLife 2018;7:e36861. DOI: https://doi.org/10.7554/eLife.36861





cryoDRGN: The first NN approach.

ARTICLES
https://doi.org/10.1038/s41592-020-01049-4

nature methods

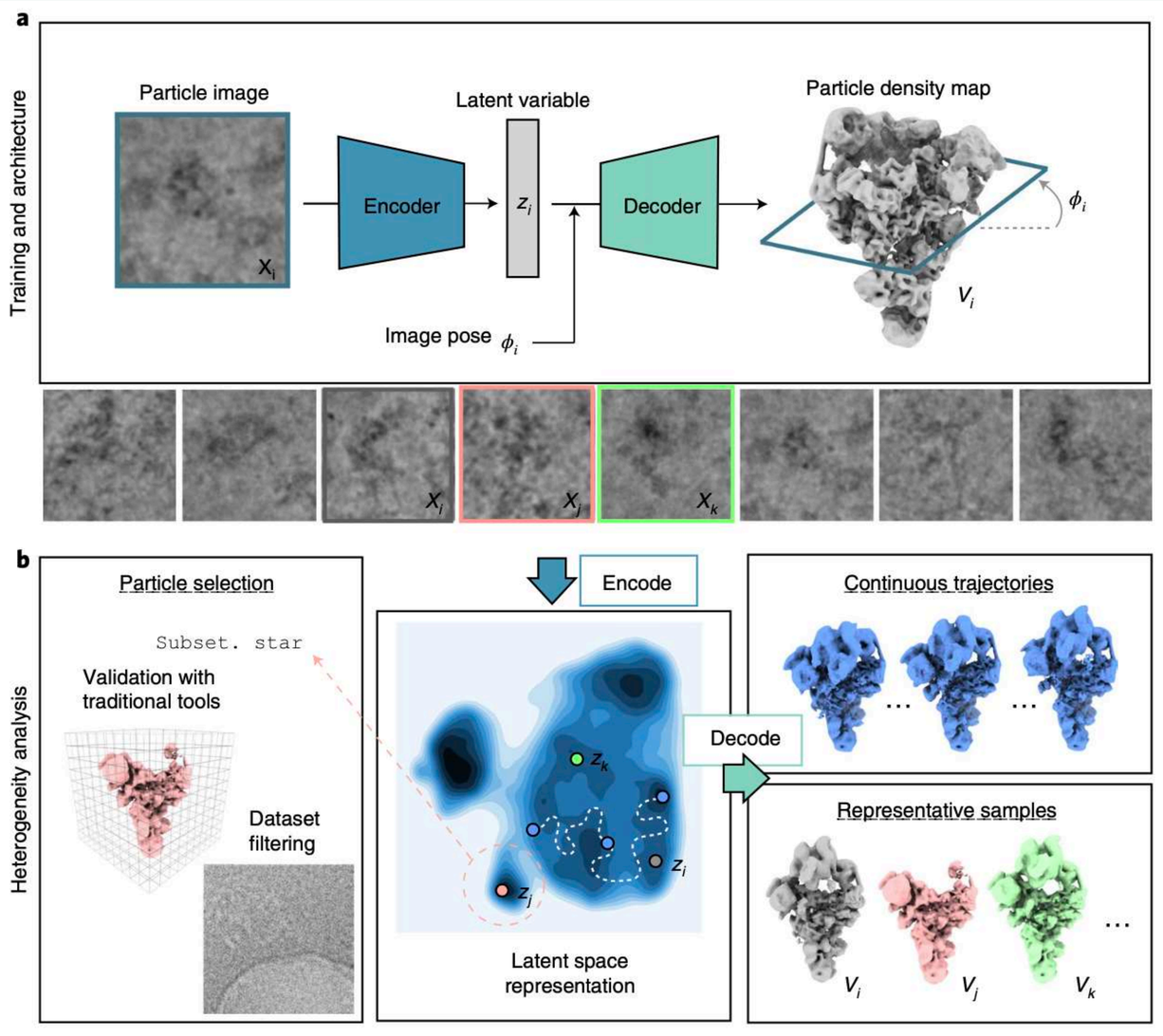
Check for updates

CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks

Ellen D. Zhong[1,2], Tristan Bepler[1,2], Bonnie Berger[2,3] ✉ and Joseph H. Davis[1,4] ✉

Cryo-electron microscopy (cryo-EM) single-particle analysis has proven powerful in determining the structures of rigid macromolecules. However, many imaged protein complexes exhibit conformational and compositional heterogeneity that poses a major challenge to existing three-dimensional reconstruction methods. Here, we present cryoDRGN, an algorithm that leverages the representation power of deep neural networks to directly reconstruct continuous distributions of 3D density maps and map per-particle heterogeneity of single-particle cryo-EM datasets. Using cryoDRGN, we uncovered residual heterogeneity in high-resolution datasets of the 80S ribosome and the RAG complex, revealed a new structural state of the assembling 50S ribosome, and visualized large-scale continuous motions of a spliceosome complex. CryoDRGN contains interactive tools to visualize a dataset's distribution of per-particle variability, generate density maps for exploratory analysis, extract particle subsets for use with other tools and generate trajectories to visualize molecular motions. CryoDRGN is open-source software freely available at http://cryodrgn.csail.mit.edu.

NATURE METHODS | VOL 18 | FEBRUARY 2021 | 176–185 | www.nature.com/naturemethods



3DVA in cryoSPARC

Journal of Structural Biology 213 (2021) 107702

Contents lists available at ScienceDirect

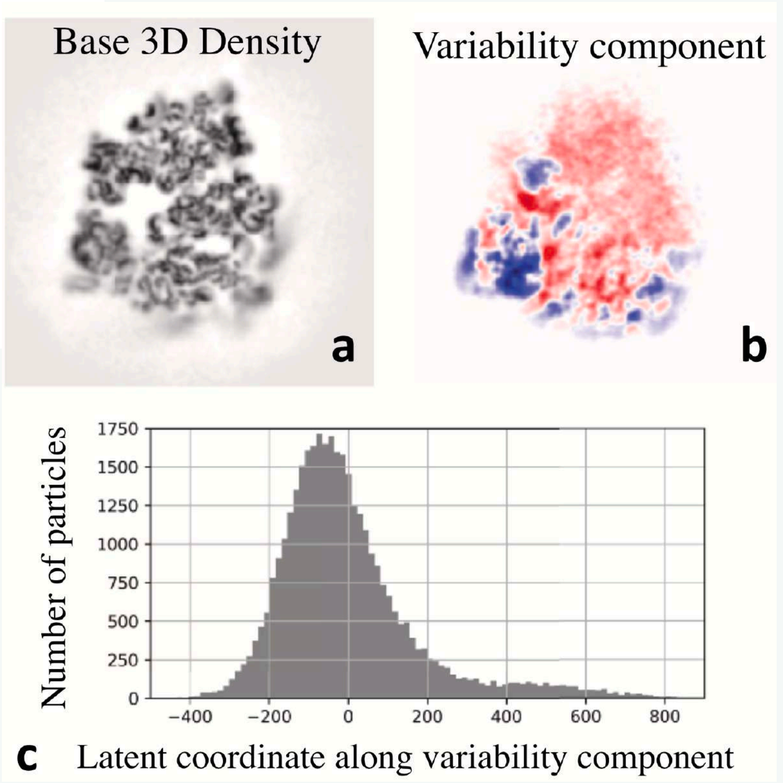
Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM

Ali Punjani [a,b,c,*], David J. Fleet [a,b,*]

[a] Department of Computer Sciences, University of Toronto M5S 3G4, Canada
[b] Vector Institute, 710-661 University Ave., Toronto M5G 1M1, Canada
[c] Structura Biotechnology Inc., 129-100 College Ave., Toronto M5G 1L5, Canada



- Very straightforward to use.
- Limited to linear motions.

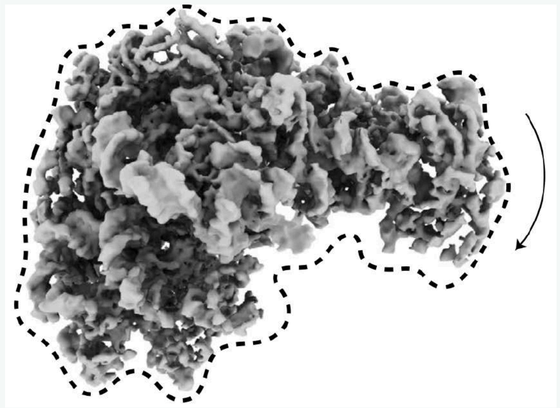
Gaussian Mixture Model method - EMAN2

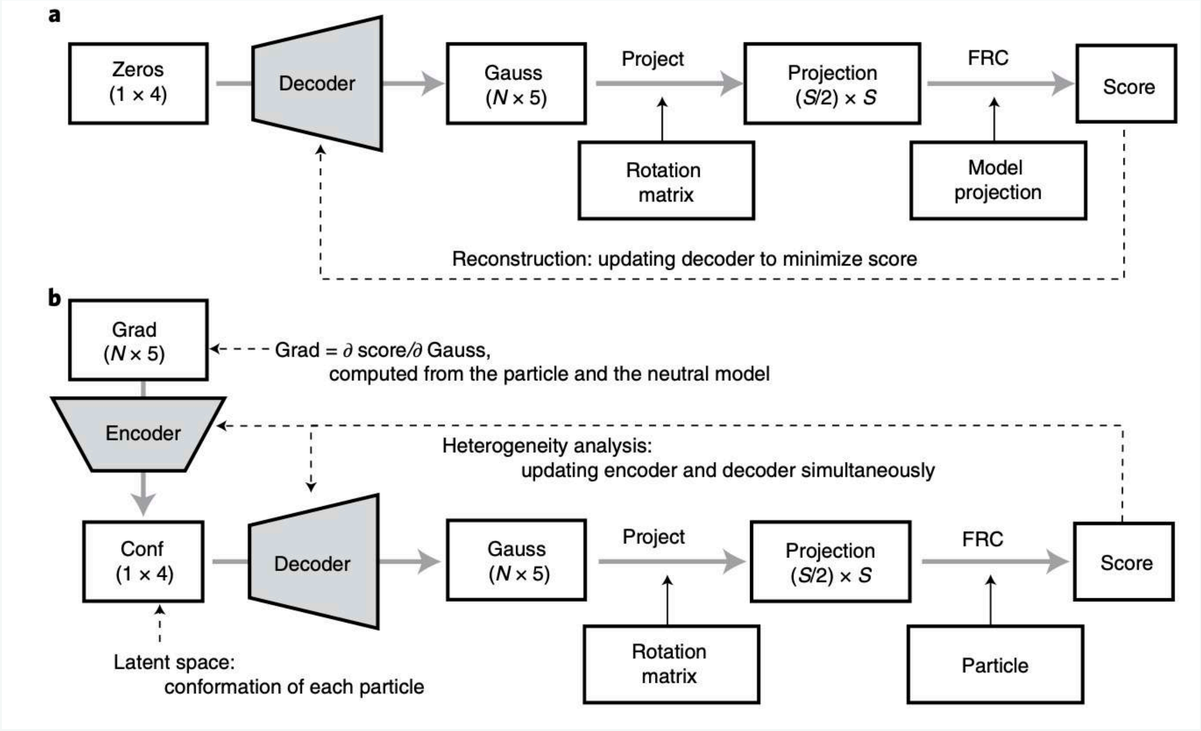
ARTICLES
https://doi.org/10.1038/s41592-021-01220-5
nature methods
Check for updates

Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM

Muyuan Chen and Steven J. Ludtke ✉

930 NATURE METHODS | VOL 18 | AUGUST 2021 | 930–936 | www.nature.com/naturemethods





- Incorporating the Gaussian Mixture Model (GMM) gets us toward atomic / molecular information of the dynamics.

CryoSparc's more complex tool for heterogeneity analysis:

nature methods

Article

https://doi.org/10.1038/s41592-023-01853-8

3DFlex: determining structure and motion of flexible proteins from cryo-EM

Received: 28 July 2022

Accepted: 16 March 2023

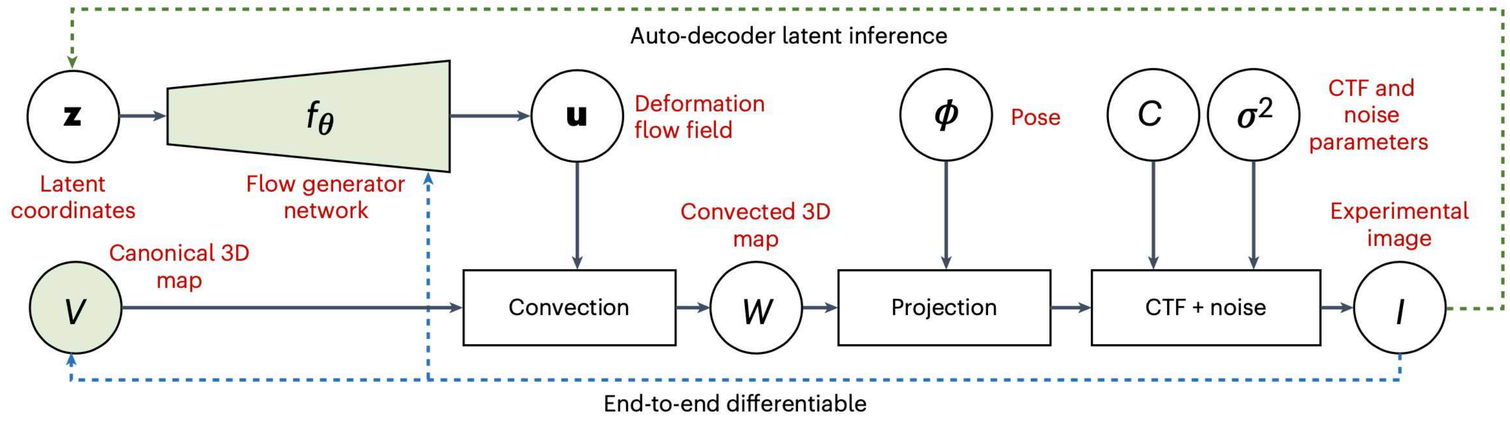
Published online: 11 May 2023

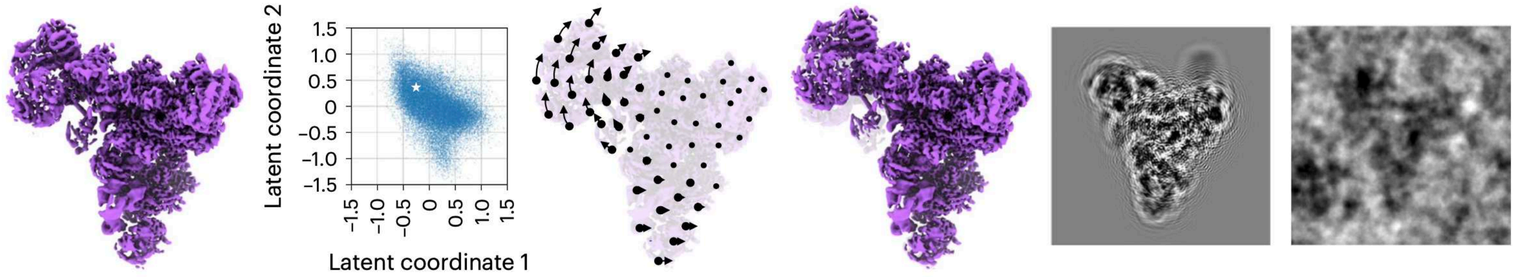
Check for updates

Ali Punjani [1,2,3] & David J. Fleet [1,2,4]

Modeling flexible macromolecules is one of the
single-particle cryogenic-electron microscopy (
to illuminate fundamental questions in structur
Three-Dimensional Flexible Refinement (3DFlex

Nature Methods | Volume 20 | June 2023 | 860–870





CryoSparc's more complex tool for heterogeneity analysis:

nature methods

Article

3DFlex: deterr
flexible protei

Received: 28 July 2022

Accepted: 16 March 2023

Published online: 11 May 2023

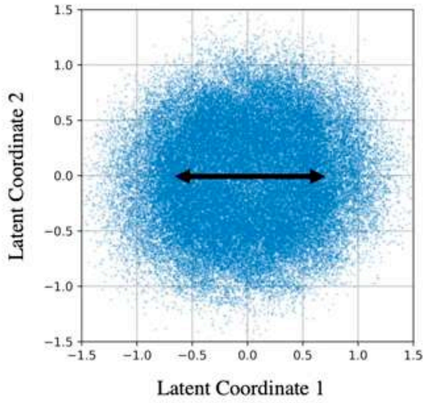
Check for updates

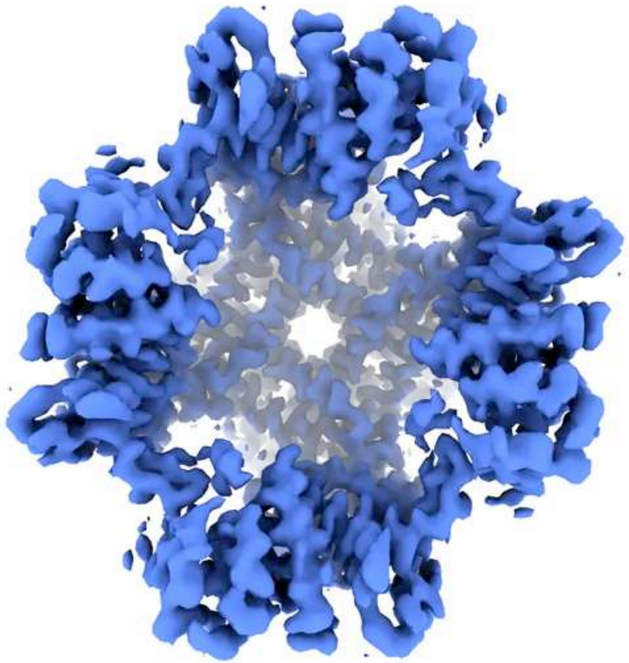
Nature Methods | Volume 20 | June

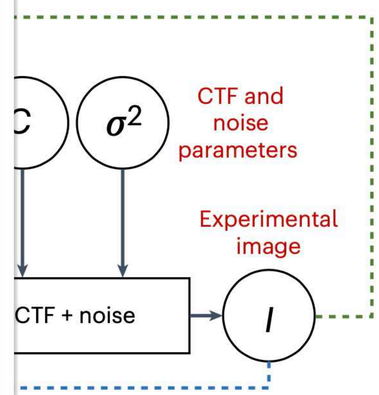
3D Flexible Refinement

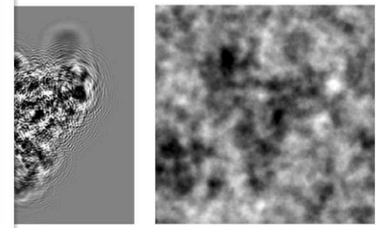
TRPV1 Ion Channel
EMPIAR-10059

Latent coordinate 1



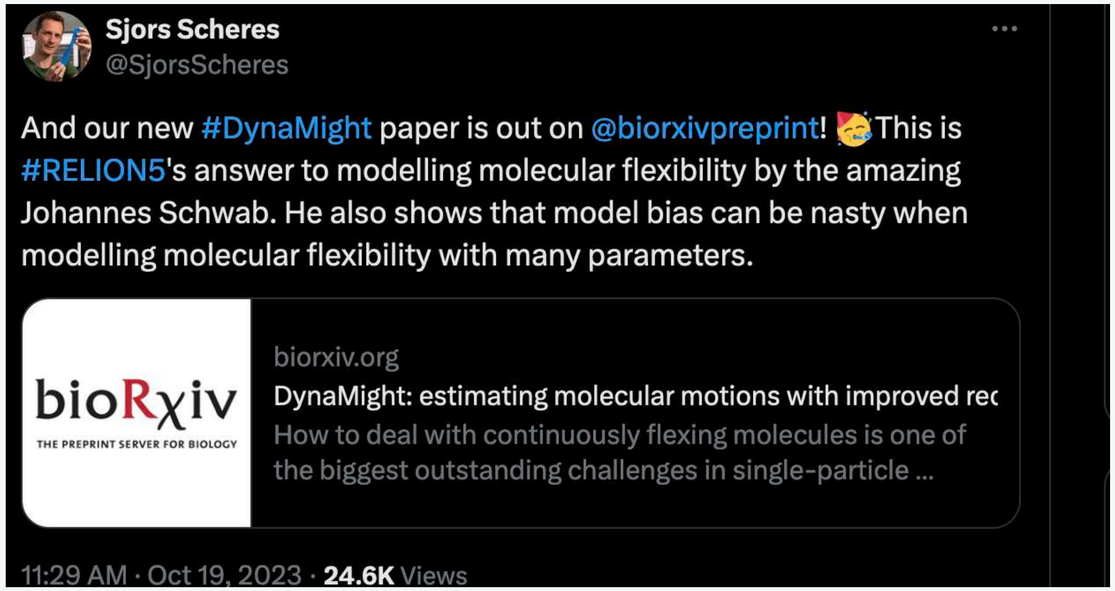


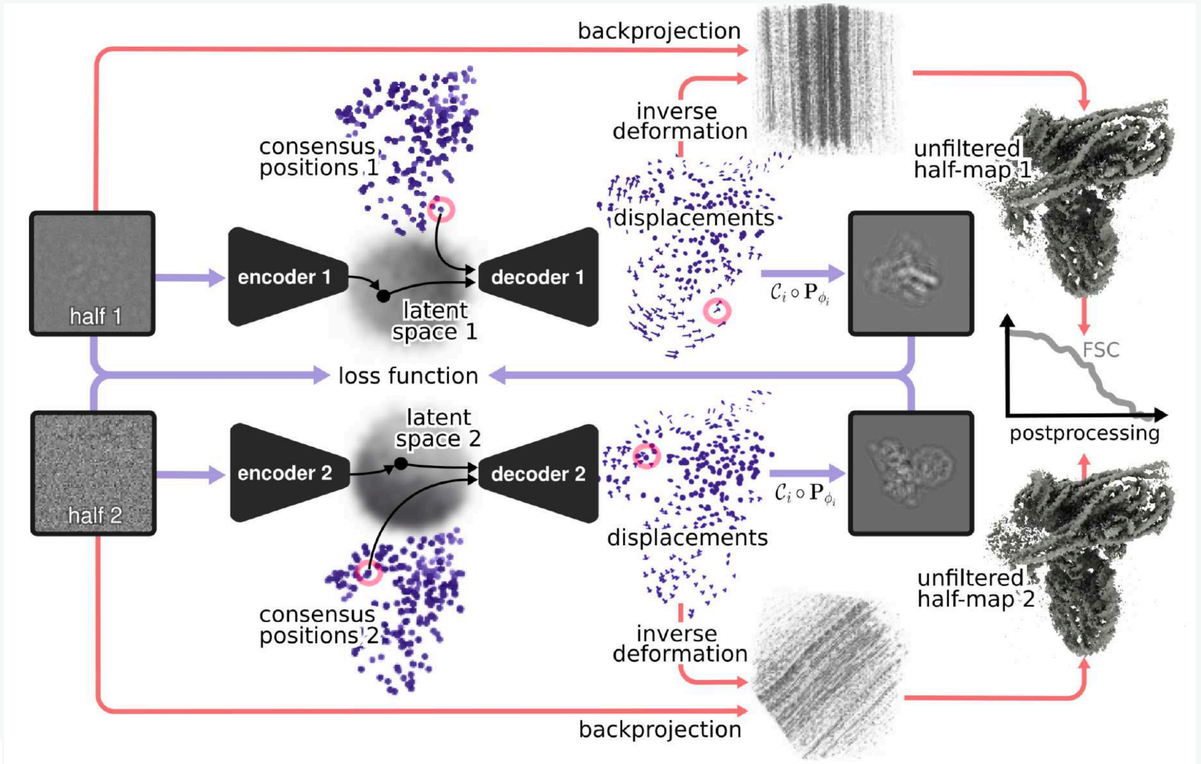




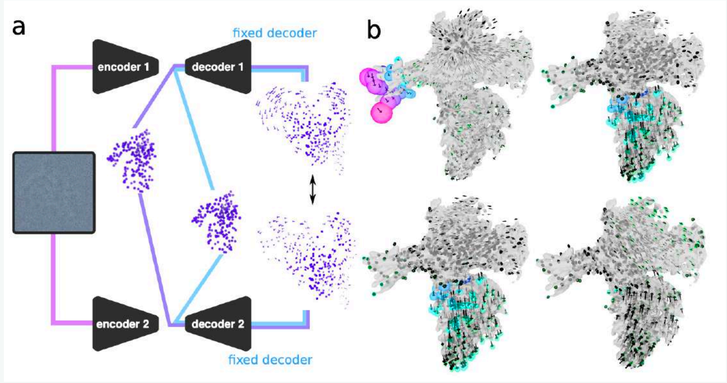
Canonical 3D map | Latent space | Deformation flow field | Convected 3D map | Projected, CTF-corrupted image | Experimental image

RELION also has a new ML-based solution for this problem





Has a method for error estimation of the deformations.



Implemented in RELION-5

Multiple methods available in Scipion

https://github.com/scipion-em/scipion-em-continuousflex

nature communications

Nature Communications | (2023)14:154

Article

https://doi.org/10.1038/s41467-023-35791-y

Estimating conformational landscapes from Cryo-EM particles by 3D Zernike polynomials

Received: 1 June 2022

Accepted: 29 December 2022

Published online: 11 January 2023

Check for updates

D. Herreros [1], R. R. Lederman [2], J. M. Krieger [1], A. Jiménez-Moreno [1], M. Martínez [1], D. Myška [3], D. Strelak [1,4], J. Filipovic [3], C. O. S. Sorzano [1,5] & J. M. Carazo [1,5]

The new developments in Cryo-EM Single Particle Analysis are helping us to understand how the macromolecular structure and function meet to drive

jmb

Journal of Molecular Biology

Volume 435, Issue 9, 1 May 2023, 167951

Research Article

MDSPACE: Extracting Continuous Conformational Landscapes from Cryo-EM Single Particle Datasets Using 3D-to-2D Flexible Fitting based on Molecular Dynamics Simulation

Rémi Vuillemot [1 6], Alex Mirzaei [1], Mohamad Harastani [1], Ilyes Hamitouche [1], Léo Fréchin [2], Bruno P. Klaholz [2], Osamu Miyashita [3], Florence Tama [3 4 5], Isabelle Rouiller [6], Slavica Jonic [1]

TikTok and Cows?

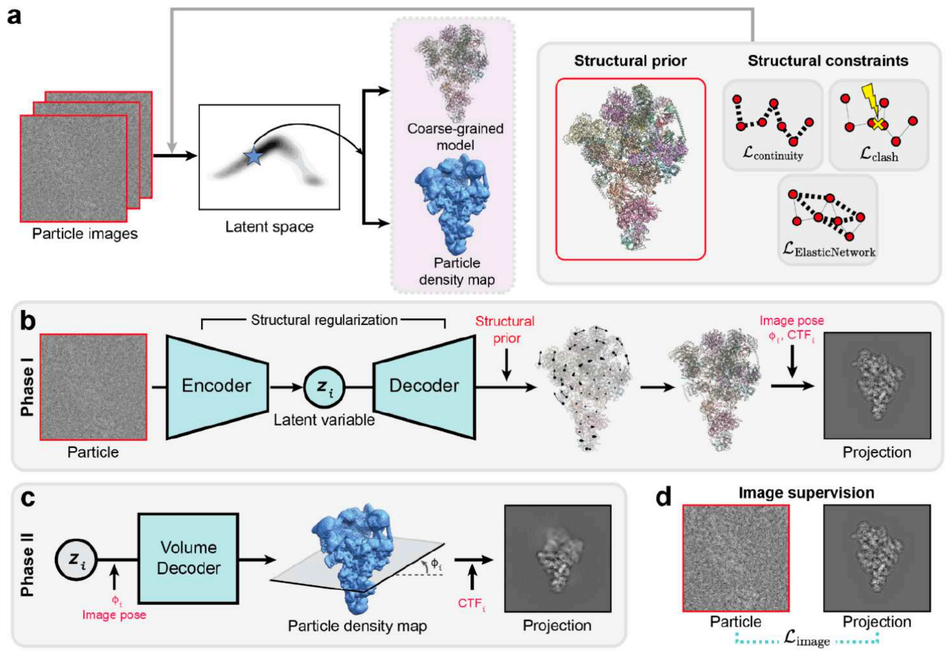
CryoSTAR: Leveraging Structural Prior and Constraints for Cryo-EM Heterogeneous Reconstruction

Yilai Li[1#], Yi Zhou[1#], Jing Yuan[1#], Fei Ye[1], Quanquan Gu[1*]

[1]ByteDance Research

[#]Contributed Equally

[*]Correspondence to: quanquan.gu@bytedance.com



CowScape: Quantitative reconstruction of the conformational landscape of biological macromolecules from cryo-EM data

Felix Lambrecht [1], Andreas Kröpelin [2], Mario Lüttich [1], Michael Habeck [2,1,*], David Haselbach [1,3,*], Holger Stark [1,*]

February 20, 2024

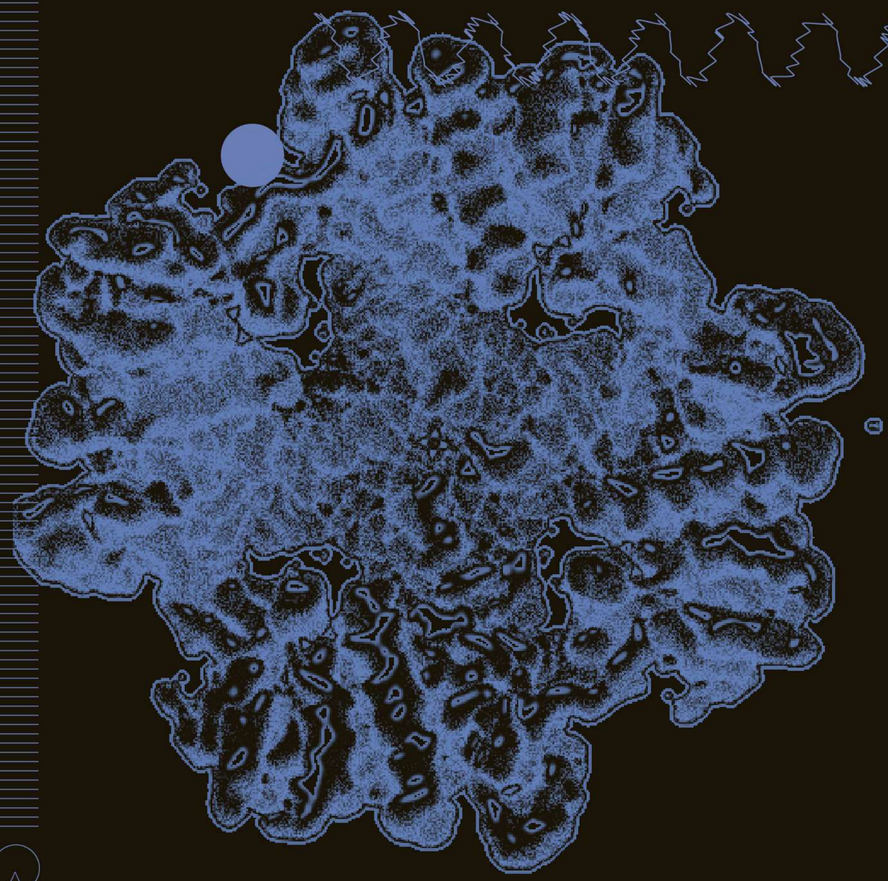
[1]Max Planck Institute for Multidisciplinary Sciences, 37077 Göttingen, Germany
[2]Microscopic Image Analysis Group, Jena University Hospital, 07743 Jena, Germany
[3]Institute for Molecular Pathology, Vienna, Austria

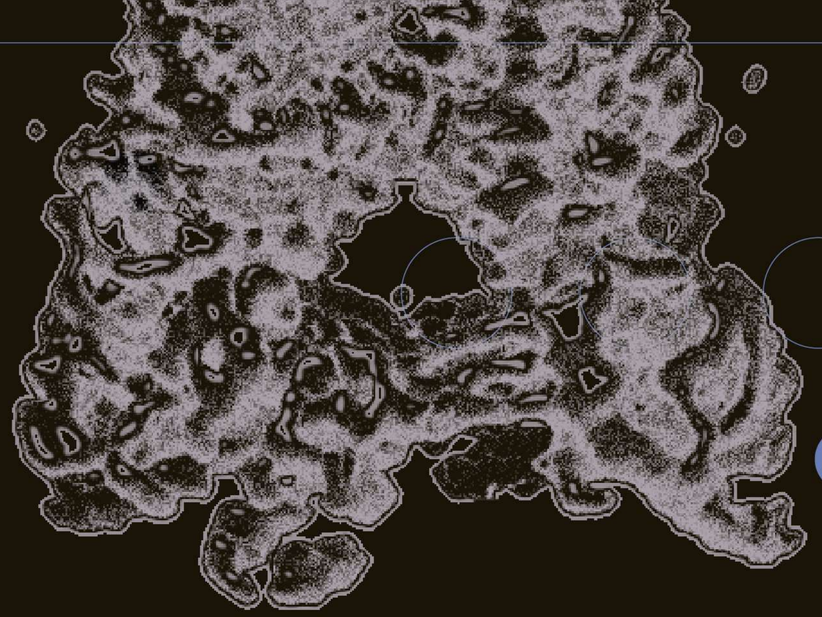
*E-Mail: hstark1@gwdg.de; david.haselbach@imp.ac.at; michael.habeck@uni-jena.de

Abstract

Cryo-EM data processing typically focuses on the structure of the main conformational state under investigation and discards images that belong to other states. This approach can reach atomic resolution, but ignores vast amounts of valuable information about the underlying conformational ensemble and its dynamics. CowScape analyzes an entire cryo-EM dataset and thereby obtains a quantitative description of structural variability of macromolecular complexes that represents the biochemically relevant conformational space. By combining extensive image classification with principal component analysis (PCA) of the classified 3D volumes and kernel density estimation, CowScape can be used as a quantitative tool to analyze this variability. PCA projects all 3D structures along the major modes spanning a low-dimensional space that captures a large portion of structural variability. The number of particle images in a given state can be used to calculate an energy landscape based on kernel density estimation and Boltzmann

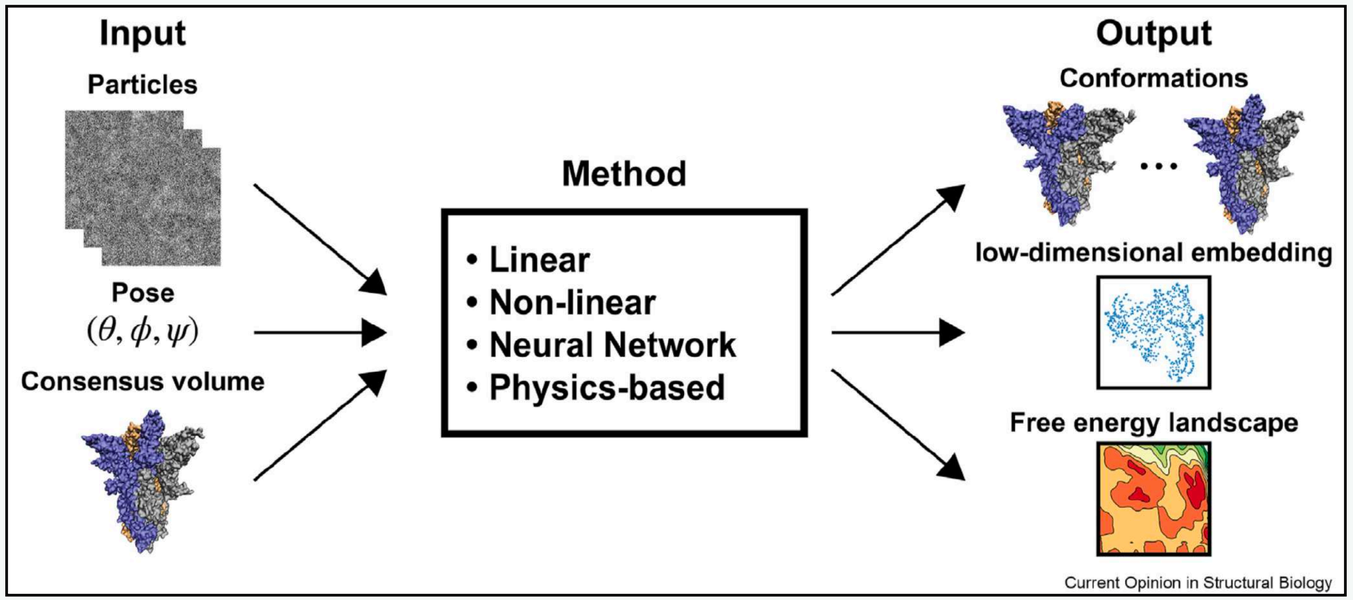
Sonya **Hanson**
shanson@flatironinstitute.org





Practical considerations when using continuous heterogeneity methods.

Most methods take in particle stacks after 3D Refinement from a previous tool.



- Many methods are already integrated into other tools that already do 3D Refinement (cryoSPARC/RELION/EMAN2) so this is straightforward
- cryoDRGN has a friendly easily installable Python pipeline and plays a little nicer with cryoSPARC inputs

Procedures to jointly optimize pose and conformation are still being improved.

ICCV 2021

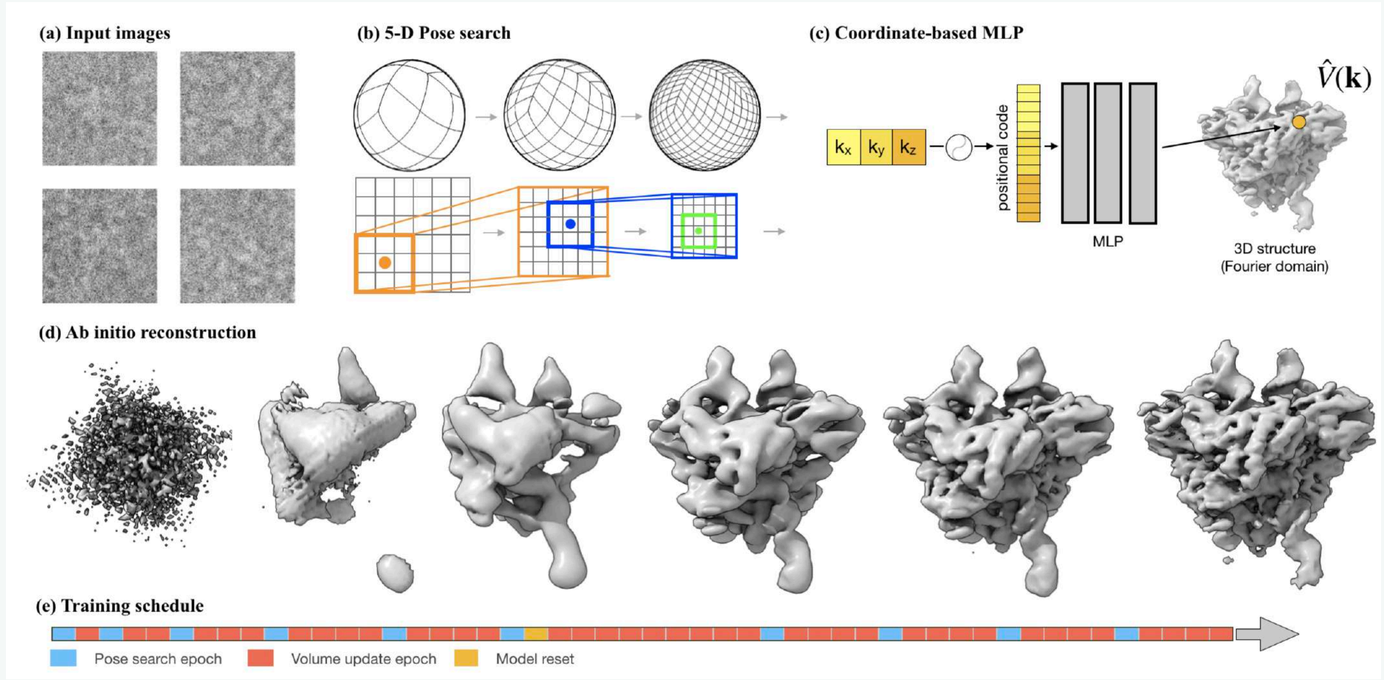
CryoDRGN2: *Ab initio* neural reconstruction of 3D protein structures from real cryo-EM images

Ellen D. Zhong
MIT
zhonge@mit.edu

Adam Lerer
Facebook AI
alerer@fb.com

Joseph H. Davis
MIT
jhdavis@mit.edu

Bonnie Berger
MIT
bab@mit.edu



Procedures to jointly optimize pose and conformation are still being improved.

IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, VOL. 7, 2021 759

CryoGAN: A New Reconstruction Paradigm for Single-Particle Cryo-EM Via Deep Adversarial Learning

Harshit Gupta, Michael T. McCann, *Member, IEEE*, Laurène Donati, and Michael Unser, *Fellow, IEEE*

Amortized Inference for Heterogeneous Reconstruction in Cryo-EM

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Axel Levy
Stanford University

Gordon Wetzstein
Stanford University

Julien Martel
Stanford University

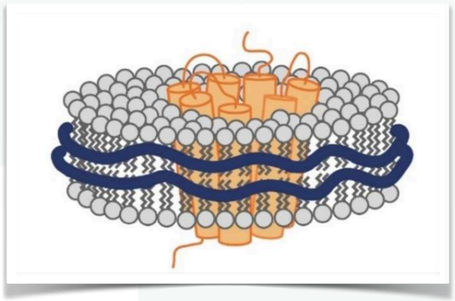
Frédéric Poitevin
SLAC National Accelerator Laboratory

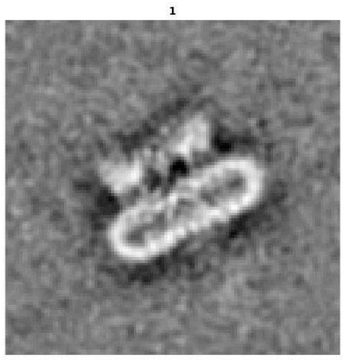
Ellen D. Zhong*
Princeton University

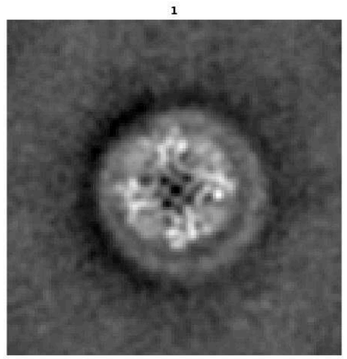
Nanodiscs can also be a challenge for continuous heterogeneity methods.

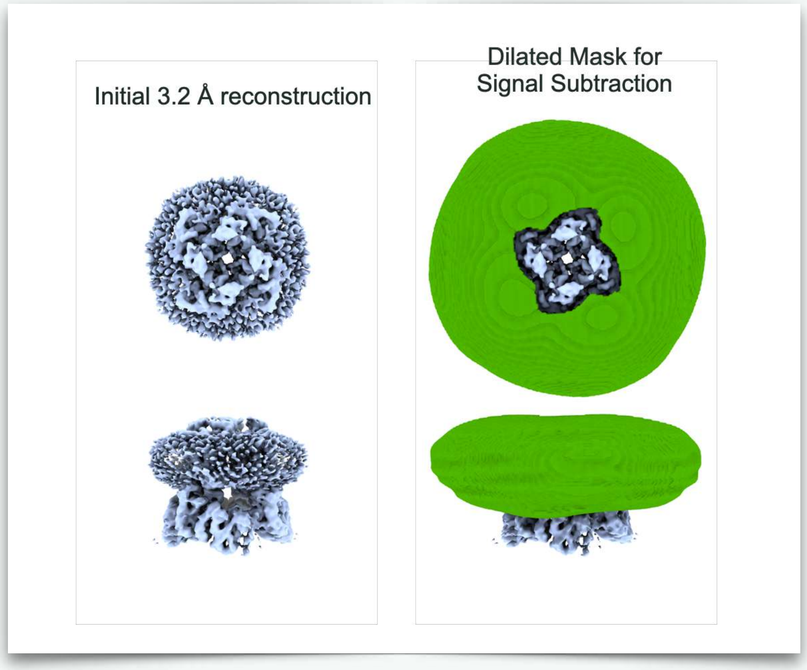
Analysis with ManifoldEM Matlab Code

rTRPV1 + DkTx/RtX
in nanodiscs
(EMPIAR 10059)



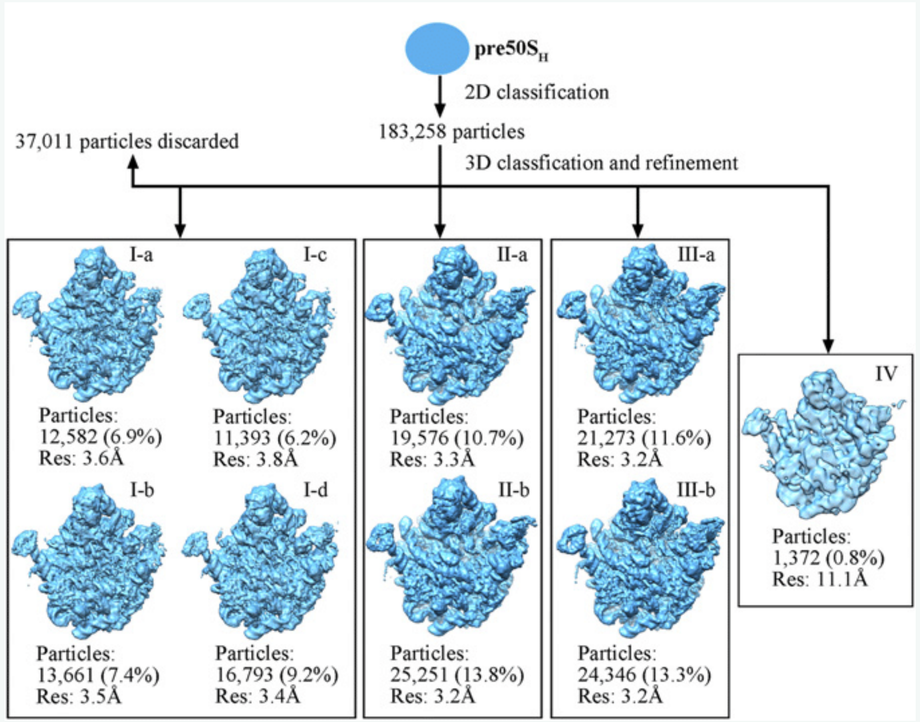


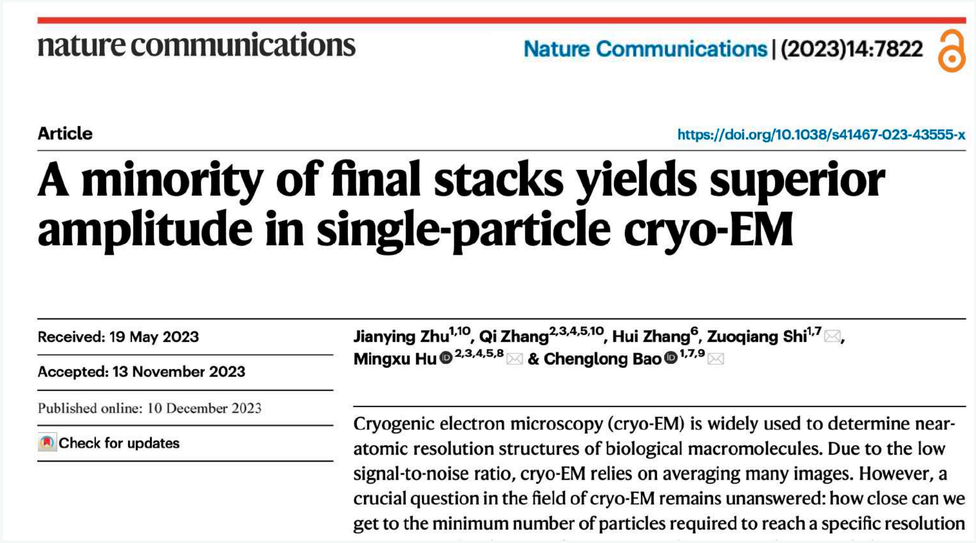




Using all vs. a subset of your data.

- Getting rid of junk is important.
- Making sure there isn't residual compositional heterogeneity is important
- Intuitively you wouldn't want particles too tightly classified in a single conformational state, but this is a little tricky to assess




nature communications
Nature Communications | (2023)14:7822
Article
https://doi.org/10.1038/s41467-023-43555-x

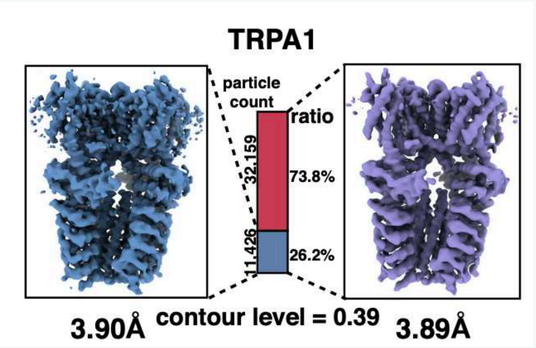
A minority of final stacks yields superior amplitude in single-particle cryo-EM

Received: 19 May 2023
Accepted: 13 November 2023
Published online: 10 December 2023
Check for updates

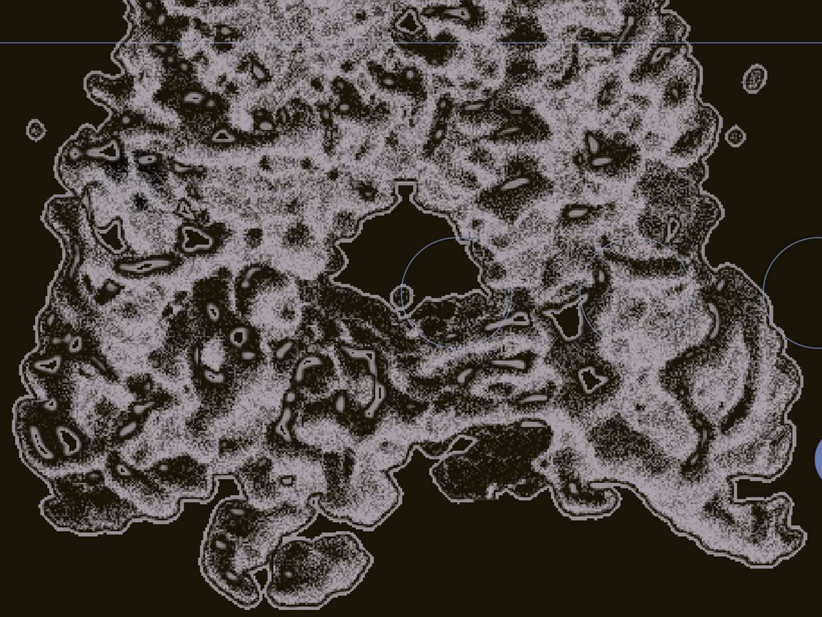
Jianying Zhu[1,10], Qi Zhang[2,3,4,5,10], Hui Zhang[6], Zuoqiang Shi[1,7] ✉, Mingxu Hu[2,3,4,5,8] ✉ & Chenglong Bao[1,7,9] ✉

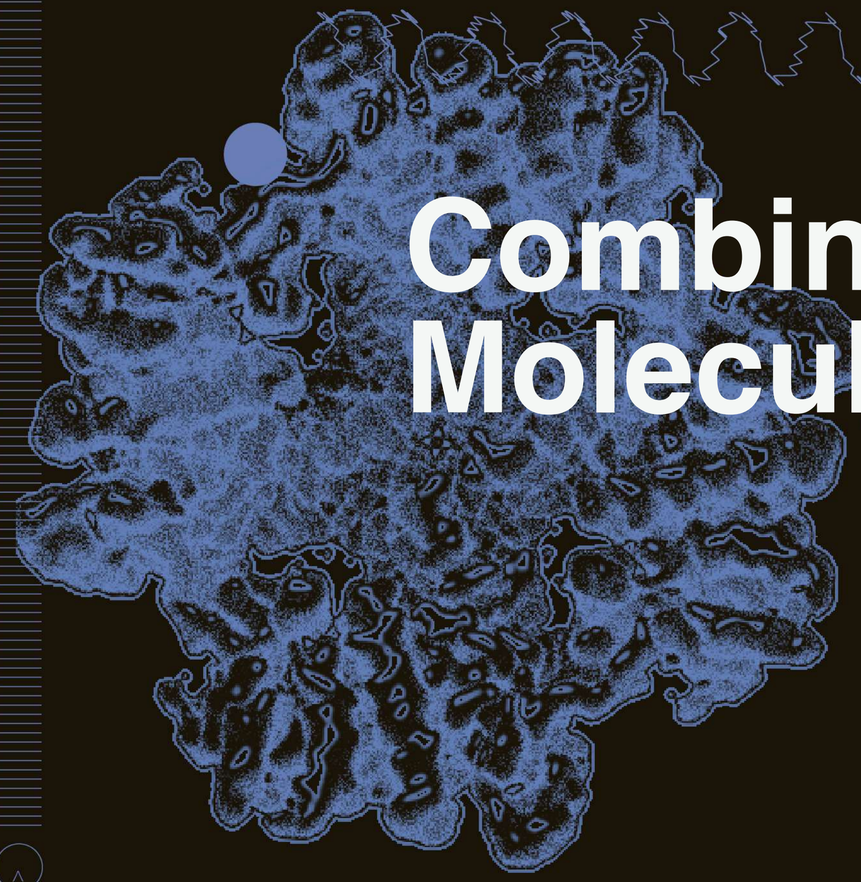
Cryogenic electron microscopy (cryo-EM) is widely used to determine near-atomic resolution structures of biological macromolecules. Due to the low signal-to-noise ratio, cryo-EM relies on averaging many images. However, a crucial question in the field of cryo-EM remains unanswered: how close can we get to the minimum number of particles required to reach a specific resolution

CryoSieve is a new tool that allows you to build a same resolution reconstruction with a fraction of the particles.



Sonya **Hanson**
shanson@flatironinstitute.org

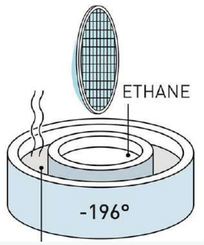




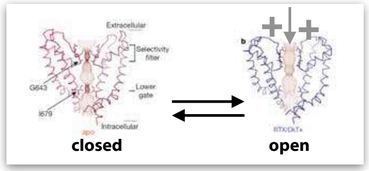
Combining cryo-EM and Molecular Dynamics Simulations

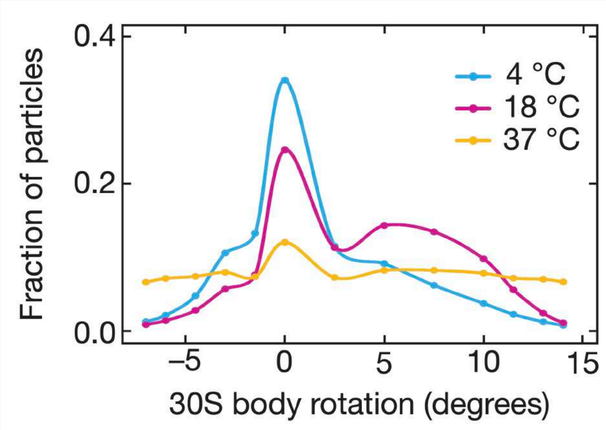
FLATIRON INSTITUTE

How much to worry about the ‘cryo’ of cryo-EM?

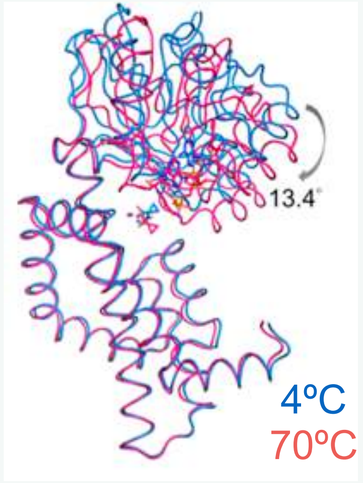


time to freeze during vitrification ~ 0.1 of ms
current estimates of time scale of TRPV1 opening ~ 2-5 ms

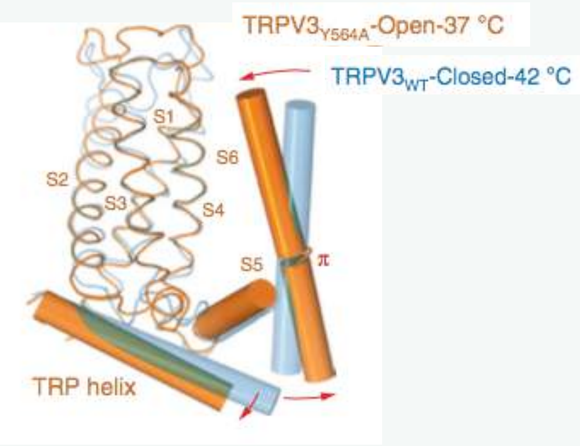




Fischer, Konevega, Wintermeyer, Rodnina, & Stark, *Nature* (2010)



Chen, Chang, Lin, Huang & Tsai *JACS* (2019).



At 37°C, the majority of the TRPV3$_{Y564A}$ particles were classified into low-resolution reconstructions, likely **representing an ensemble of heterogeneous conformations**.

Singh… Sobolevsky, *Nature Struc Mol Biol* (2019).

MD Simulations can give some insight…

nature COMMUNICATIONS

ARTICLE

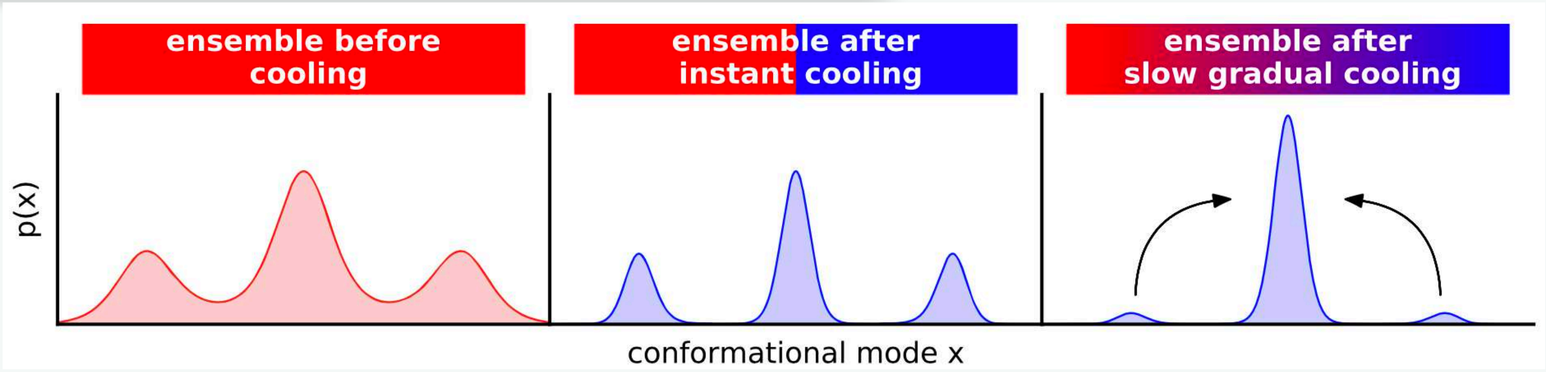
https://doi.org/10.1038/s41467-022-29332-2 OPEN

Check for updates

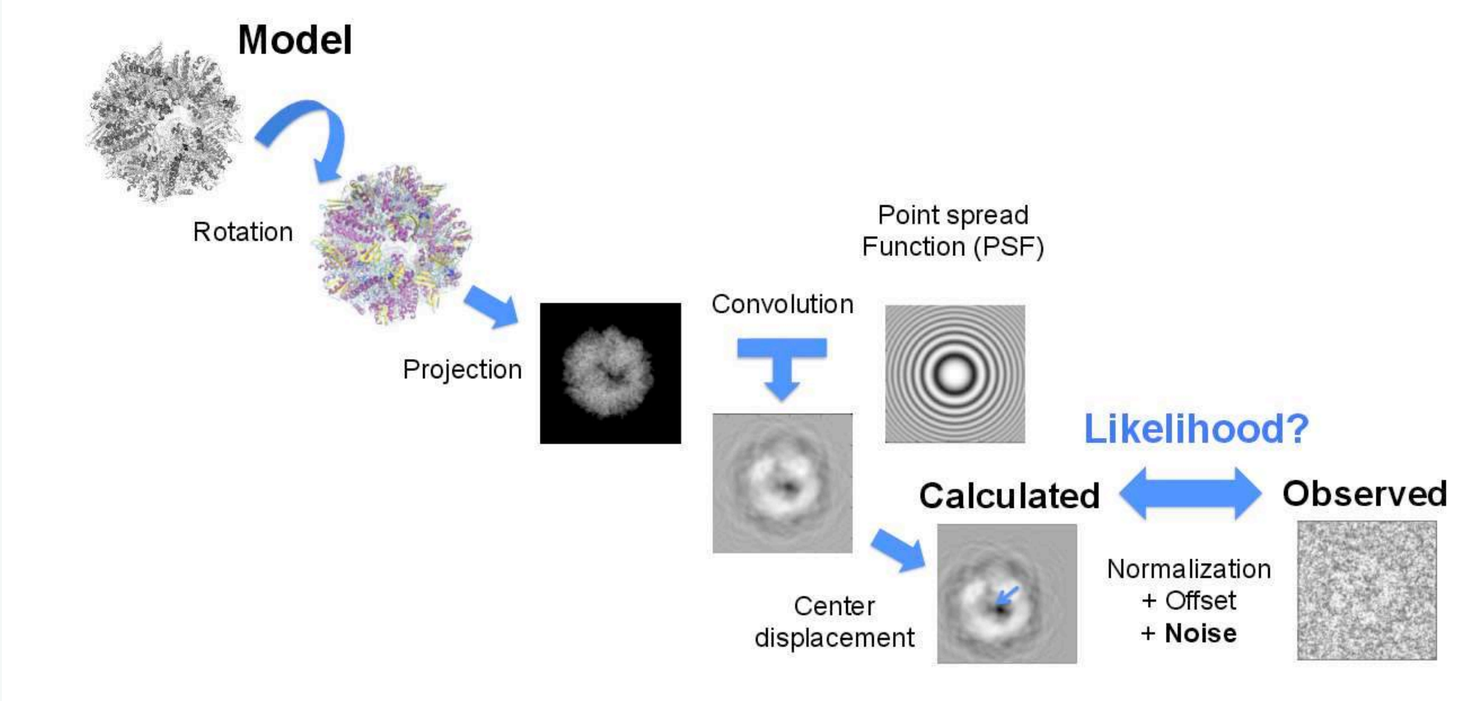
Effects of cryo-EM cooling on structural ensembles

Lars V. Bock [1]✉ & Helmut Grubmüller [1]

NATURE COMMUNICATIONS | (2022)13:1709 | https://doi.org/10.1038/s41467-022-29332-2 | www.nature.com/naturecommunications

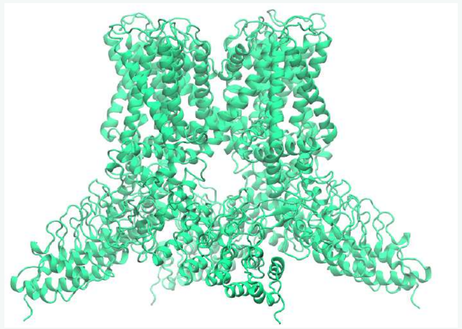


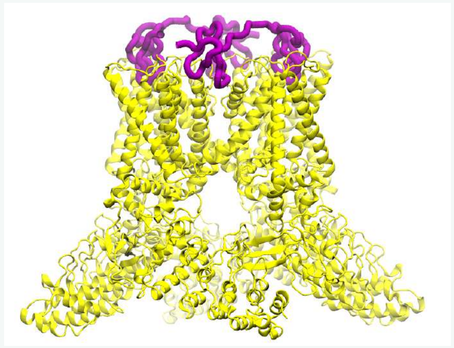
Bayesian Inference Methods For Comparing Single Particles to Atomic Models

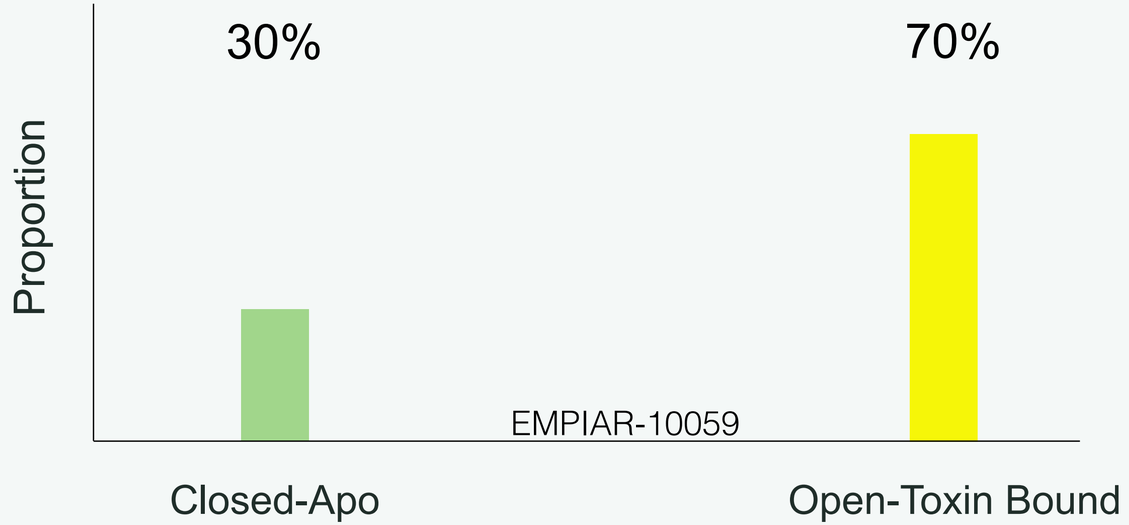


Pilar Cossio et al. Computer Physics Comunications (2017)

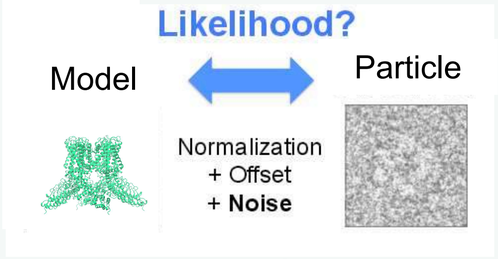
Using Bayesian Inference to Classify Particles into Closed and Open States.





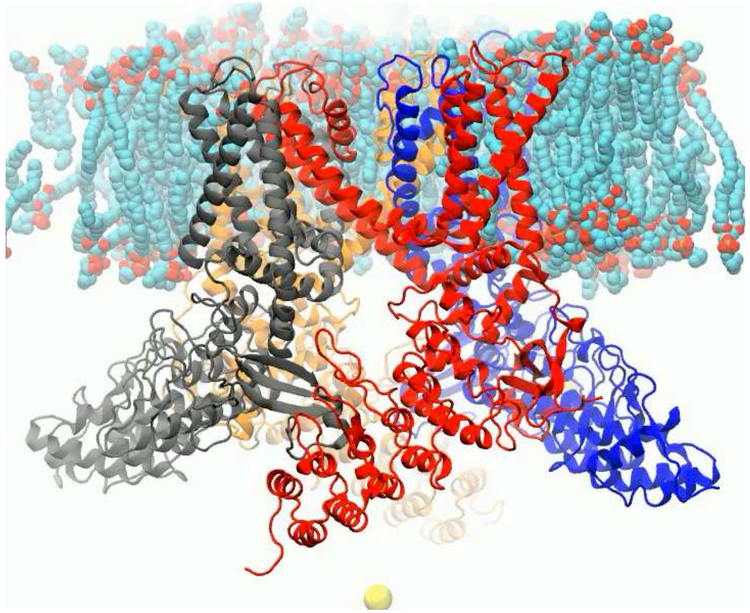


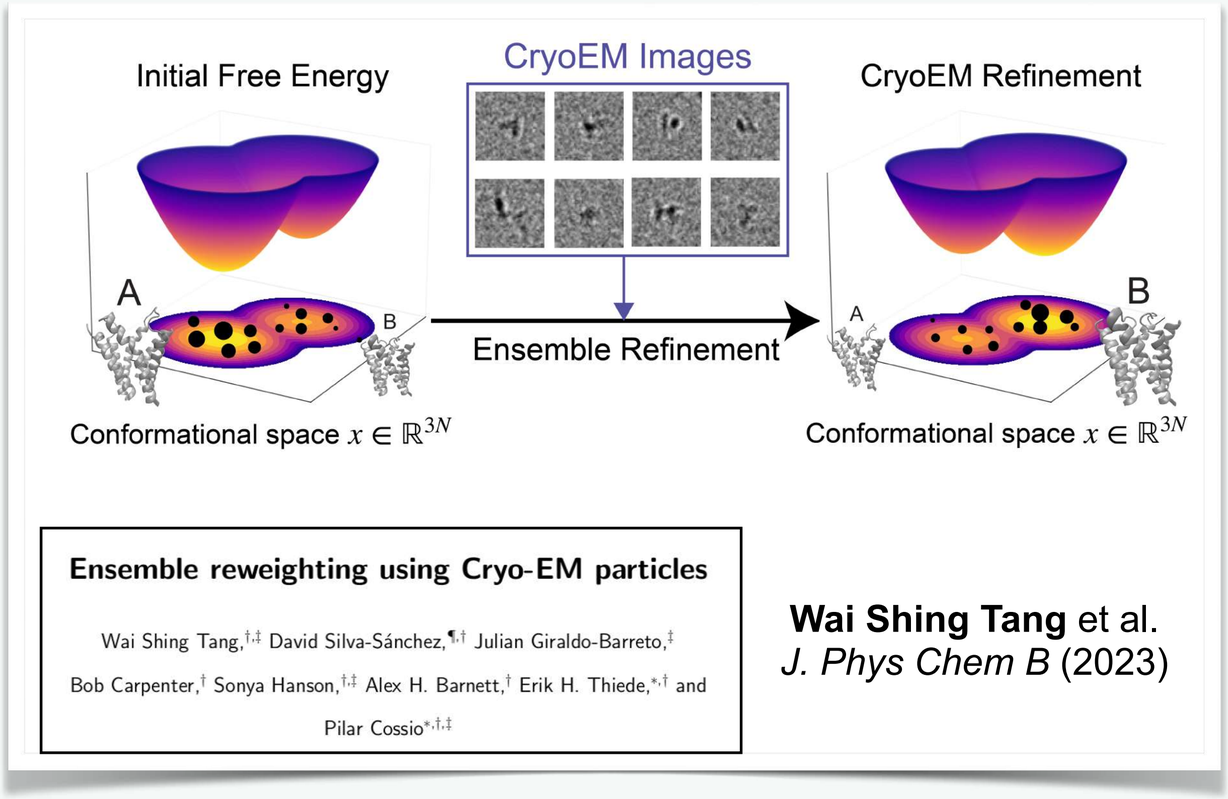
BioEM



Cossio et al. Computer Physics Transactions (2017)

The future: **Bridging the gap between cryo-EM data and long molecular dynamics simulations**





Ensemble reweighting using Cryo-EM particles

Wai Shing Tang,[†,‡] David Silva-Sánchez,[¶,†] Julian Giraldo-Barreto,[‡] Bob Carpenter,[†] Sonya Hanson,[†,‡] Alex H. Barnett,[†] Erik H. Thiede,[*,†] and Pilar Cossio[*,†,‡]

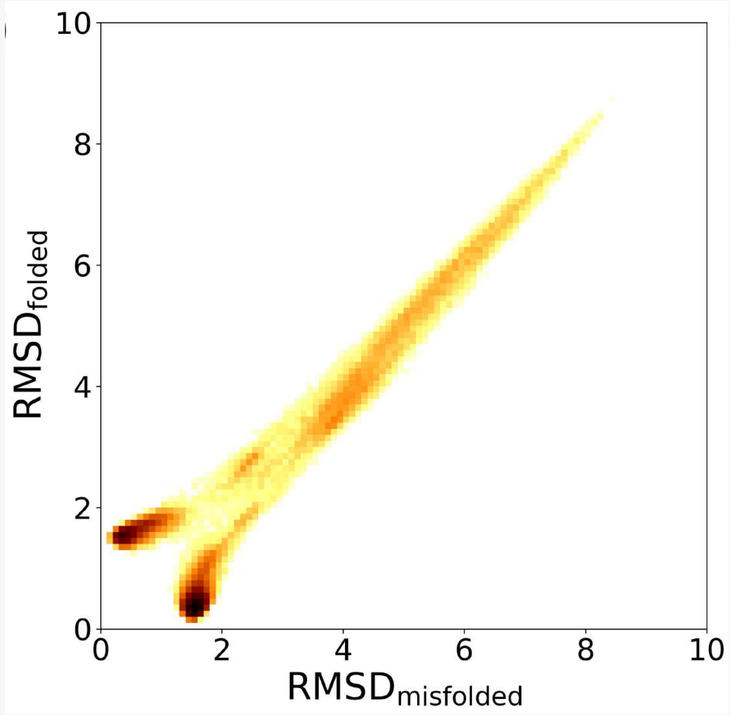
Wai Shing Tang et al.
J. Phys Chem B (2023)

Cryo-EM ensemble reweighting

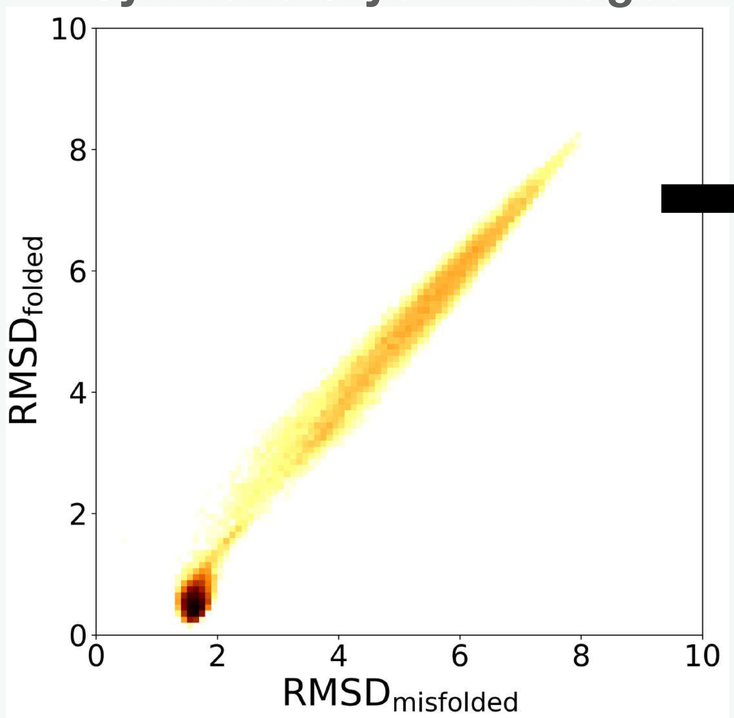

EH Thiede


WS Tang

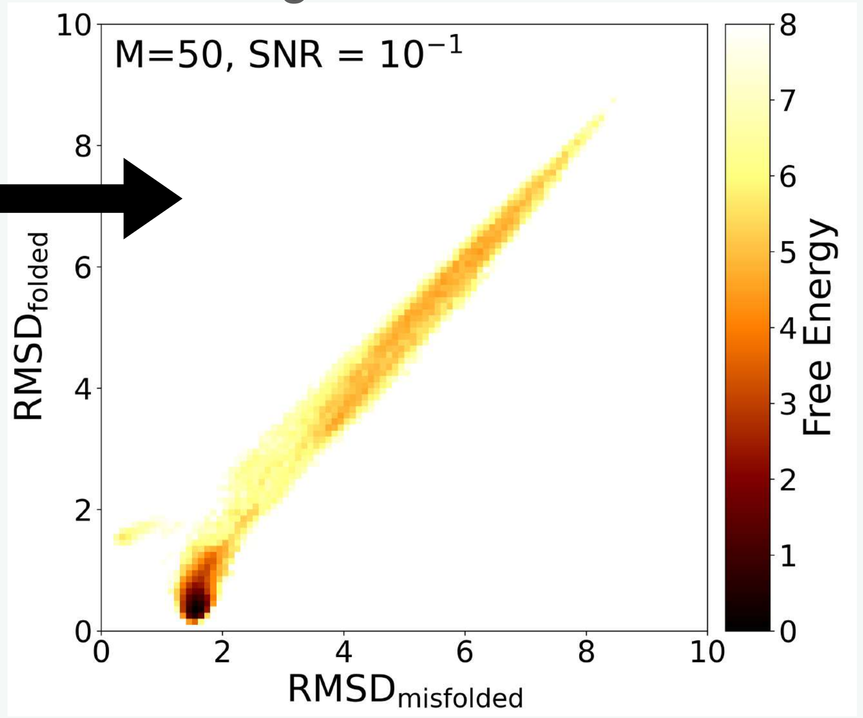
MD Structural Ensemble



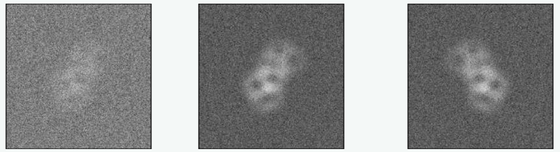
Synthetic Cryo-EM Images



Reweighted Ensemble



(our simulation)

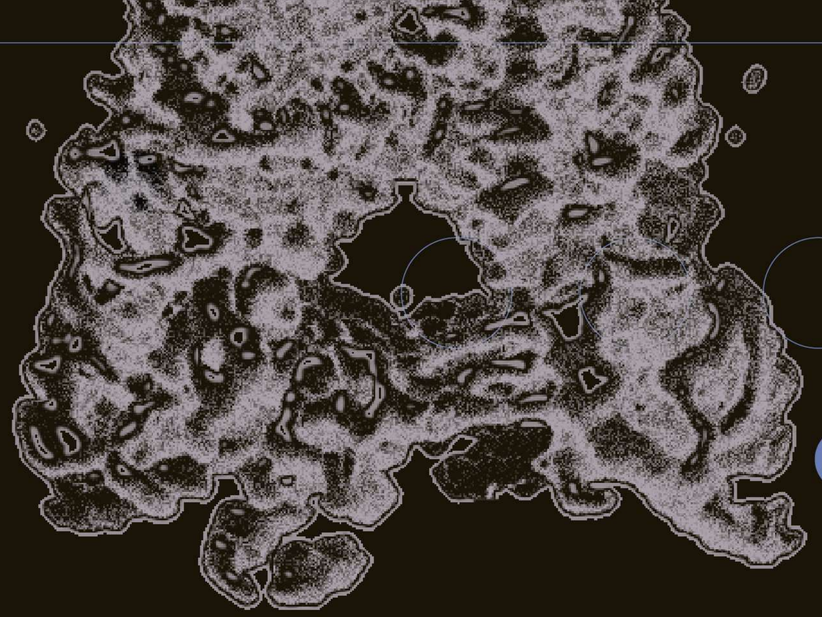


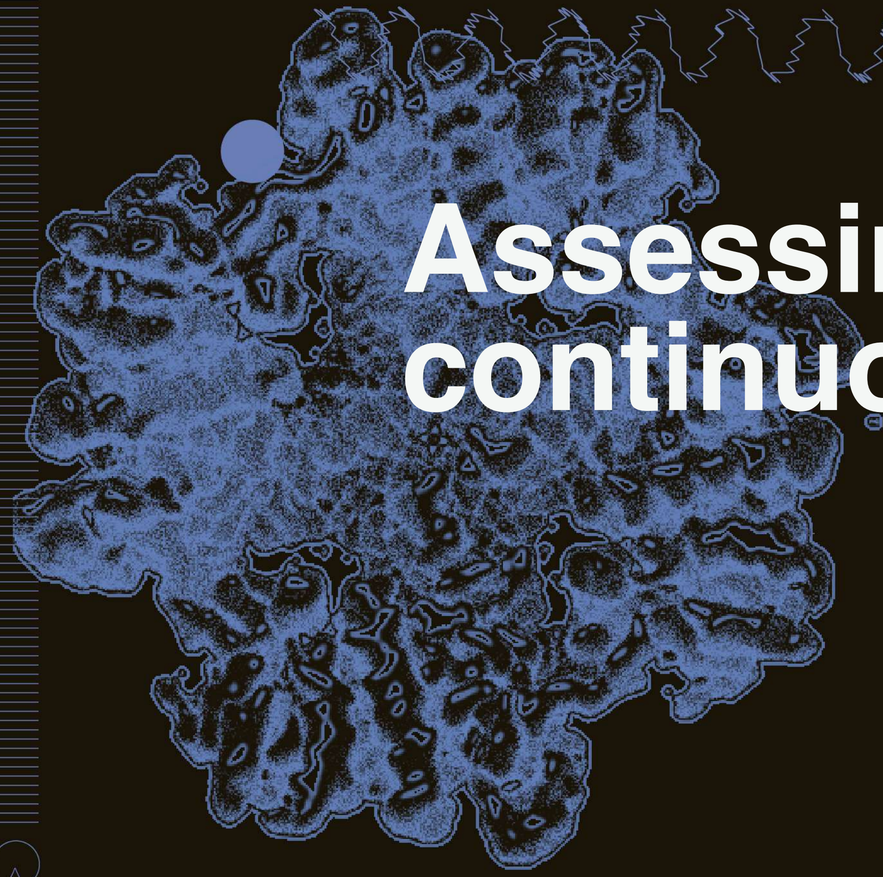
(DE Shaw data)

MD Simulation reweighted by cryo-EM images

WS Tang et al. *J. Phys. Chem. B* **2023.**

Sonya **Hanson**
shanson@flatironinstitute.org





Assessing the performance of continuous heterogeneity methods.

Are these motions real?

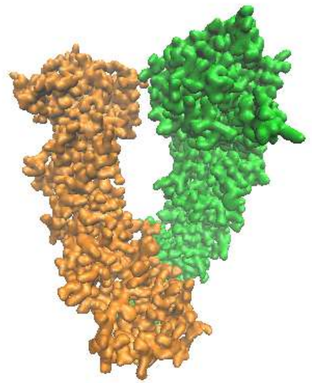
Recovery of Conformational Continuum From Single-Particle Cryo-EM Images: Optimization of ManifoldEM Informed by Ground Truth

Evan Seitz, Francisco Acosta-Reyes, Suvrajit Maji, Peter Schwander, *Member, IEEE*, and Joachim Frank

***Abstract*—This work is based on the manifold-embedding approach to study biological molecules exhibiting continuous conformational changes. Previous work established a method—now termed ManifoldEM—capable of reconstructing 3D movies and accompanying free-energy landscapes from single-particle cryo-EM images of macromolecules exercising multiple conformational degrees of freedom. While ManifoldEM has proven its viability in several experimental studies, critical limitations and uncertainties have been found throughout its extended development and use. Guided by insights from studies with cryo-EM ground-truth data, simulated from atomic structures undergoing conformational changes, we have built a novel framework, ESPER, able to retrieve**

I. Introduction

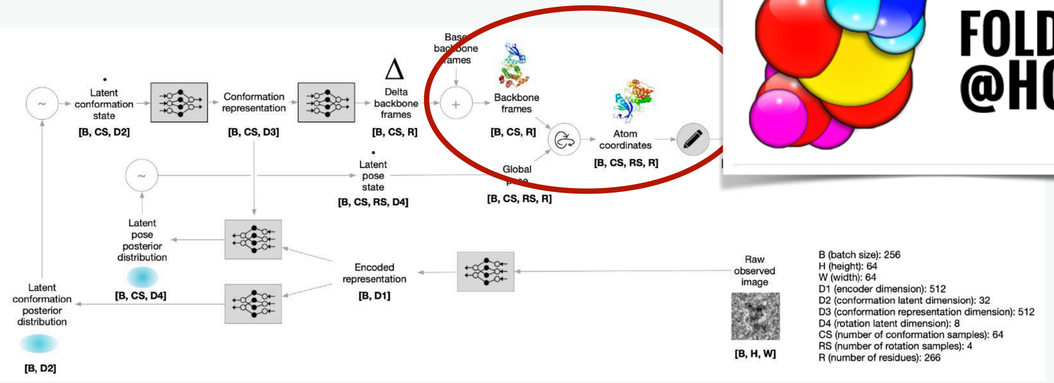
MOLECULAR machines—consisting of assemblies of proteins or nucleoproteins—take on a range of unique configurations or *conformational states* as they go through their functional cycles [1]. These states are typically characterized by different spatial constellations of relatively rigid domains, and can be organized in a *state space* according to the continuous motions of each domain along a unique coordinate. Specific sequences of the states in this space form pathways along which the molecular machine may transform. When the number of



Seitz et al *bioRxiv* (2019)
https://doi.org/10.1101/864116

An apoferritin for heterogeneity?

- Simulated datasets.



DeepMind

2021-06-25

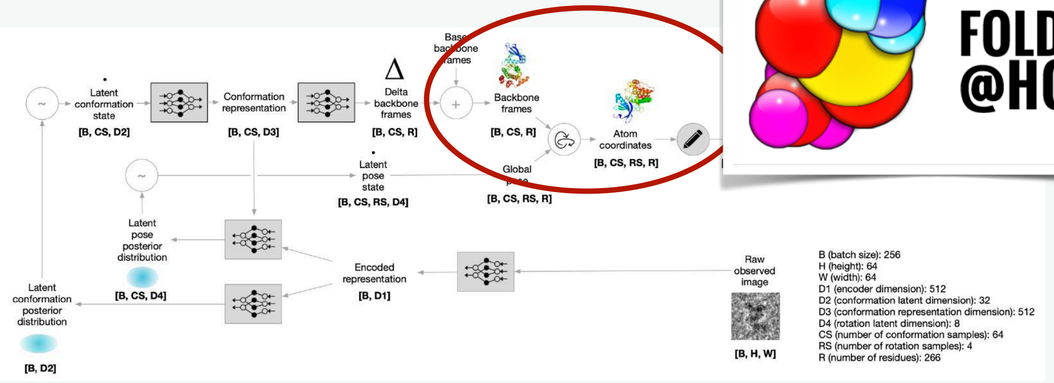
Inferring a Continuous Distribution of Atom Coordinates from Cryo-EM Images using VAEs

Dan Rosenbaum[*,1], **Marta Garnelo**[*,1], **Michal Zielinski**[*,1], **Charlie Beattie**[1], **Ellen Clancy**[1], **Andrea Huber**[1], **Pushmeet Kohli**[1], **Andrew W. Senior**[1], **John Jumper**[1], **Carl Doersch**[1], **S. M. Ali Eslami**[*,1], **Olaf Ronneberger**[*,1] **and Jonas Adler**[*,1]

[*]Equal contributions, [1]DeepMind

An apoferritin for heterogeneity?

- Simulated datasets.



DeepMind

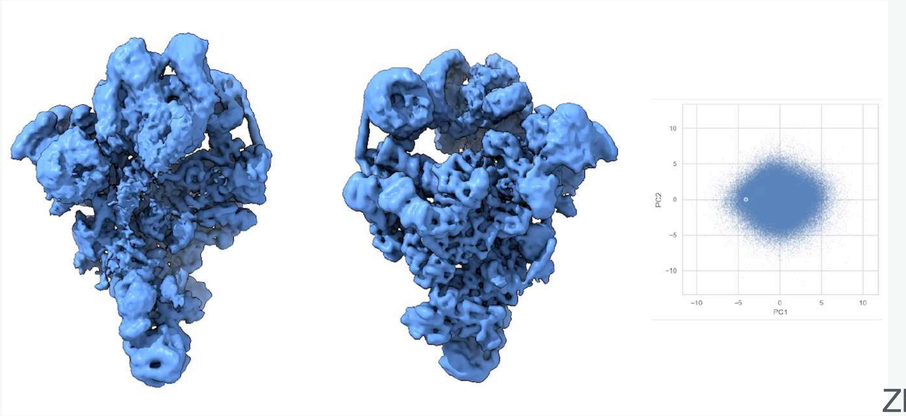
2021-06-25

Inferring a Continuous Distribution of Atom Coordinates from Cryo-EM Images using VAEs

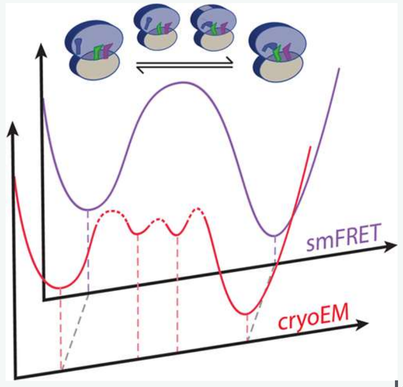
Dan Rosenbaum[*,1], **Marta Garnelo**[*,1], **Michal Zielinski**[*,1], **Charlie Beattie**[1], **Ellen Clancy**[1], **Andrea Huber**[1], **Pushmeet Kohli**[1], **Andrew W. Senior**[1], **John Jumper**[1], **Carl Doersch**[1], **S. M. Ali Eslami**[*,1], **Olaf Ronneberger**[*,1] **and Jonas Adler**[*,1]

[*]Equal contributions, [1]DeepMind

- Experimental datasets (w/ validation of the populations from another experimental method!)



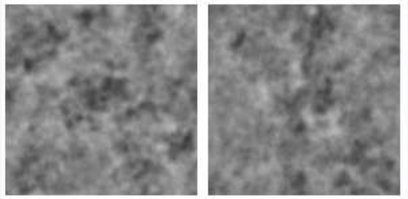
Zhong et al
Nature Methods (2021)



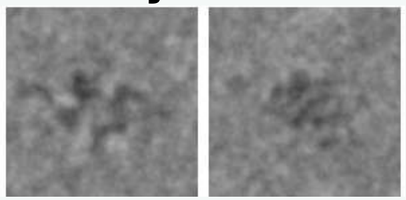
Kinz-Thompson et al
J Phys Chem B (2015)

Validating methods for heterogeneity in cryo-EM is still needed…

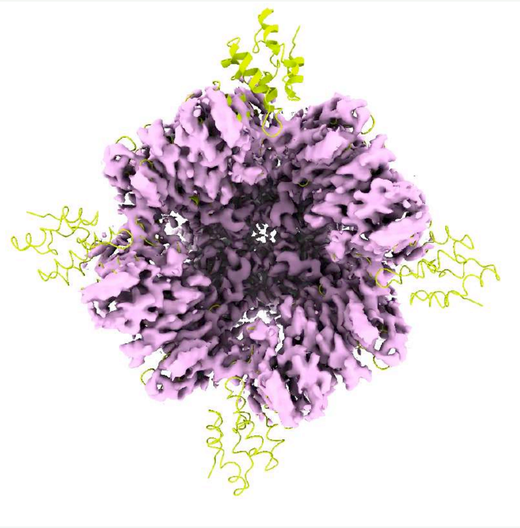
real



synthetic



There are a variety of these methods ranging from linear spectral methods to deep neural nets to reweighting of MD ensembles, and the each have different **strengths and weakness** that we are trying to tease out.



… especially on real experimental datasets.

Deep generative modeling for volume reconstruction in cryo-electron microscopy

Claire Donnat [a,1], Axel Levy [b,c], Frédéric Poitevin [c], Ellen D. Zhong [d], Nina Miolane [e,1]

(i) **Lack of benchmark datasets:** Current methods are developed and tested on a wide range of synthetic and experimental datasets that differ in the nature of the biomolecule being imaged, the dataset size, image size and associated resolution — with very little overlap across methods - see Table summarizing existing experiments in the supplementary material. There is unfortunately no MNIST (Deng, 2012) or Imagenet (Deng et al., 2009) for cryo-EM. Most methods resort to evaluating their performance on synthetic data, yet no cryo-EM simulator acts as a standard to generate simulated images in a unified way. Synthetic datasets vary in the realism of the image formation model used for simulation, *e.g.* in the noise model, the signal-to-noise ratio or the distribution of nuisance variables (e.g. poses). Subsequent ex-

Donnat …Miolane, *J. Struct. Biol.* (2022).

2023: A blind heterogeneity challenge for cryo-EM hosted by the Flatiron Institute.

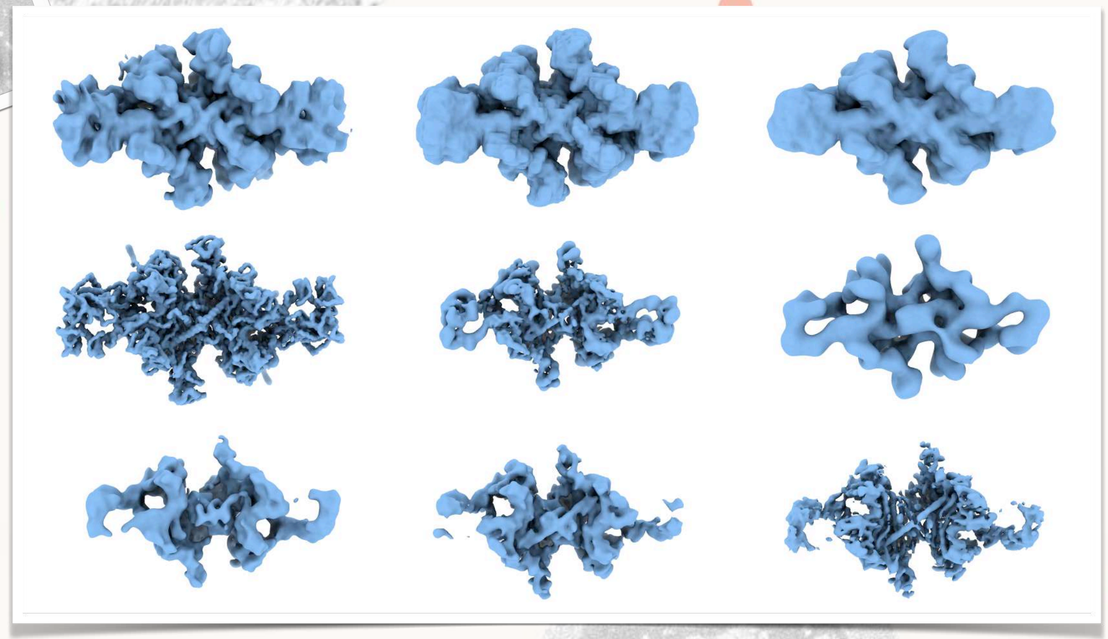
Average

Participate by analyzing the two heterogeneous datasets provided and uploading your results **between September 1 and October 31, 2023. More info at:**

(Note that this is a challenge to explore how to compare methods in the first place and no winner will be declared. We welcome feedback/questions/comments!)

Two Datasets:
1- real from NYSBC
2- simulated via MD, so we have ground truth states and their distribution

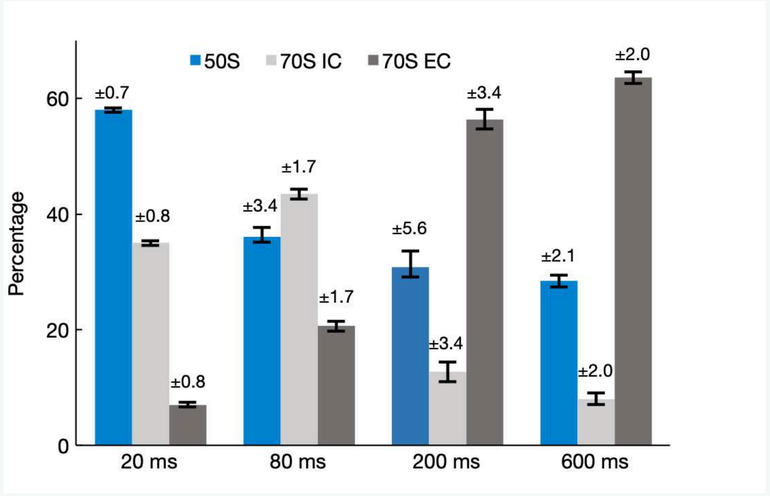
Average



https://www.simonsfoundation.org/heterogeneity-in-cryo-electron-microscopy/

One potentially valuable aspect of continuous heterogeneity methods is getting populations, which can be translated into Free Energy landscapes.

- Quantifying errors in populations of states - also relevant to discrete heterogeneity.



Kaledhonkar et al, *Nature* (2019).

Sonya **Hanson**
shanson@flatironinstitute.org

DISCUSSION TOPICS

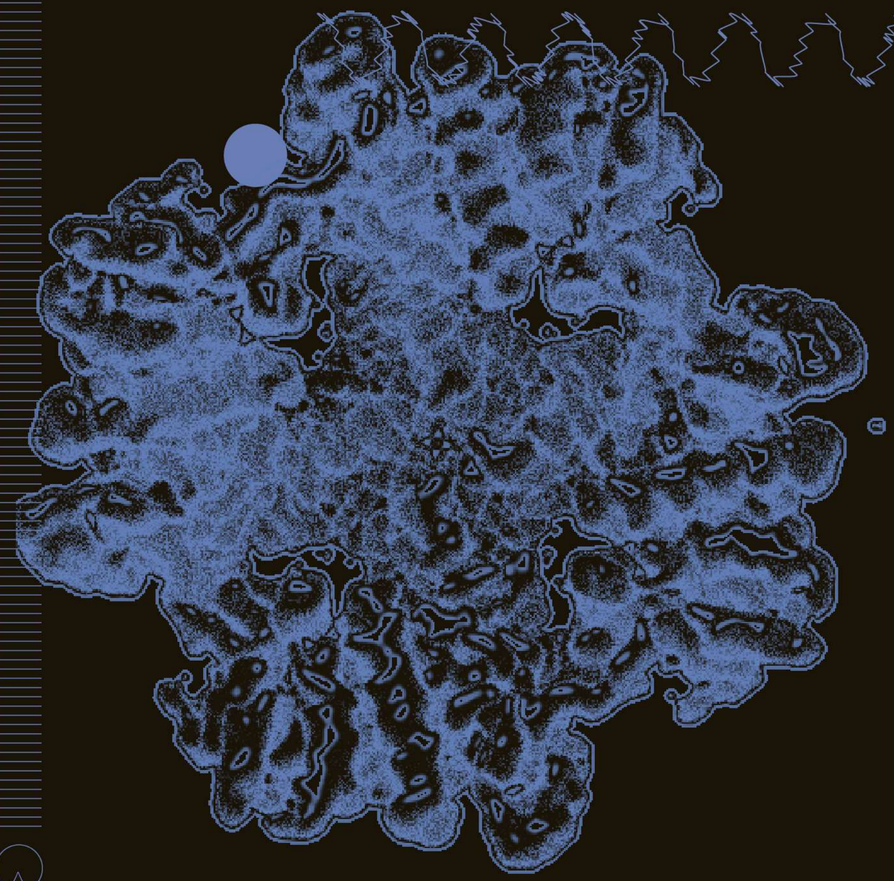
How should we benchmark methods in continuous heterogeneity for cryo-EM?

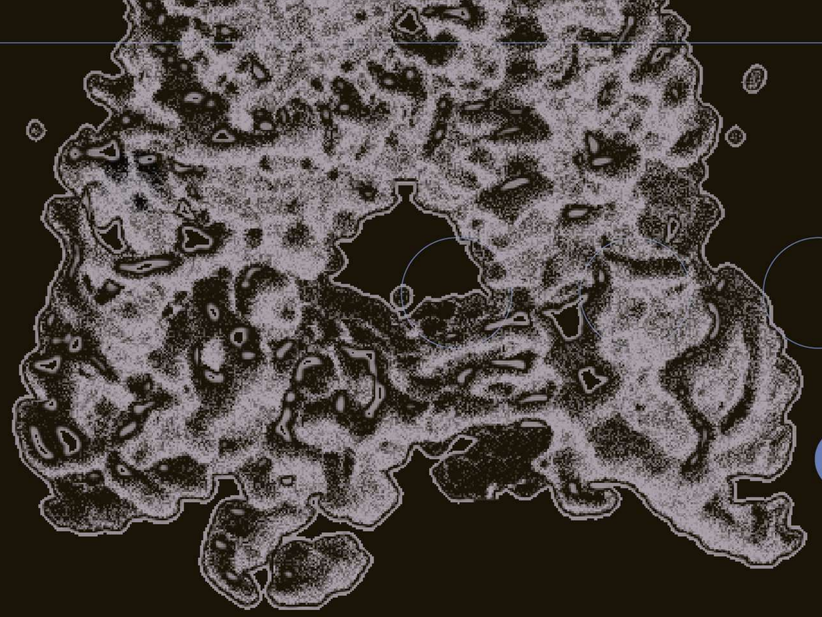
How much should we prioritize getting populations?

How much do we think the vitrification process affects our ensembles?

et cetera

Sonya **Hanson**
shanson@flatironinstitute.org





Toward continuous heterogeneity in *in situ* cryo-EM and cryo-ET.