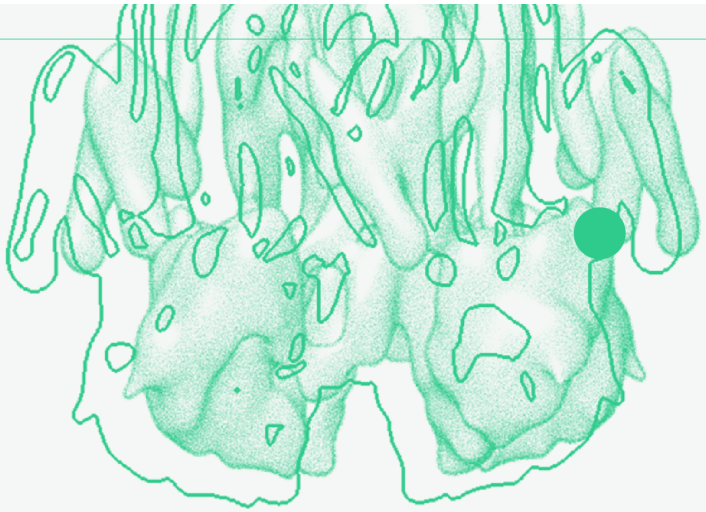


Ensemble reweighting using cryo-EM particles

Pilar Cossio
Center for Computational Mathematics
Flatiron Institute, NY

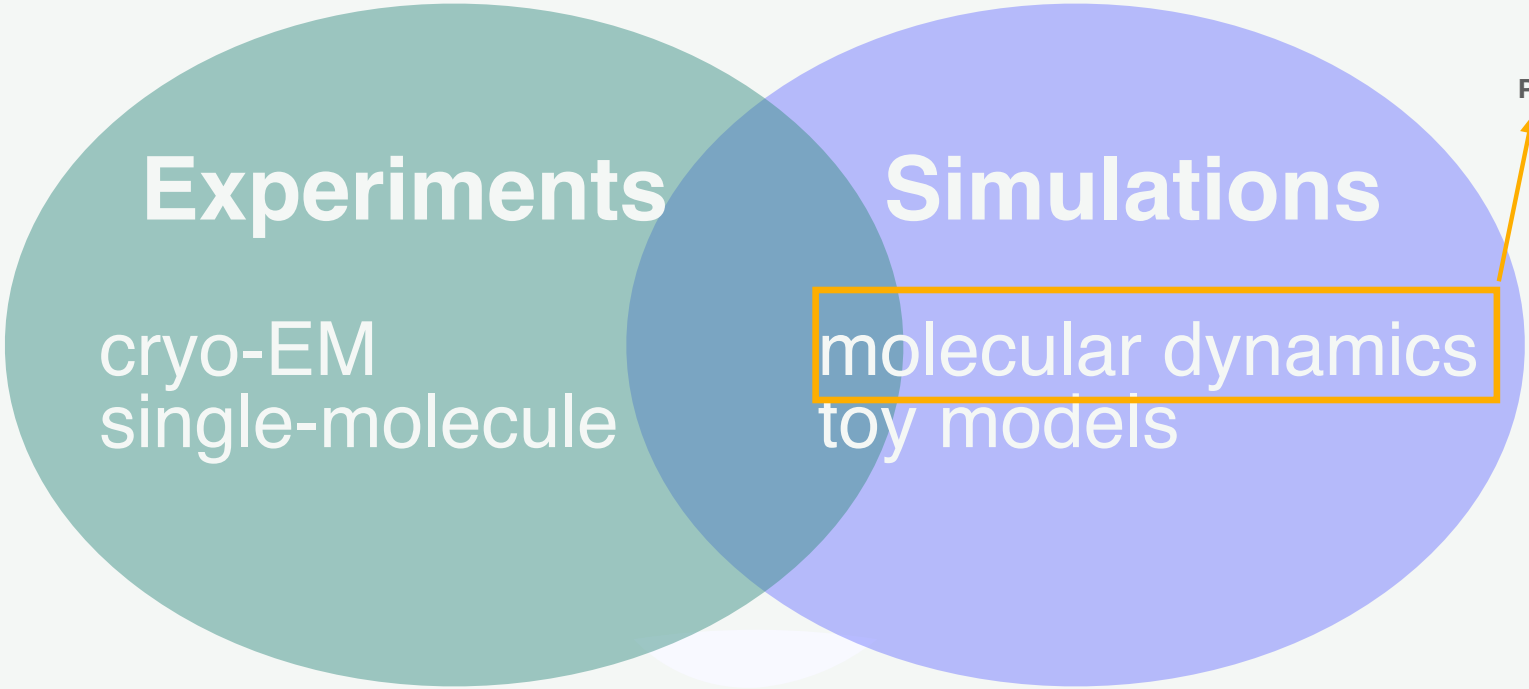
2023

FLATIRON INSTITUTE

Outline:

1. Motivation
2. Cryo-EM as a single-molecule technique
3. Free-energies profiles along a path
4. Generalization: Ensemble Reweighting
5. Conclusions and perspectives: biasing MD…

Structural and Molecular Biophysics* at the Flatiron Institute, NY.



Check out:
Palacio-Rodriguez et al. JPCL, 2022
CV efficiency and rates from biased simulations

Theory, simulation, and application to extract **detailed molecular mechanisms: free energies and dynamics.**

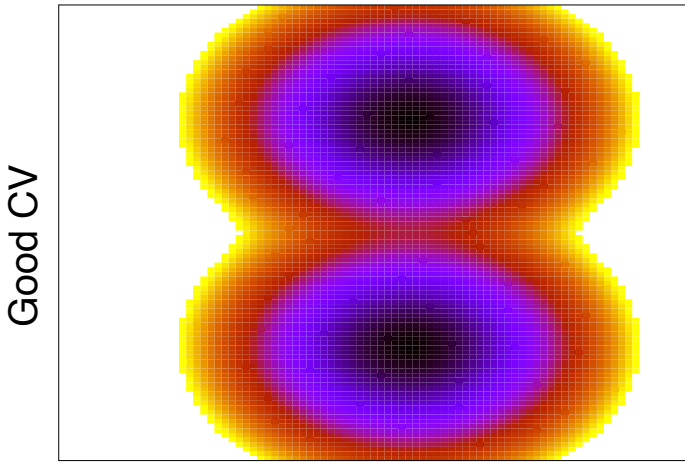
* Co-leadership with Sonya Hanson

* Builds on infrequent metadynamics. Tiwary & Parinello, PRL 2013.

Transition Rates and Efficiency of Collective Variables from Time-Dependent Biased Simulations

Karen Palacio-Rodriguez,† Hadrien Vroylandt,† Lukas S. Stelzl, Fabio Pietrucci, Gerhard Hummer, and Pilar Cossio*

Cite This: J. Phys. Chem. Lett. 2022, 13, 7490–7496 Read Online



We introduce a measure of the ***efficiency of CV*** **(γ)**:

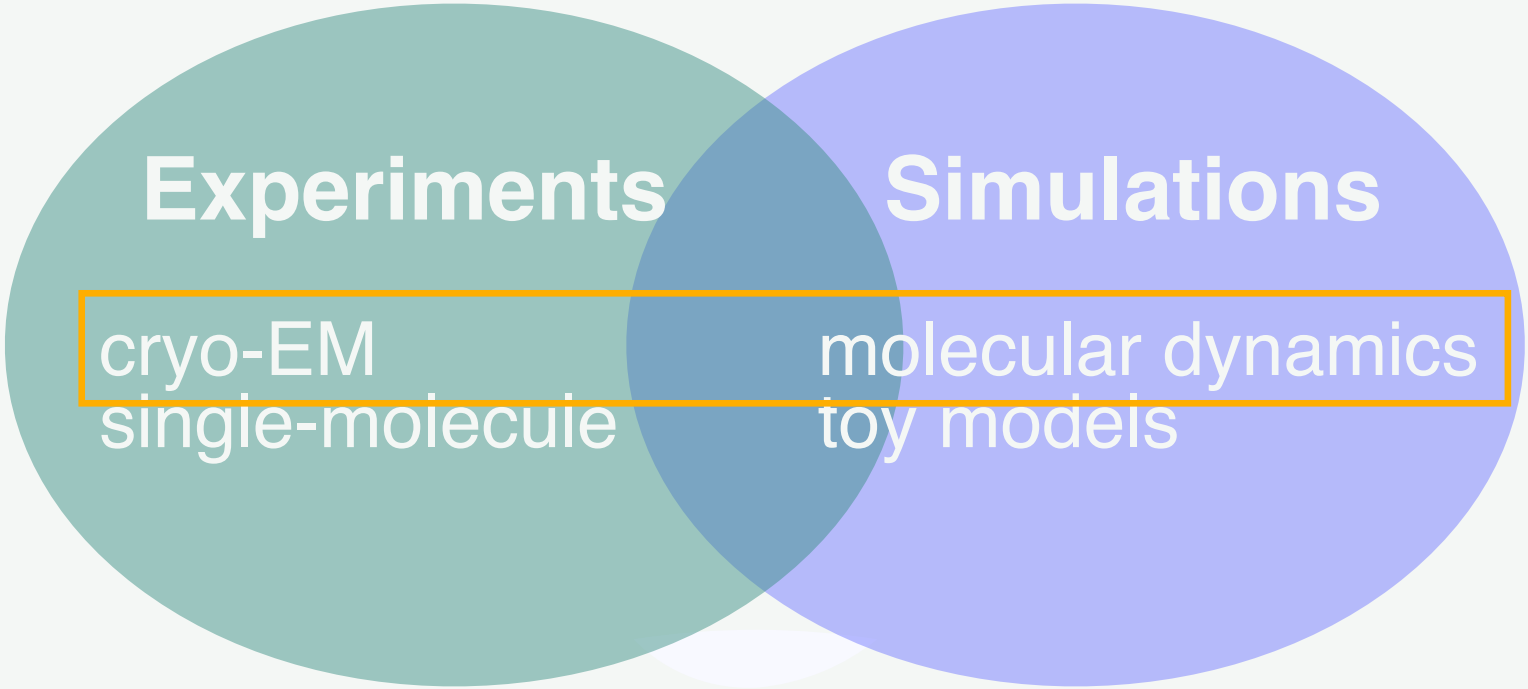
$$k(t) = k_{\mathrm{pre}} e^{-\beta \Delta G^{\ddagger} + \beta \gamma V_{\mathrm{MB}}(t)} = k_0 \, e^{\beta \gamma V_{\mathrm{MB}}(t)}$$

Scales the bias

Fitting parameters:
γ in [0,1] & γ =1 —> good CV

We can extract the unbiased rate (for any CV) and have a new measure of the efficiency of the CV.

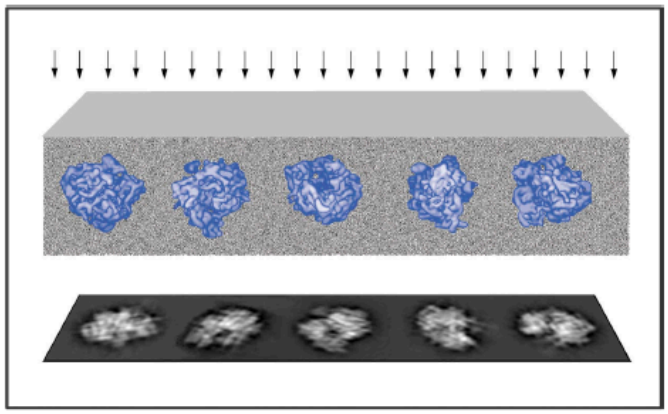
Structural and Molecular Biophysics* at the Flatiron Institute, NY.



Theory, simulation, and application to extract **detailed molecular mechanisms: free energies and dynamics.**

* Co-leadership with Sonya Hanson

Cryo-EM images of biomolecules have two main unknowns: **pose and conformation**

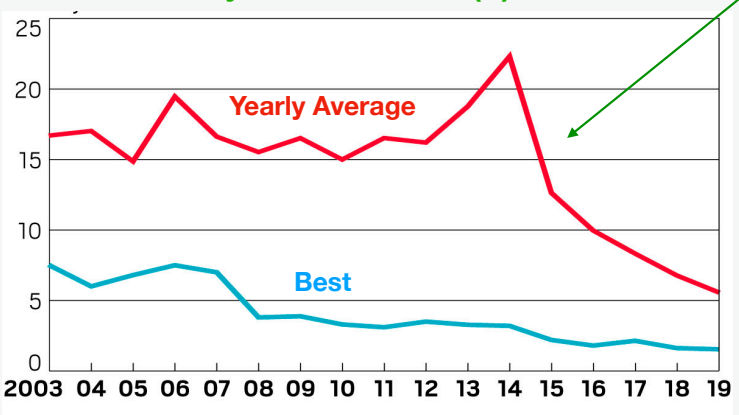


Projection direction?

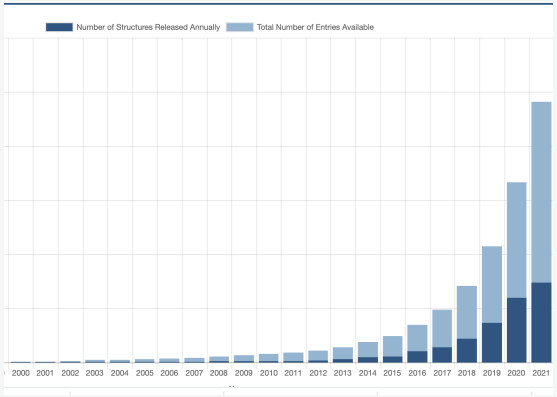
~~Conformation?~~

The CryoEM resolution revolution

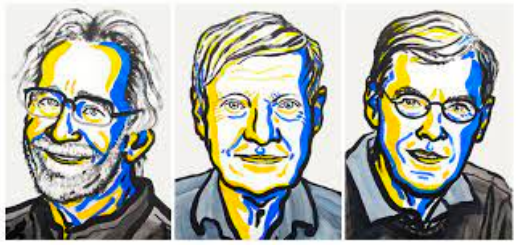
CryoEM resolution (A)



CryoEM structures in the PDB



Nobel Prize in Chemistry

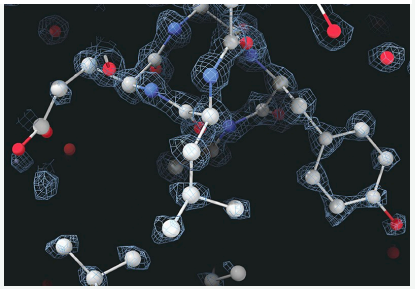


Dubochet, Frank & Henderson

NEWS AND VIEWS | 21 October 2020

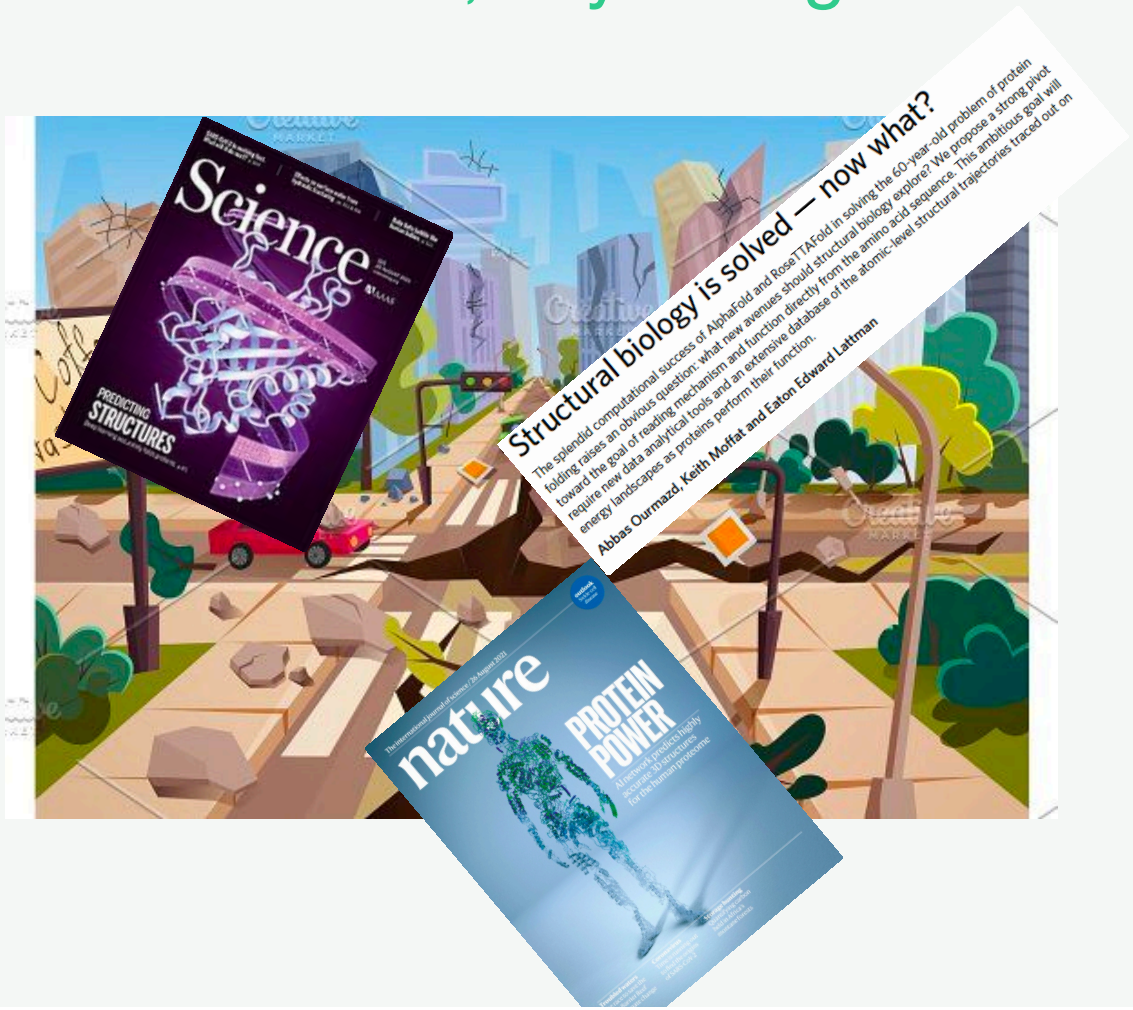
Cryo-electron microscopy reaches atomic resolution

2020
~1.2A Resolution!



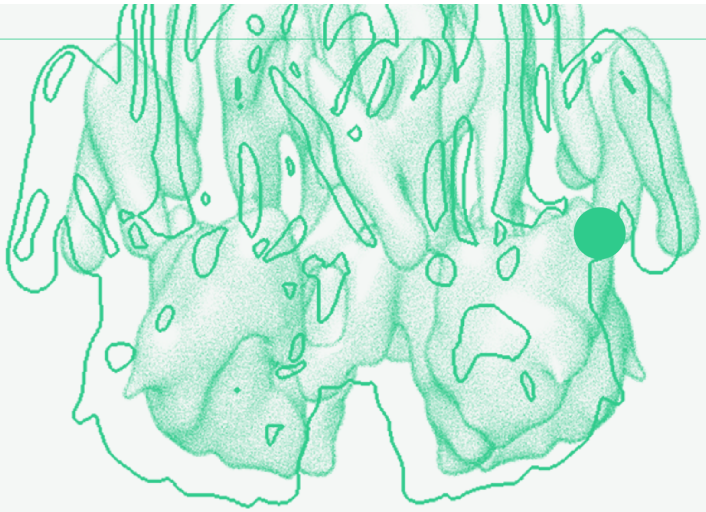
2023

However, CryoEM grounds have been shaken:



Where is the field going?

- Faster, better and cheaper structure determination.
- Conformational variability, free energies and environment
- *In situ*

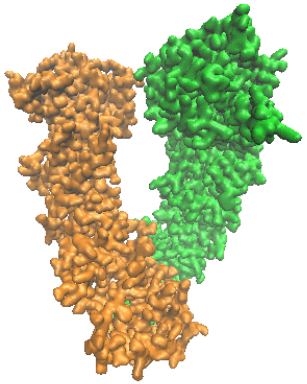


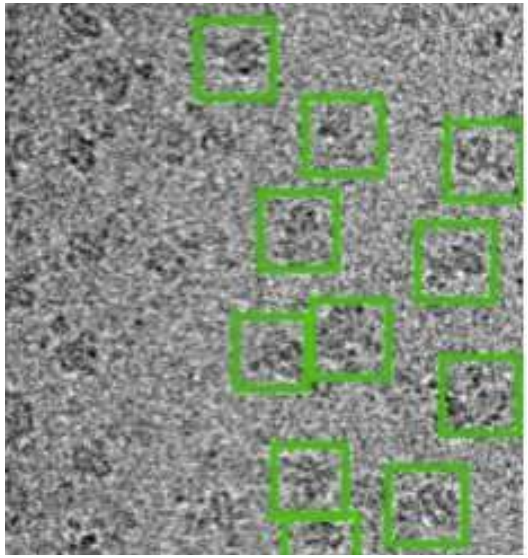
CryoEM: conformational variability, free energies and environment

Conformational heterogeneity from single-particle cryo-EM

Freezing is done very rapidly*.

It's possible to trap individual conformations!!





More information than just an average.

* Bock & Grubmuller, Nat Commun. 2022

Dealing with cryoEM heterogeneity

A typical case:

EMPIAR-10278

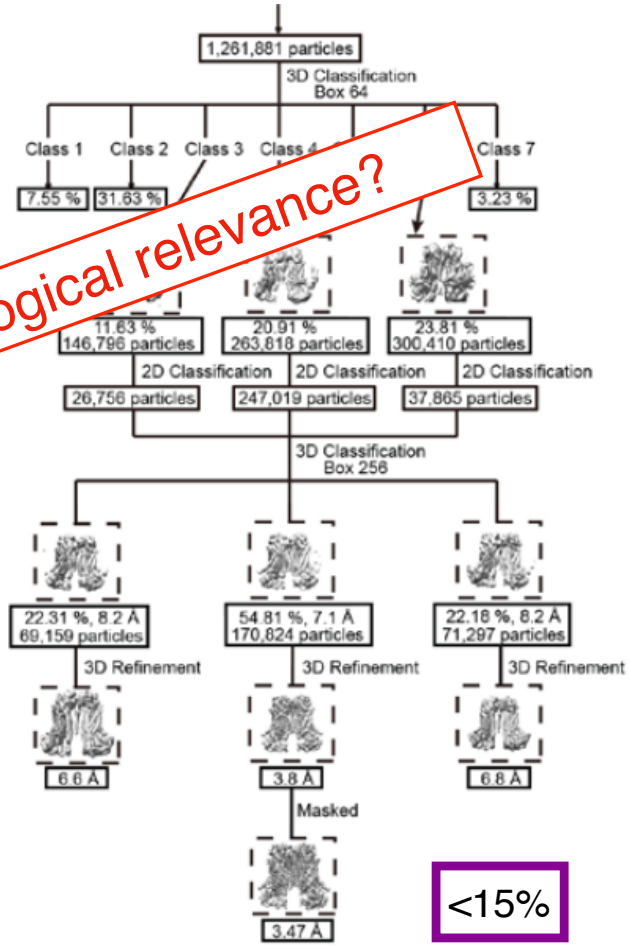
Cryo-EM structure of TMEM16F in digitonin with calcium bound

Publication: Cryo-EM Studies of TMEM16F Calcium-Activated Ion Channel Suggest Features Important for Lipid Scrambling
Feng S, Han TW, Ye W, Jin P, Cheng T, Li J, Jan LY, Cheng Y
Cell Rep **28** 567-579.e4 (2019)

Contains: picked particles

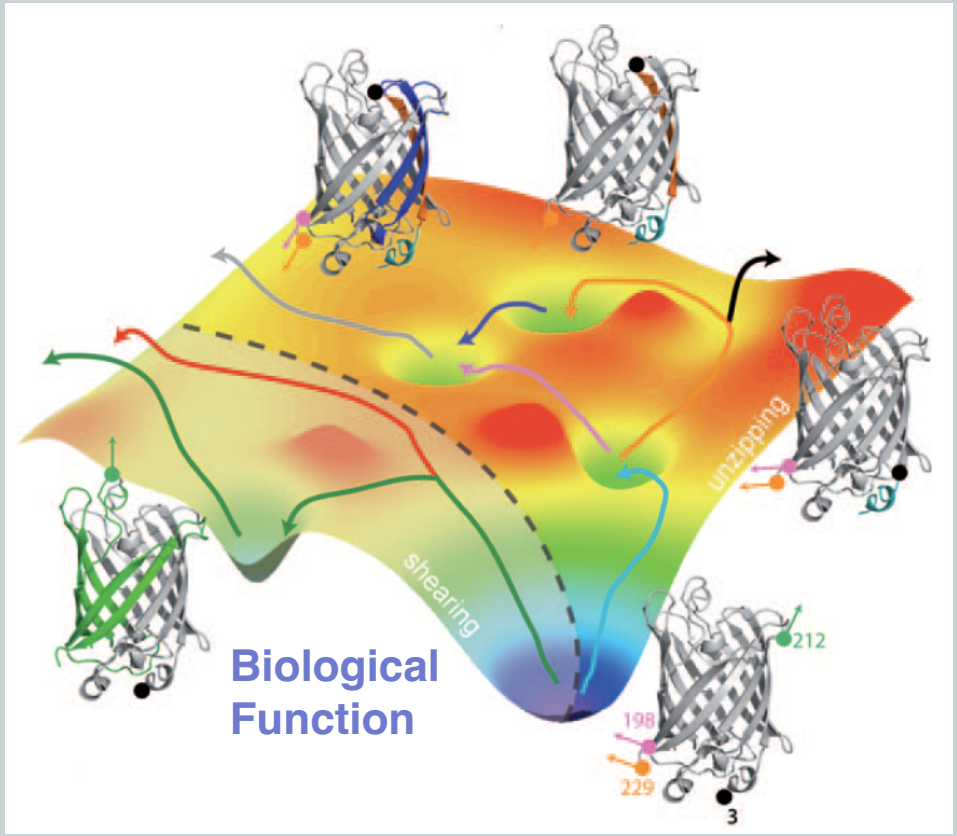
3D classificati…one to filter-out …les and reconstruct high-resolution density maps

Why discard so much data? Biological relevance?



Extracting free-energies from cryo-EM (?)

Related to the probability distribution



Bertz et al. Angew. Chem. 2008, 47, 8192 –8195

The free energy (FE) is related to the probability distribution of the configurations.

A configuration $x \in \mathbb{R}^{3N}$ and the Boltzmann factor gives the probability

$$\rho(x) = \frac{1}{Z} e^{-\beta H(x)}$$

Hamiltonian

Too complex! Project along some collective variables (s):

$$\rho(s) = \int \delta(S(x) - s)\rho(x)dx = \frac{1}{Z_1} e^{-\beta G(s)}$$

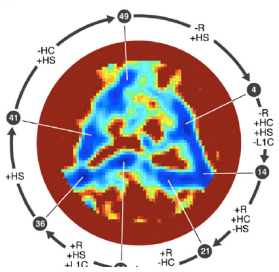
Free energy

Free-energy surfaces from cryo-EM

Machine learning & diffusion maps

Trajectories of the ribosome as a Brownian nanomachine (2014)

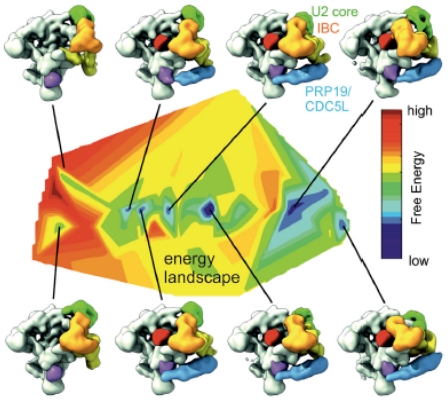
Ali Dashti[a,1], Peter Schwander[a,1], Robert Langlois[b], Russell Fung[a], Wen Li[b], Ahmad Hosseinizadeh[a], Hstau Y. Liao[b], Jesper Pallesen[c,2], Gyanesh Sharma[b,3], Vera A. Stupina[d], Anne E. Simon[d], Jonathan D. Dinman[d], Joachim Frank[b,c,4], and Abbas Ourmazd[a,1,4]



(2020) 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✉]

Principal component analysis (PCA)



(2021) Energy Landscape of the SARS-CoV-2 Reveals Extensive Conformational Heterogeneity

Ghoncheh Mashayekhi,[1#] John Vant,[2#] Abhishek Singharoy[2*], Abbas Ourmazd[1*]

Structure and Conformational Dynamics of the Human Spliceosomal B^{act} Complex (2018)

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

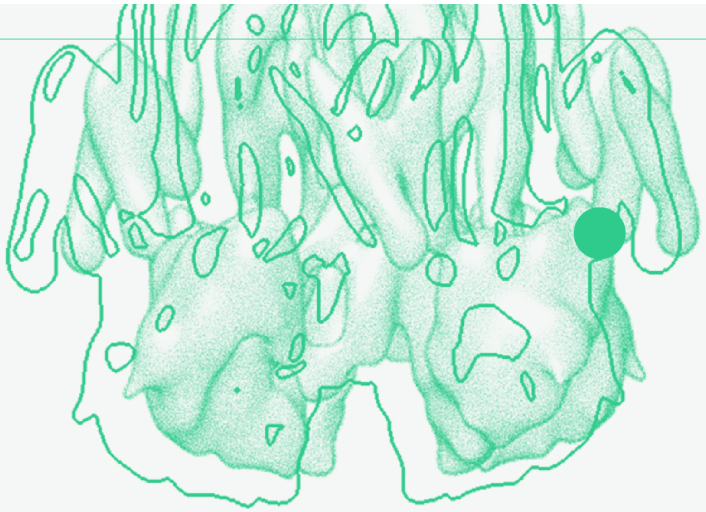
(2020)

Energy landscape of domain motion in glutamate dehydrogenase deduced from cryo-electron microscopy

Mao Oide[1,2], Takayuki Kato[3], Tomotaka Oroguchi[1,2] and Masayoshi Nakasako[1,2]

Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy

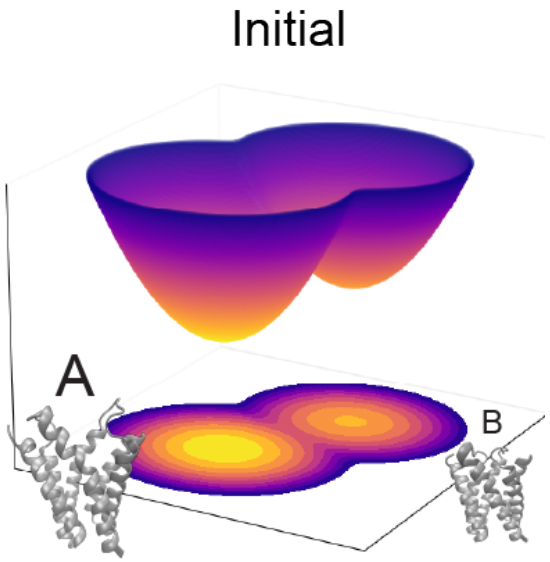
Niels Fischer[1], Andrey L. Konevega[2,3], Wolfgang Wintermeyer[2], Marina V. Rodnina[2] & Holger Stark[1]



General framework

Free energies from CryoEM

Inference of free energies given cryoEM particles.



Conformational space $x \in \mathbb{R}^{3N}$

Ensemble of models (e.g. AlphaFold, MD trajectory)

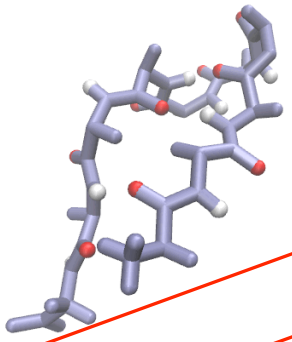
Theory

data

CryoEM images Y={y_i}

Conformation

$x \in \mathbb{R}^{3N}$



distribution

$$p(x|H) = \frac{1}{Z} e^{-\beta H(x)}$$

Hamiltonian

Bayes' theorem

$$p(H|Y) \propto p(H)p(Y|H)$$

we would like to know

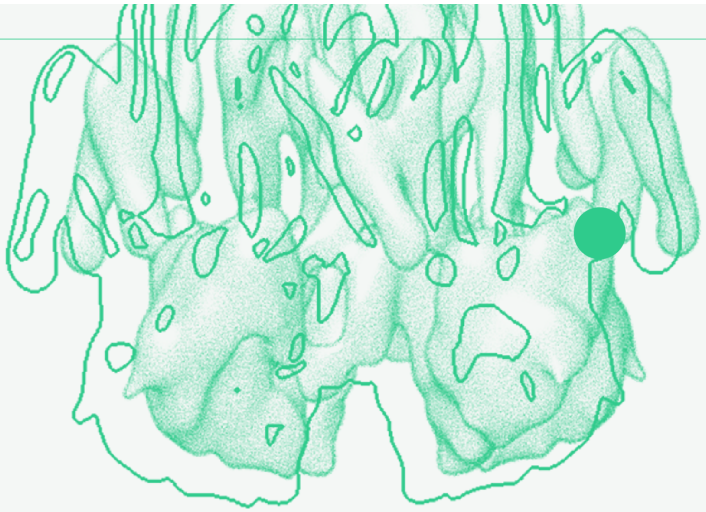
Too complex!

a set of *iid* cryo-EM particles
(not an averaged observable)

$$p(H|Y) \propto p(H) \prod_i \int p(y_i|x_i)p(x_i|H)dx_i$$

Likelihood of image y_i generated by conformation x_i*

Probability of x_i



Free-energy profile along a path

Cryo-BIFE

Cryo-EM Bayesian Inference of Free Energy

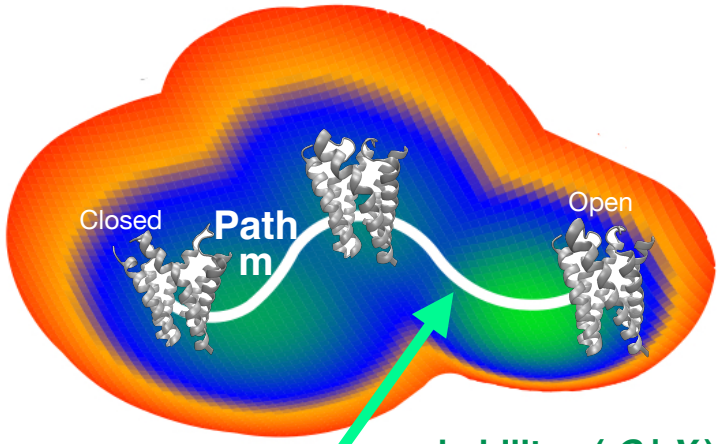
Article | Open Access | Published: 01 July 2021

A Bayesian approach to extracting free-energy profiles from cryo-electron microscopy experiments

Julian Giraldo-Barreto, Sebastian Ortiz, Erik H. Thiede, Karen Palacio-Rodriguez, Bob Carpenter, Alex H. Barnett & Pilar Cossio

We assume a low-temperature approximation and project along a path collective variable*

$$p(x|G) = \sum_m \delta(x - x_m) \frac{e^{-\beta G(s_m)}}{Z_1}$$



probability p(*G* | *Y*) ?

Cryo-EM particles Y

*Inspired by MD methods like the string method (Weinan, Ren, Vanden-Eijnden - Physical Review B, 2002)

Giraldo Barreto et al. Sci Rep 2021

Cryo-BIFE

For multiple images Y, the posterior reduces to

Single image

$$p(G|Y) \propto p(G) \prod_i \left[\sum_m p(y_i|x_m) \frac{e^{-\beta G(s_m)}}{Z_1} \right]$$

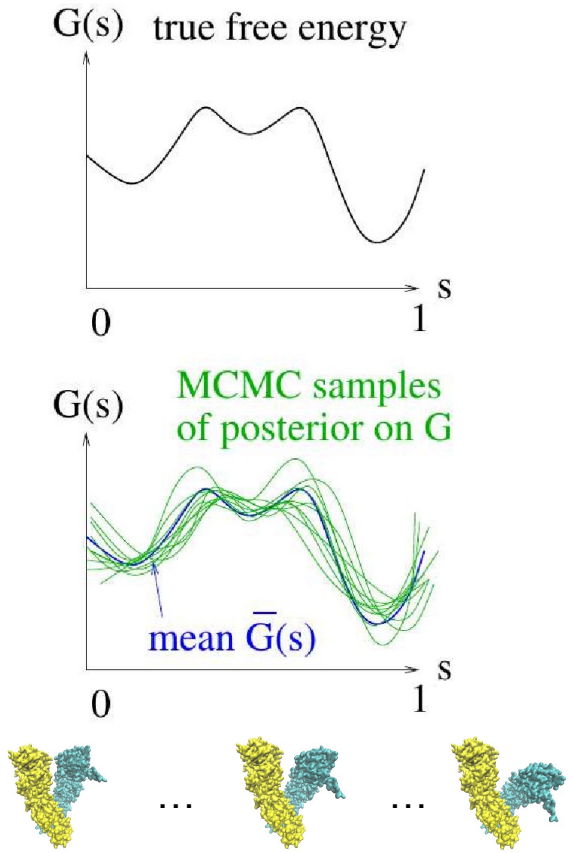
Prior

Probability of a *G* given the particles

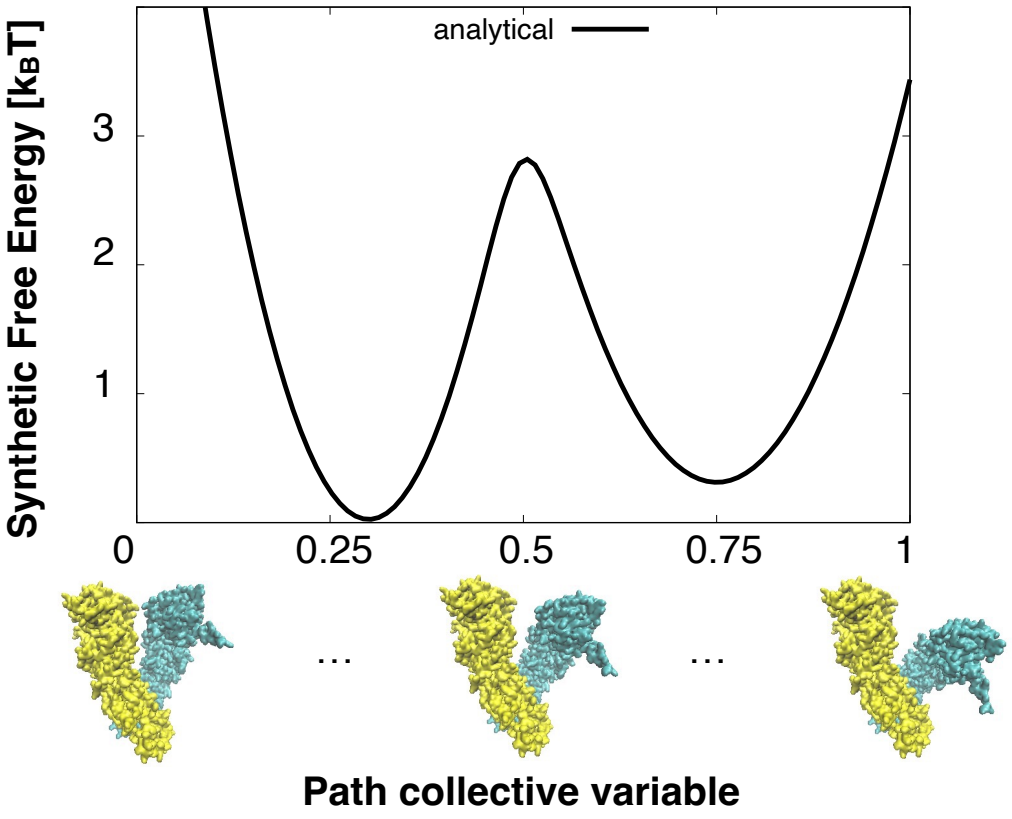
Main point:
Sampling *G* from p(*G* | Y), allows us to calculate its mean and **uncertainty** at each node given the cryo-EM particles.

Key ideas:

- Particle reconstructions are **not** crucial.
- We assume that we have a sufficiently good transition path of 3D structures (by MD, Alphafold etc).
- We use MCMC sampling (e.g. in STAN) to extract the expected G and its uncertainty.



1D Hsp90 synthetic data



Path calculation:

- 20 nodes (along the single degree of freedom)
- 2 orientation-rounds of BioEM

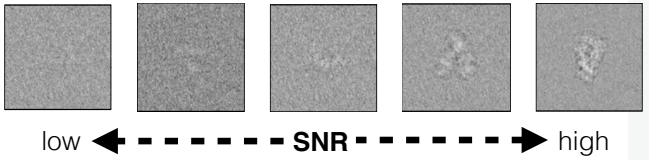
Synthetic images:

~13300 particles

- 128x128 pixels
- pixel size 2.2 A
- Coarse-grained residues

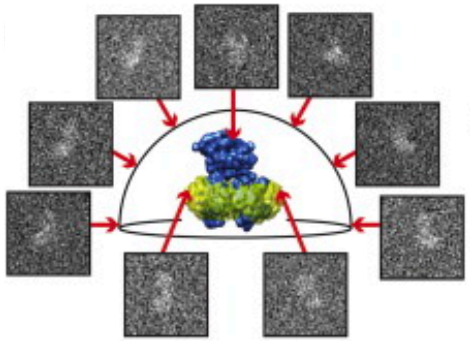
With uniformly distributed:

- random orientations
- random defocus [0.5,3] micro-m
- random signal-to-noise ratio (SNR) [0.001, 0.1]

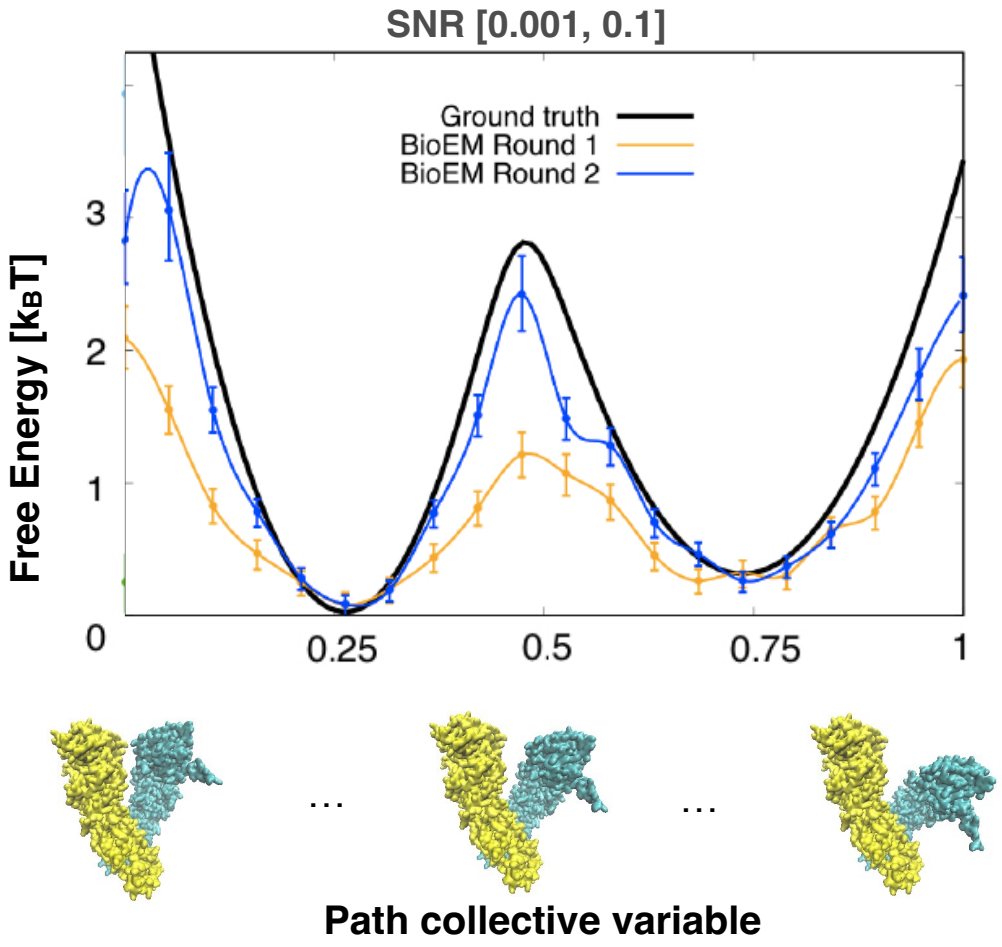


Projection direction accuracy is important

The projection direction is unknown.



Nogales & Scheres, Mol Cell 2015

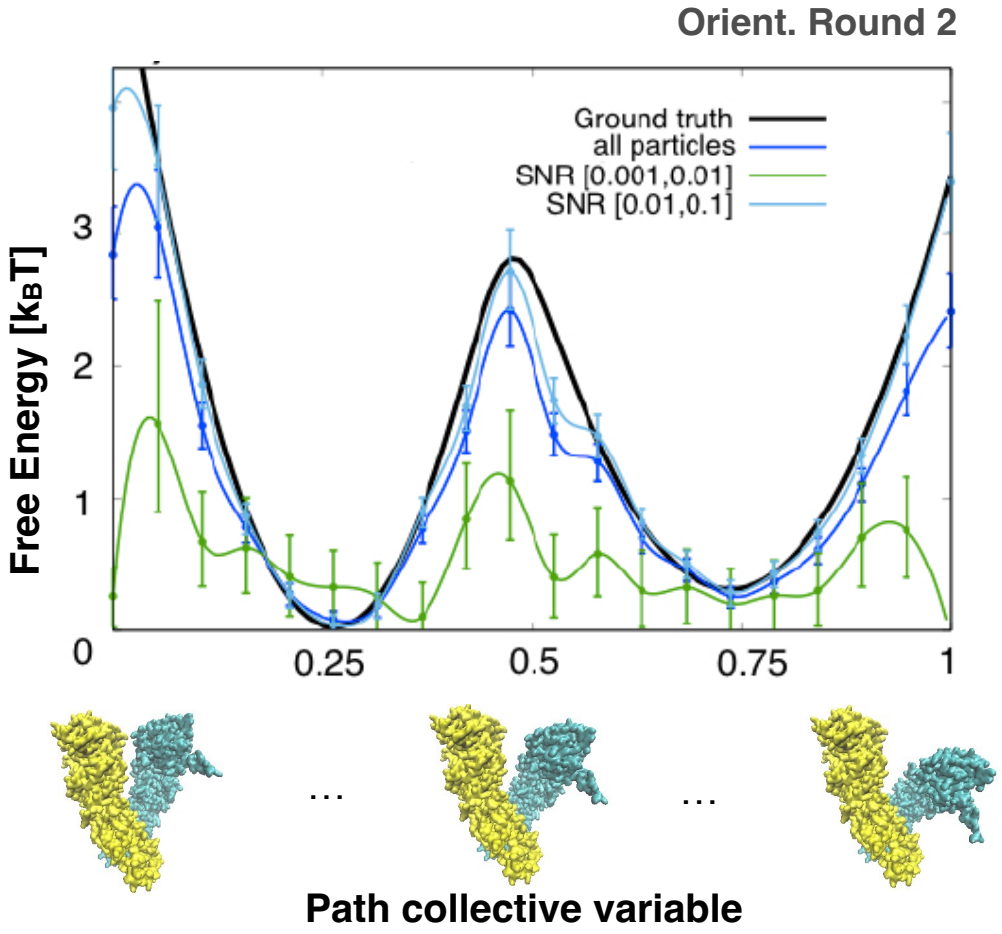


Signal-to-noise ratio (SNR)

A poor FE profile is recovered from the low SNR group

High SNR provides a slightly better profile than using all particles.

Adding “bad” particles does not hinder the FE recovery (if there are some “good” ones in the group).

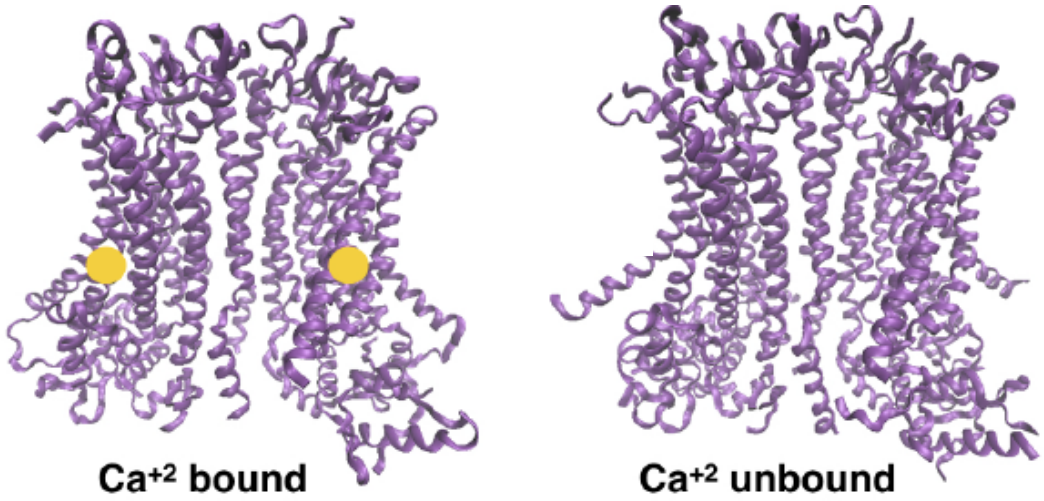


Real data:TMEM16F a calcium-activated ion channel

Cryo-EM Studies of TMEM16F Calcium-Activated Ion Channel Suggest Features Important for Lipid Scrambling (2019)

Shengjie Feng,[1,4] Shangyu Dang,[2,4] Tina Wei Han,[1] Wenlei Ye,[1] Peng Jin,[1] Tong Cheng,[1] Junrui Li,[2] Yuh Nung Jan,[1,2,3] Lily Yeh Jan,[1,2,3,*] and Yifan Cheng[2,3,5,*]
[1]Department of Physiology, University of California, San Francisco, San Francisco, CA 94158, USA
[2]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA
[3]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94158, USA

Generated from two different data sets: with and without Ca^{+2}



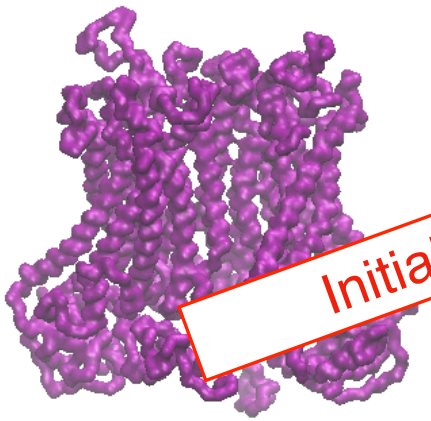
Ca^{+2} bound

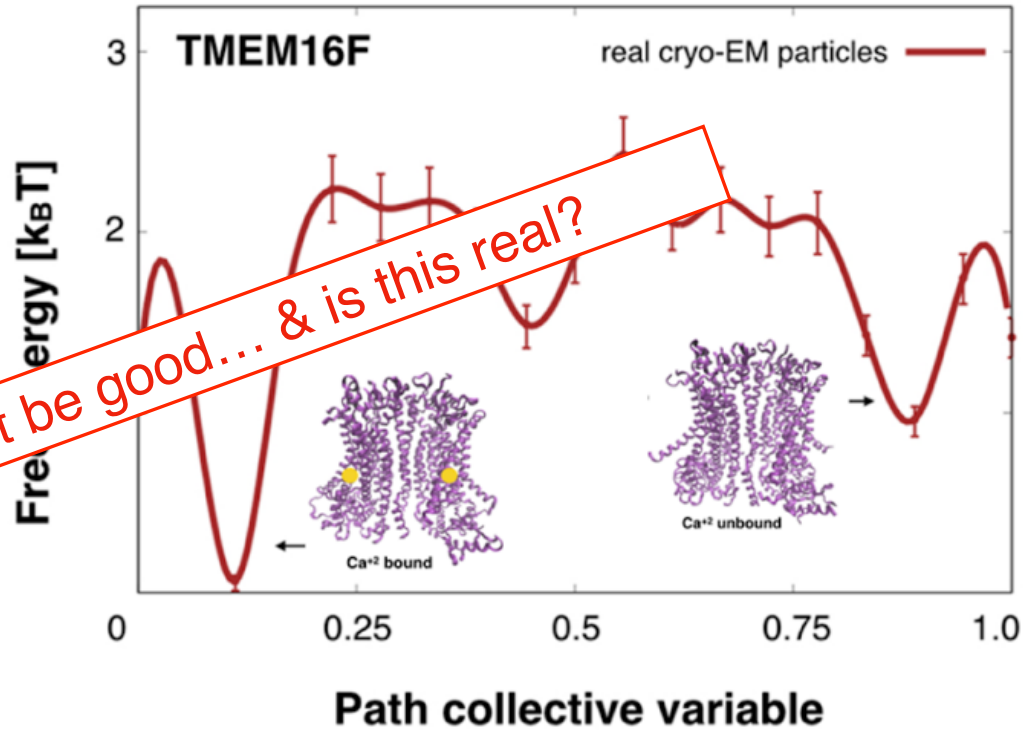
Ca^{+2} unbound

All particles available: EMPIAR 10278

Is there a population of the Ca^{+2}-unbound state in the Ca^{+2}-bound set?

Real data:TMEM16F a calcium-activated ion channel





Initial path might not be good… & is this real?

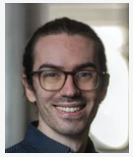
Particles from the entire set:

- EMPIAR 10278
- 15000 images
- 256x256 pixels
- pixel size 1 A

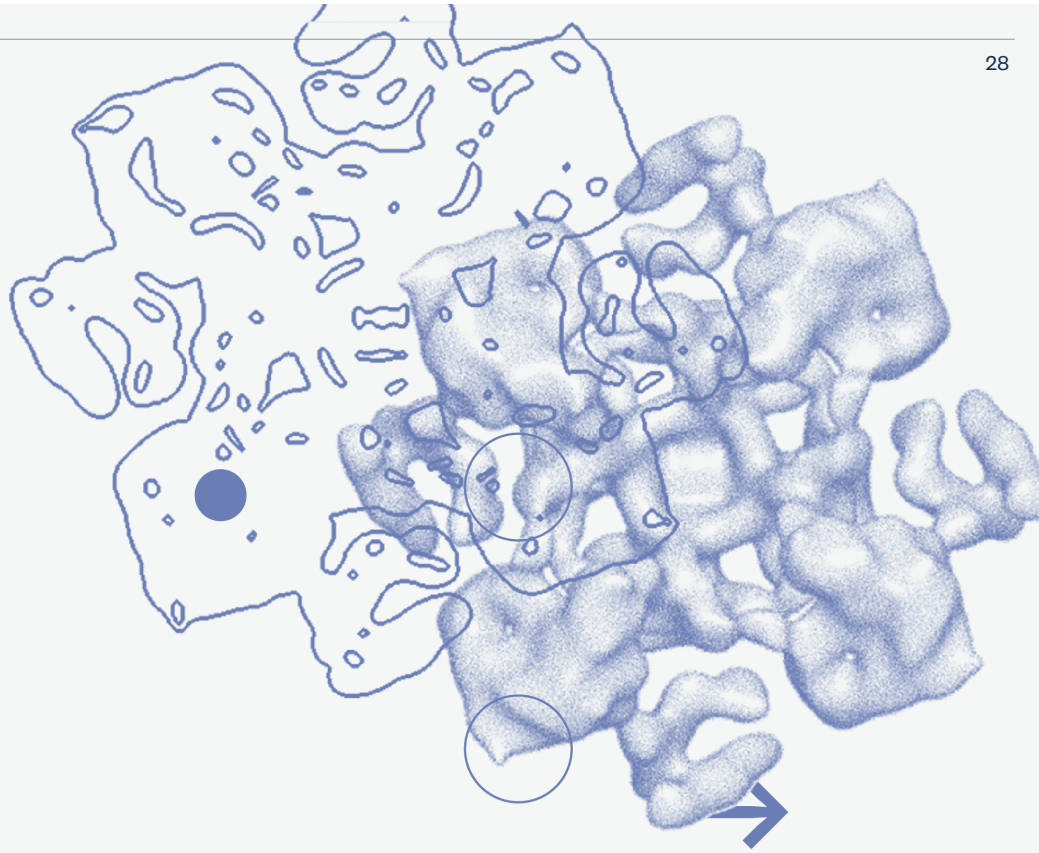
Opening Pandora's box

- ***Extension to many dimensions***
- Experimental validation system
- How to optimize the molecular path?
- Comparing many models vs many images optimization.

WS Tang



EH Thiede



Cryo-EM ensemble reweighting*

Extension to multiple dimensions: working directly in configuration space.

*Tang et al. 2022 https://arxiv.org/abs/2212.05320

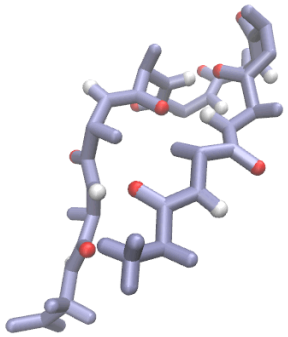
Theory

data

CryoEM images
Y={y_i}

Conformation

$$x \in \mathbb{R}^{3N}$$



Boltzmann distribution

$$p(x|H) = \frac{1}{Z} e^{-\beta H(x)}$$

Hamiltonian

Bayes' theorem

$$p(H|Y) \propto p(H)p(Y|H)$$

we would like to know

For a set of *iid* cryo-EM particles
(not an averaged observable)

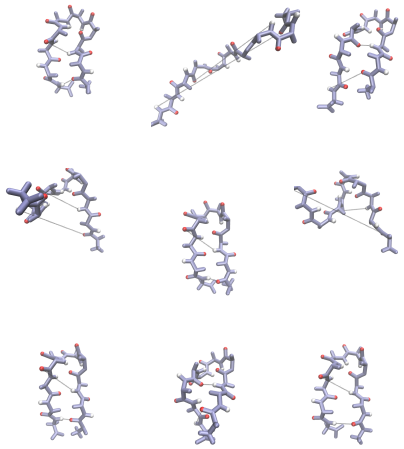
$$p(H|Y) \propto p(H) \prod_i \int p(y_i|x_i)p(x_i|H)dx_i$$

Approximation:
we parametrize the density by a set of $\{x_m\}$ with weights $\{\alpha_m\}$.

data

CryoEM images
Y=$\{y_i\}$

Set of $\{x_m\} \in \mathbb{R}^{3N}$



Approximate probability density

$$p(x|\{\alpha_m\}) = \sum_m \delta(x - x_m)\alpha_m$$

↑
Sum to one

The posterior reduces to

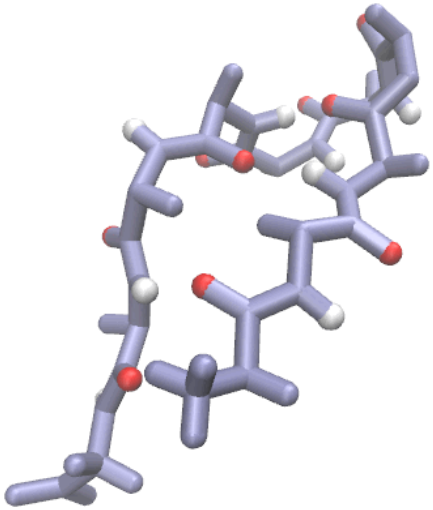
$$p(\{\alpha_m\}|Y) \propto p(\{\alpha_m\}) \prod_i \sum_m p(y_i|x_m)\alpha_m$$

↑
Goal: extract the weights

*Same expression as in cryoBIFE (Sci Rep 2021), so we can use STAN!

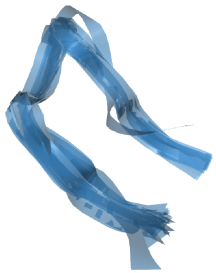
Test system Chignolin: Synthetic data

Sufficiently small to have a converged ensemble with MD.



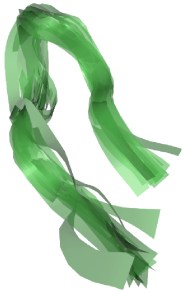
But complex enough with three metastable states:

Folded



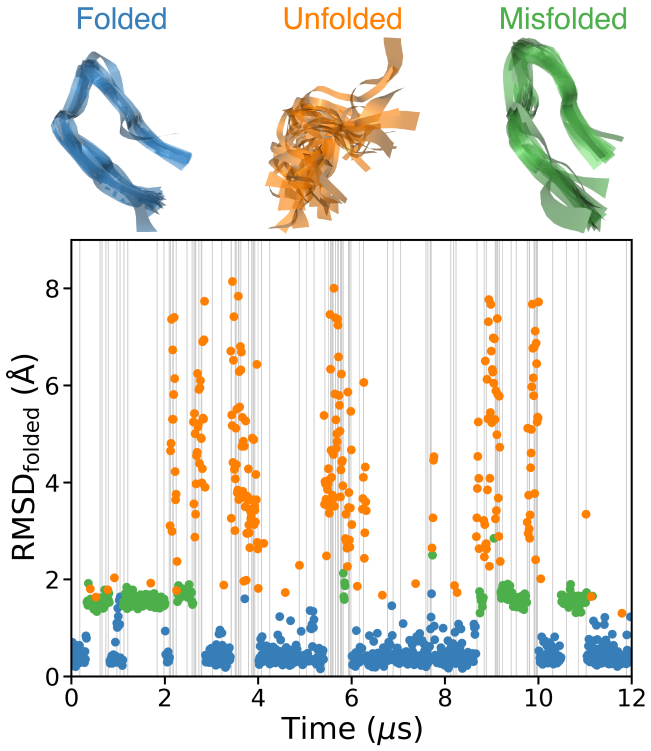
Unfolded

Misfolded



Two independent ensembles:

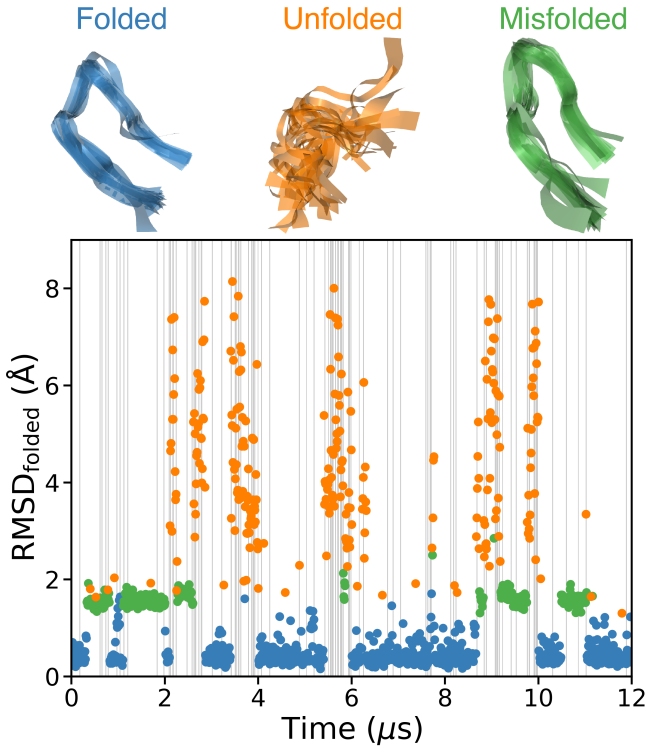
Structure-generating "in house"
MD trajectory



Cluster conformations with K-medoids
RMSD (50 centers) = $\{x_m\}$

Two independent ensembles:

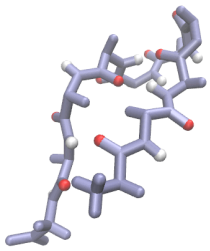
Structure-generating "in house" MD trajectory



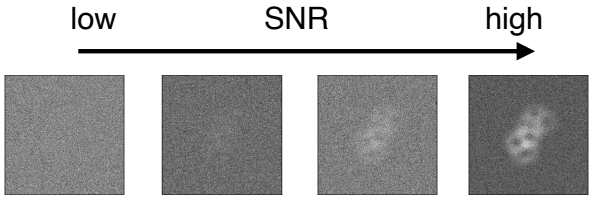
Cluster conformations with K-medoids
RMSD (50 centers) = $\{x_m\}$

Image-generating DE Shaw MD trajectory

—> Different force field
—> Different ensemble
—> Structures to generate images are not in the ensemble for refinement.



Example images with different SNR (random pose & defocus)



For simplicity, we assume that the pose and defocus known.

Can we recover the state populations?

The population is defined as the sum of the weights of the members (i.e., cluster centers) of each state.

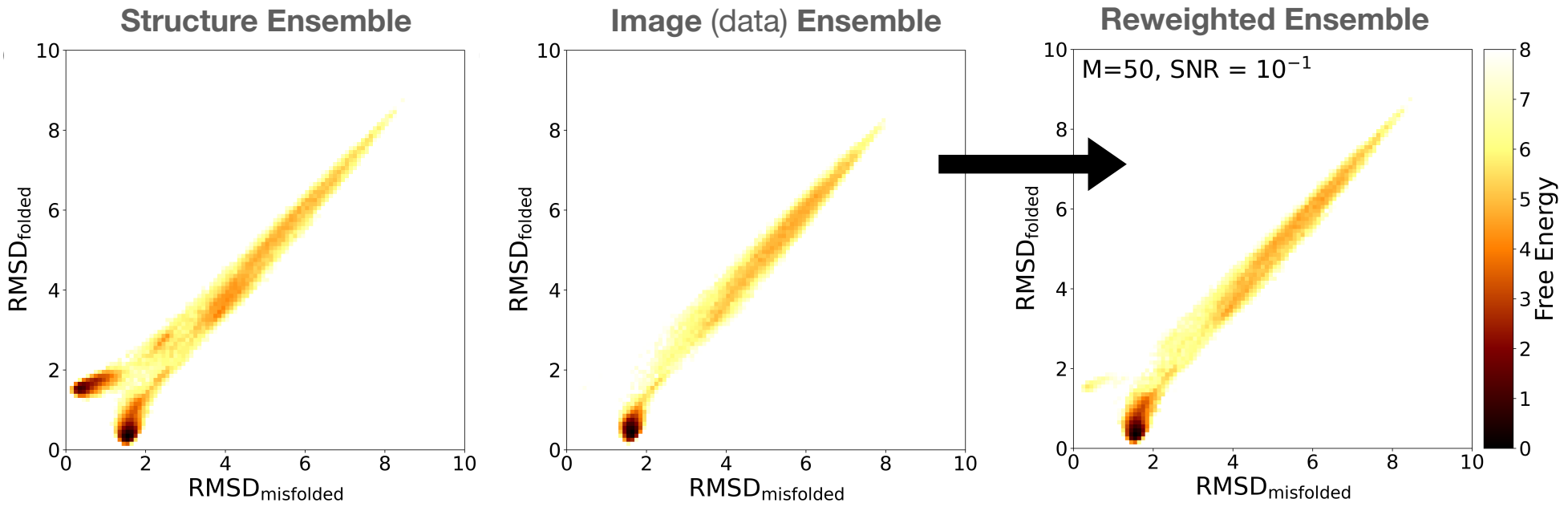
Population recovery*:

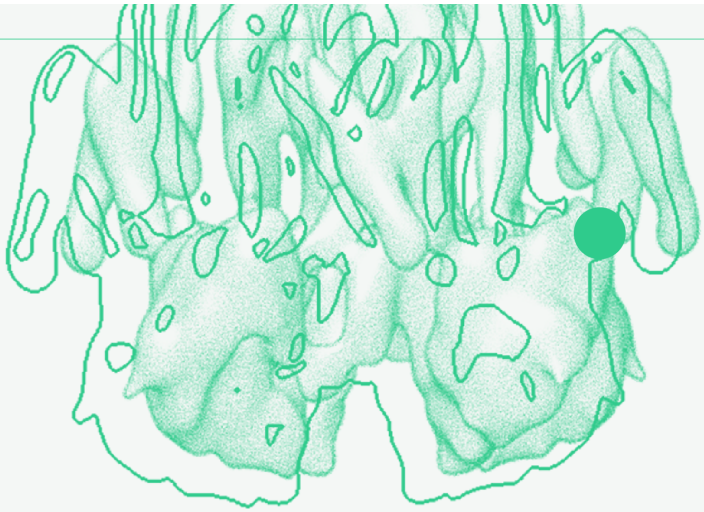
	%fld	%msfd	%unfd
Ground Truth	**0.7707**	**0.0004**	**0.2289**
No noise	0.7784	0.0058	0.2152
SNR = 1.0	0.7786	0.0055	0.2160
SNR = 0.1	0.7787	0.0058	0.2155
SNR = 0.01	0.7693	0.0129	0.2178

Yes, even at low SNRs*

*If we have to optimize the pose then it becomes more challenging

Post-processing of the refined ensemble: Free-energy over CVs of choice*.



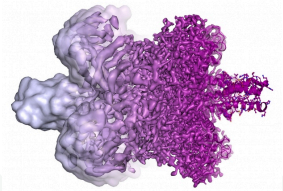


Open questions:

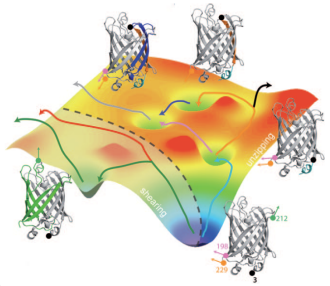
- Validation system for conformational variability?
- How to generate a ‘good’ ensemble —> bias MD

A stepping stone toward conformational landscapes *in situ*.

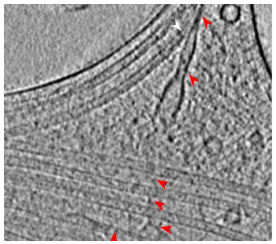
Single-particle



BIOLOGICAL ENVIRONMENT



In situ



Challenging!

Acknowledgments

FLATIRON INSTITUTE

Structural and Molecular Biophysics, CCB/CCM

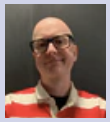
CCM

Leslie Greengard

Marylou Gabrie

Erik Thiede

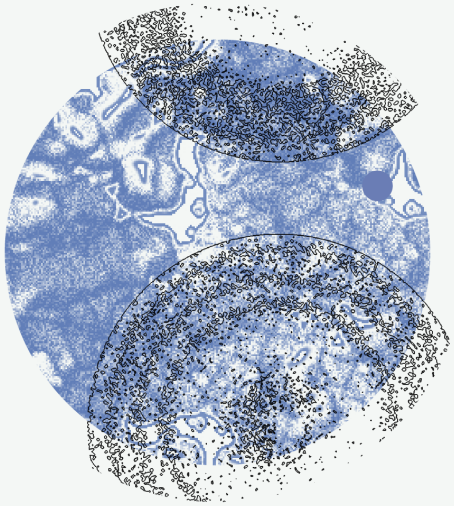
Alex Barnett



Bob Carpenter

Collaborators:
Steve Bonilla, Roberto Covino, Attila Szabo, Fabio Pietrucci, Karen Palacio-Rodriguez, Hadrien Vroylandt,….

THANKS FOR YOUR ATTENTION



FLATIRON INSTITUTE