CryoBoltz

NEURAL INFORMATION PROCESSING SYSTEMS

Multiscale Guidance of Protein Structure Prediction with Heterogeneous Cryo-EM Data

Rishwanth Raghu

Axel Levy

Gordon Wetzstein

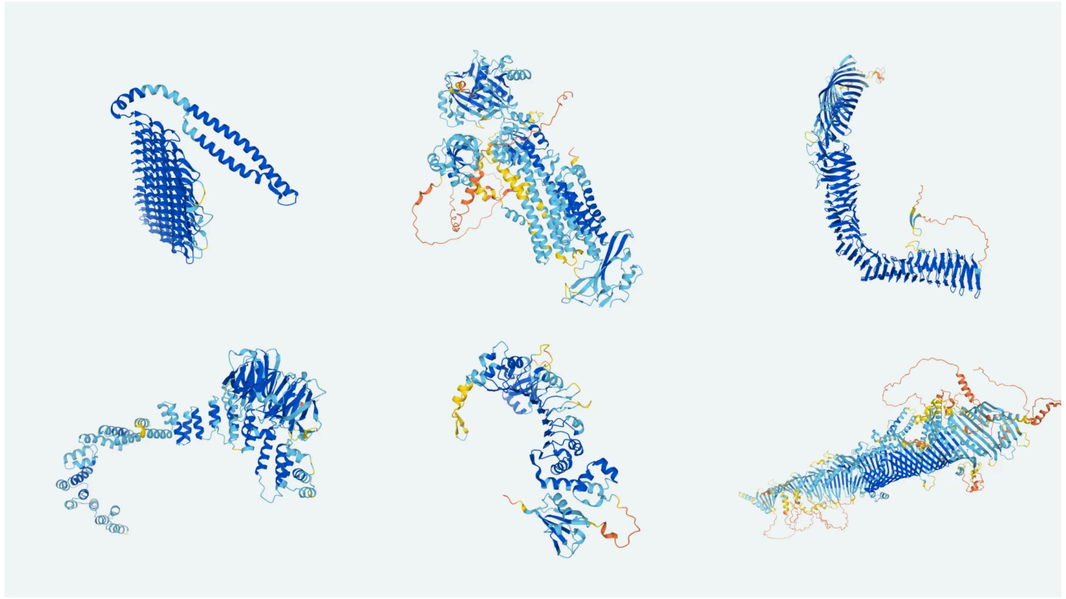
Ellen Zhong

PRINCETON UNIVERSITY

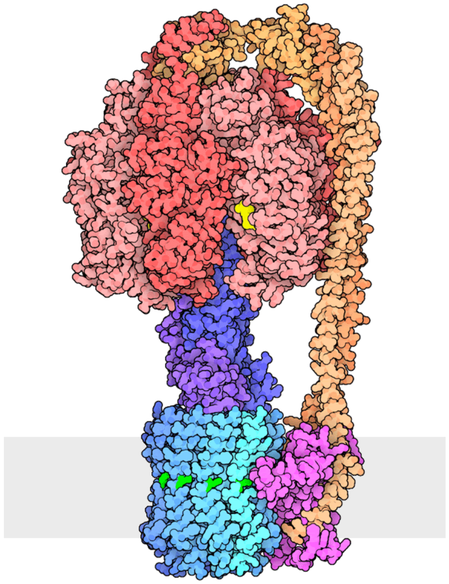
Stanford University

Protein Structure Prediction

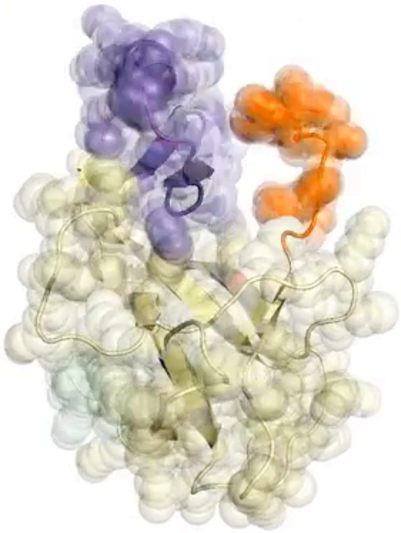
- Protein structure prediction models, e.g. AlphaFold [1], predict static 3D protein structure with high accuracy
- Protein structure is highly dynamic
- Predictive models are limited in their ability to predict multiple conformational states



AlphaFold Database. Tunyasuvunakool et al, 2021.



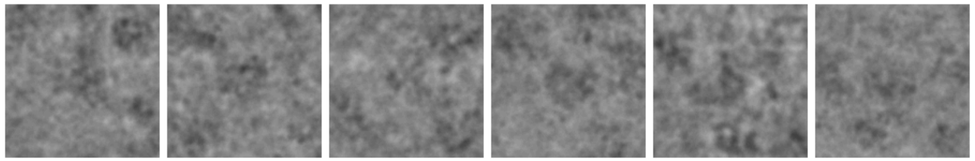
PDB 101



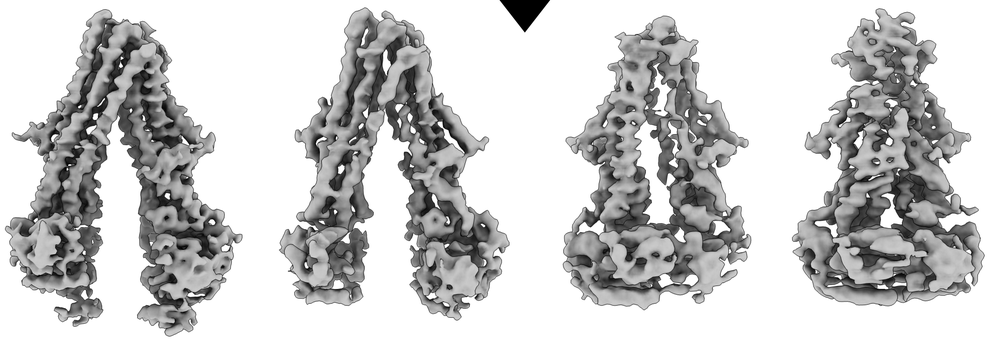
Chen et al eLife 2019

[1] Jumper et al. “Highly accurate protein structure prediction with AlphaFold.” *Nature* 2021.

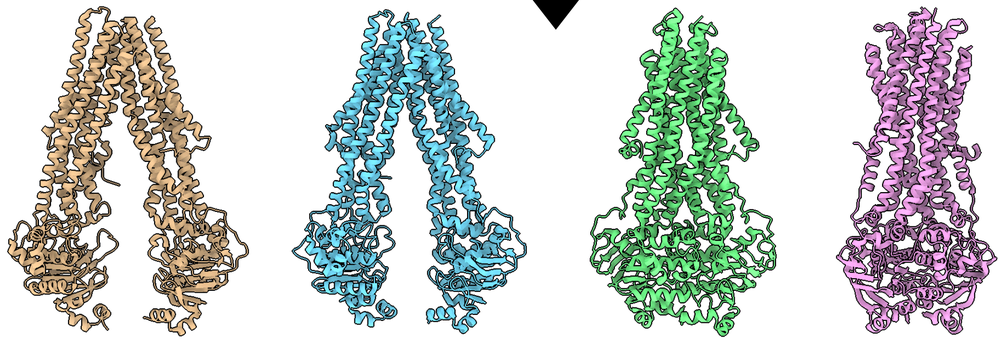
Conformational Ensembles from Cryo-EM



2D cryo-EM images



Ensemble of 3D density maps



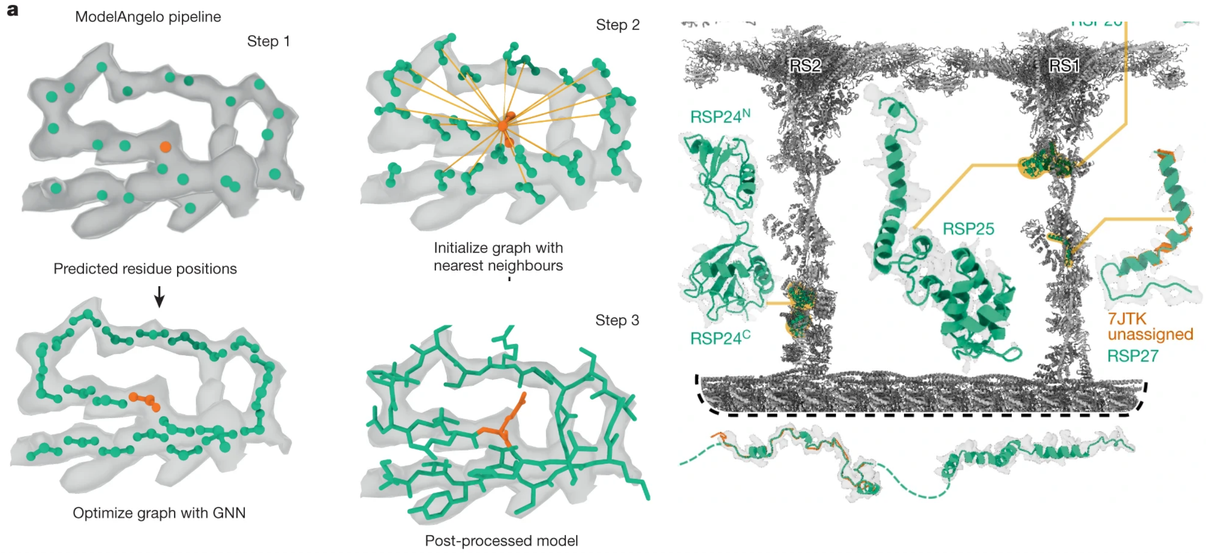
Ensemble of atomic models

✅ Captures near-native conformational ensembles

❌ Model building is challenging and unscalable

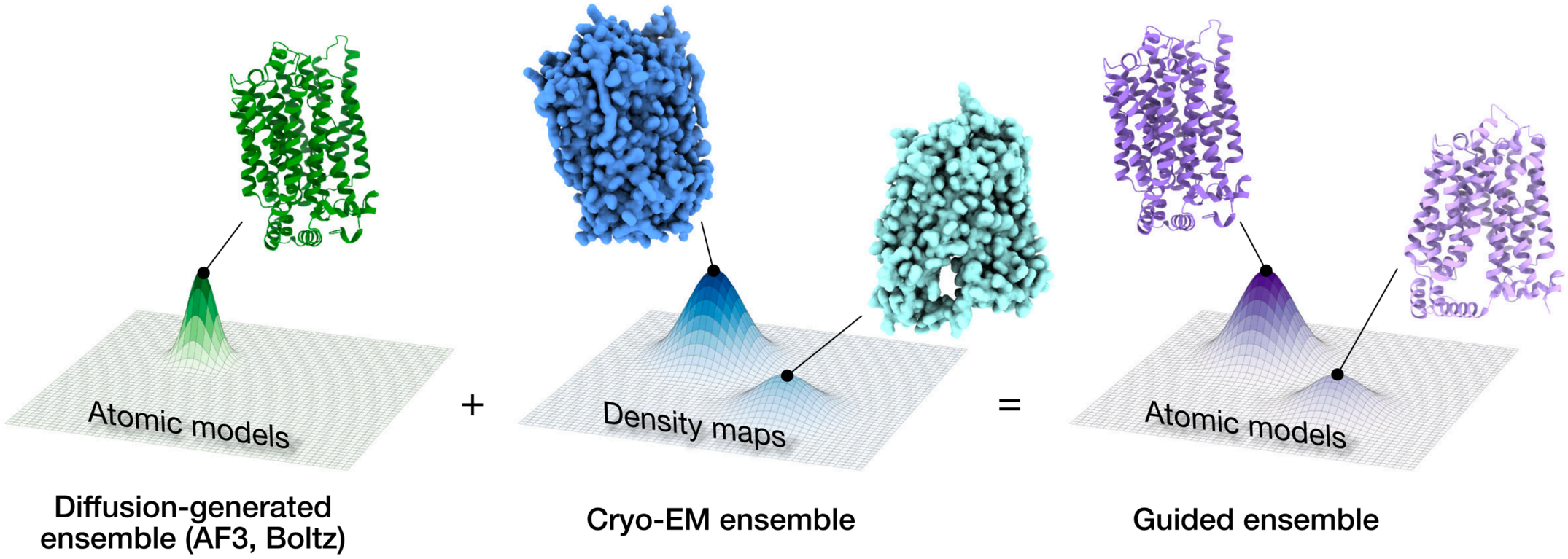
Model building

- Manual building
- MD simulation based flexible fitting
- ModelAngelo - Automated building for high-resolution maps [Jamali *et al.* 2024]



CryoBoltz Approach

We combine **experimental cryo-EM data** with the biophysical priors learned by **protein structure prediction** models, to produce **ensembles of atomic models**

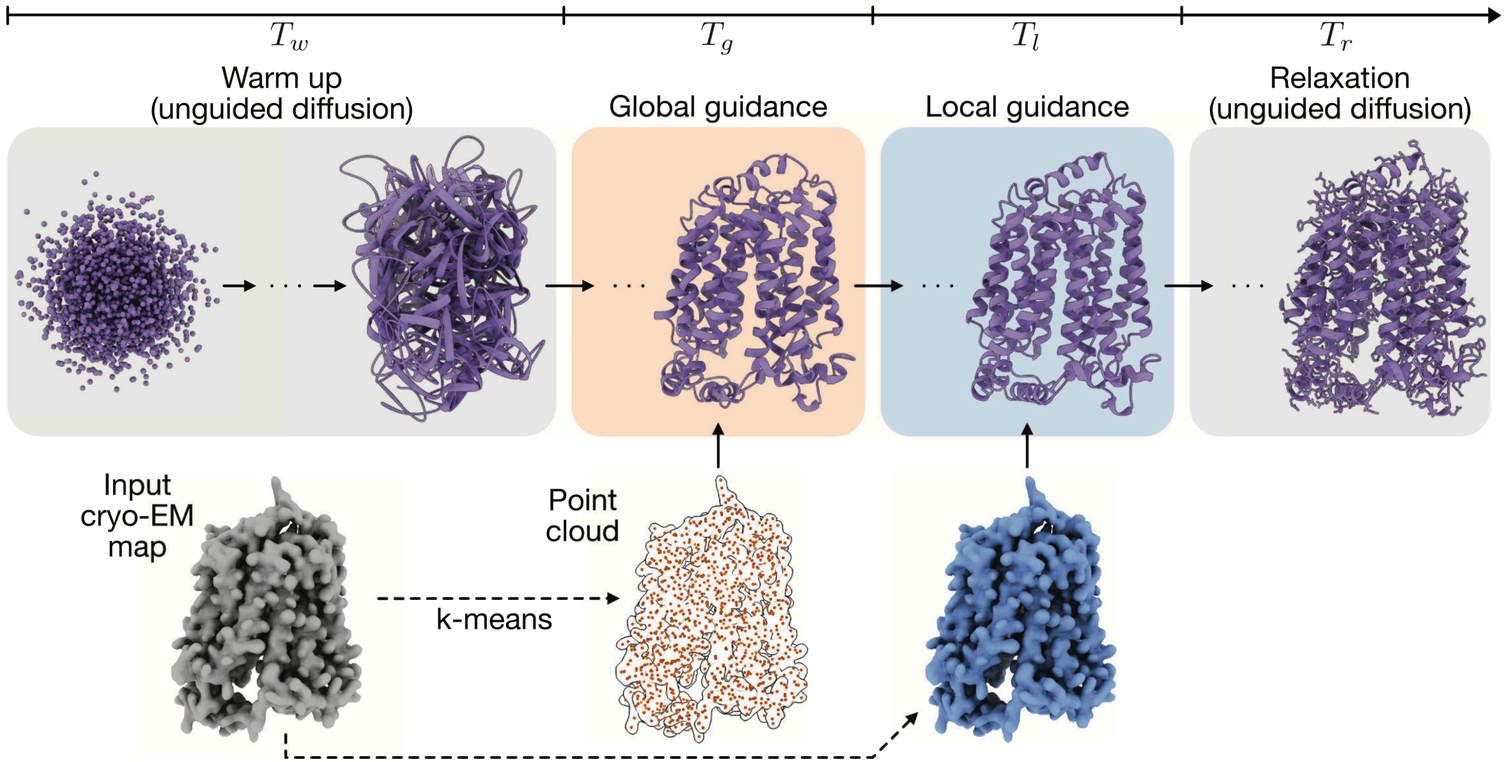


Diffusion Guidance with Cryo-EM Maps

- We use diffusion posterior sampling [1] to bias the sampling trajectory of Boltz-1 [2] towards structures consistent with an input cryo-EM map

- A coarse-to-fine fitting strategy combines optimal transport and physics-based guidance terms

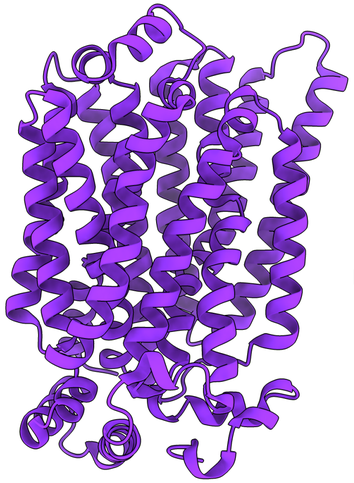
$$\mathrm{d}\mathbf{x} = \left(\mathbf{f}(\mathbf{x}, t) - g(t)^2 s_\theta(\mathbf{x}, t) - \underbrace{\lambda(t)\, \nabla_{\mathbf{x}} \log p(\mathbf{y} | \mathbf{x}_0 = \hat{\mathbf{x}}_\theta(\mathbf{x}, t))}_{\text{Experimental guidance}} \right) \mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$



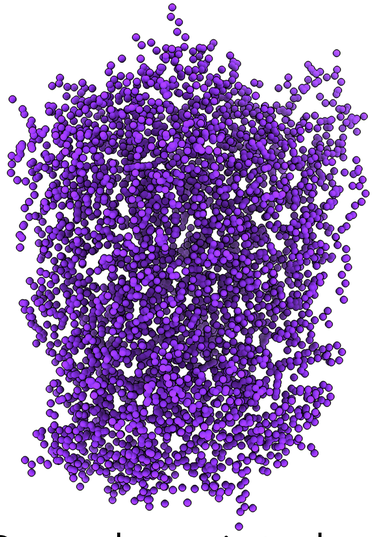
[1] Chung et al. "Diffusion posterior sampling for general noisy inverse problems." *ICLR* 2023.
[2] Wohlwend et al. "Boltz-1 democratizing biomolecular interaction modeling." *bioRxiv*, 2024.

Diffusion Guidance with Cryo-EM Maps

Global

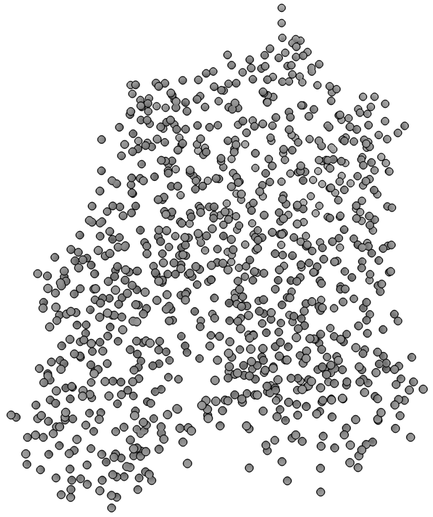


Diffusion sample



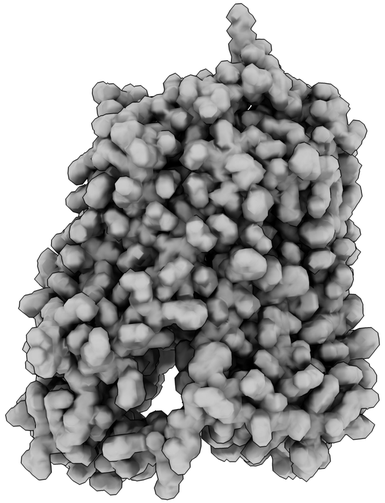
Sample point cloud

Sinkhorn divergence



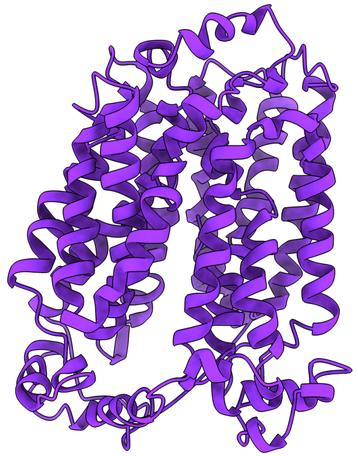
Target point cloud

Weighted k-means



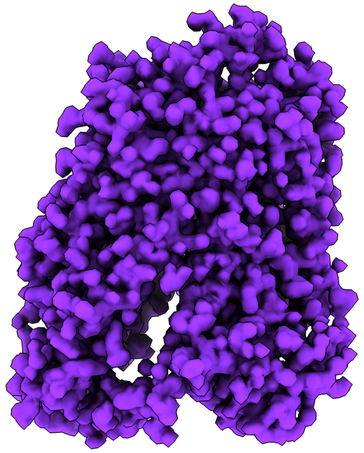
Input map

Local



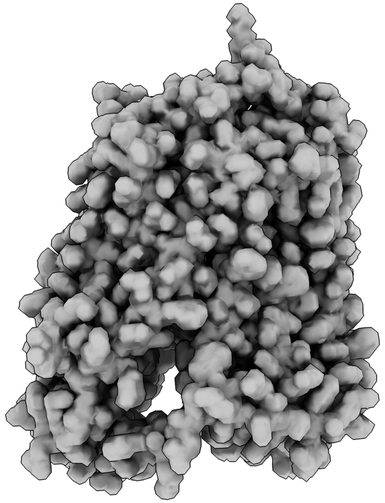
Diffusion sample

Cryo-EM forward model



Simulated map

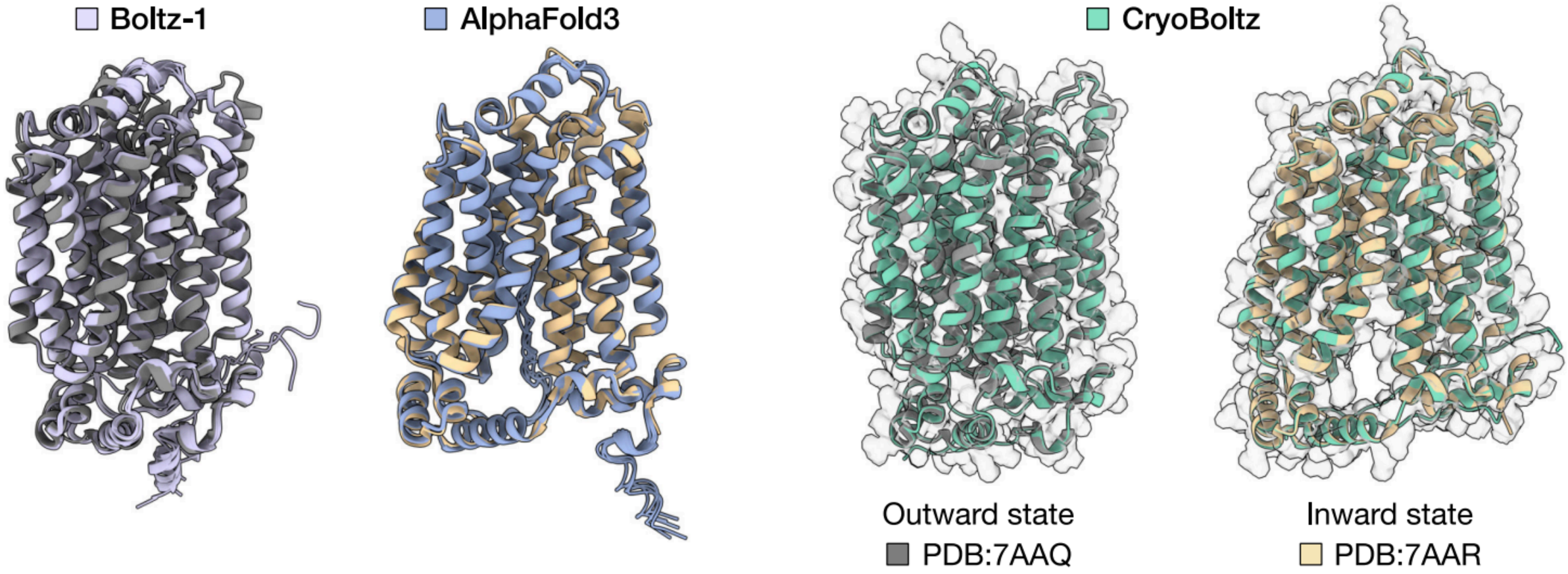
Mean squared error



Input map

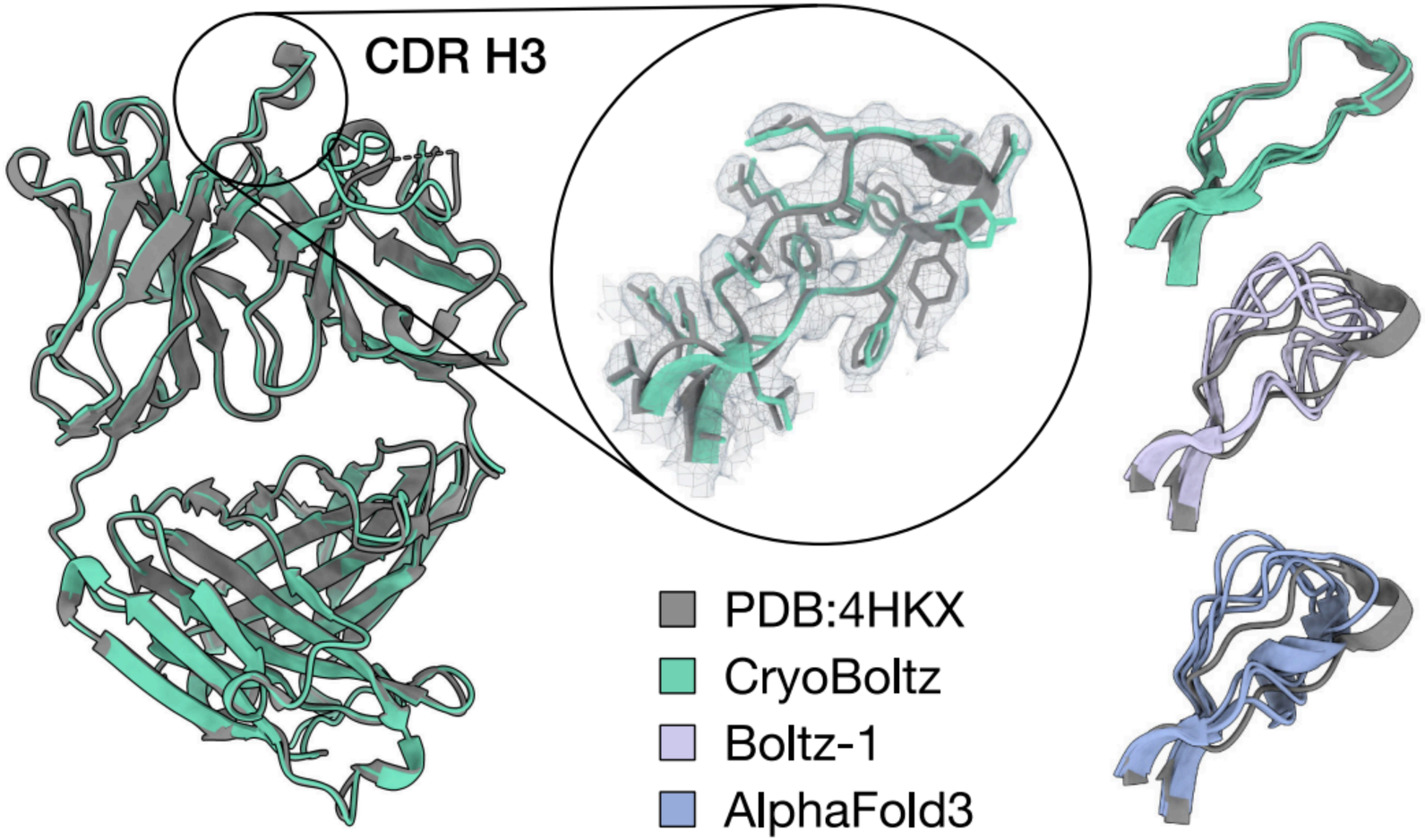
Guidance Overcomes Single-Structure Bias

Boltz-1 and AF3 predict only one conformation of the STP10 transporter, whereas guidance with synthetic maps achieves both



Local Fitting of Antibody CDR Loops

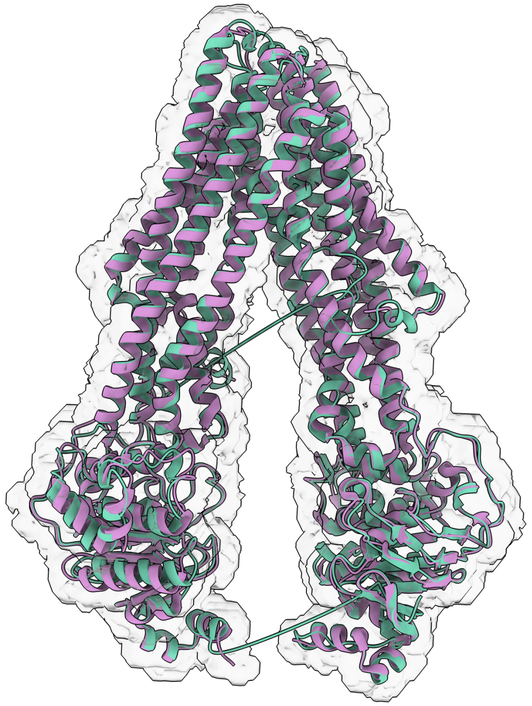
CryoBoltz accurately fits loops and side chains in a synthetic antibody map



Recovering Conformations from Real Maps

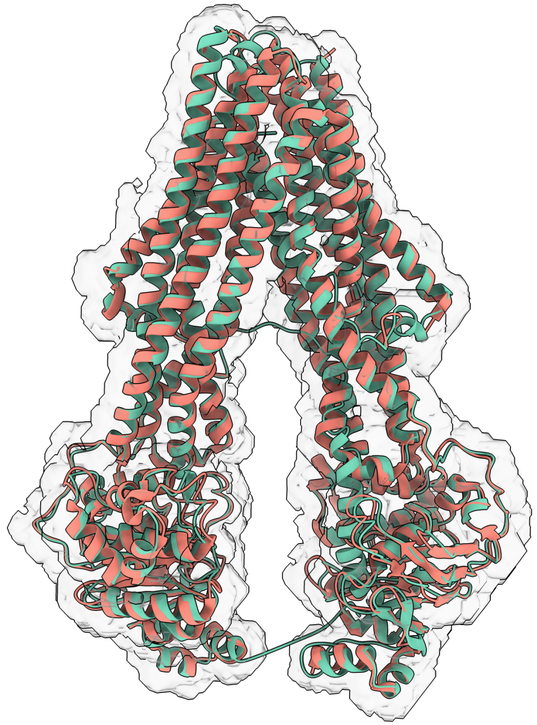
CryoBoltz accurately models 4 conformational states of P-glycoprotein

Apo State



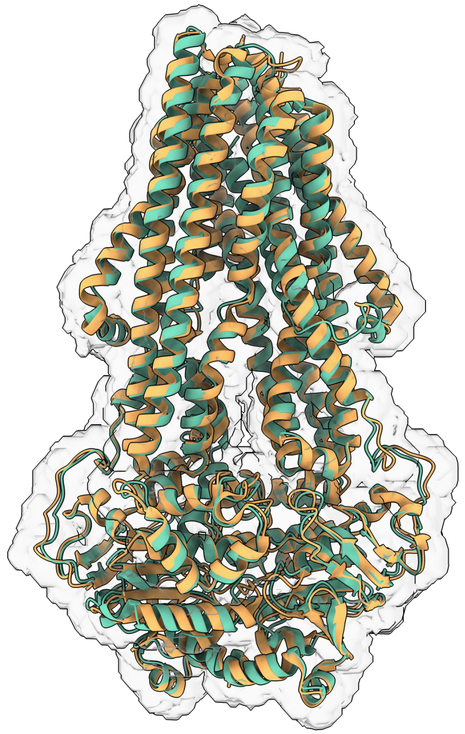
PDB 8GMG

Inward State



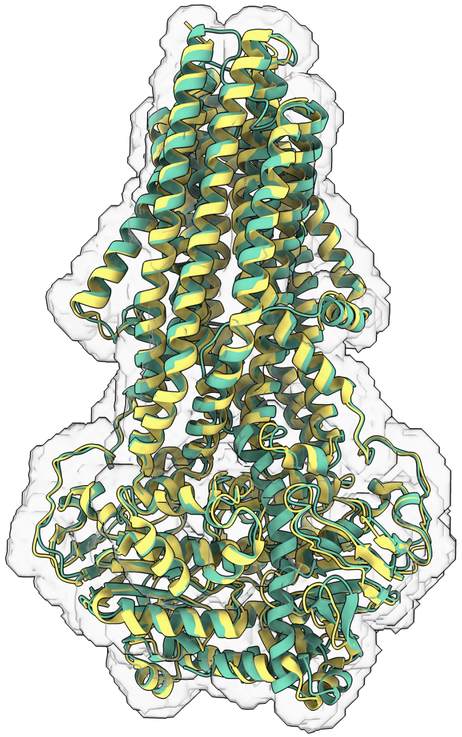
PDB 8SA1

Occluded State



PDB 8SA0

Collapsed State



PDB 8GMJ

CryoBoltz

Quantitative Evaluation

CryoBoltz models structures more accurately than unguided predictors and model building tools

		CryoBoltz			Boltz-1			AlphaFold3			ModelAngelo	
Structure	Res. (Å)	RMSD all	RMSD Cα	TM score	RMSD all	RMSD Cα	TM score	RMSD all	RMSD Cα	TM score	Comp. (%)	TM score
Pgp (apo)	4.3	**1.382**	**1.208**	**0.989**	6.994	7.194	0.767	3.827	3.865	0.904	40.3	0.361
Pgp (inward)	4.4	**1.348**	**1.187**	**0.989**	5.630	5.692	0.828	2.692	2.663	0.947	18.3	0.134
Pgp (occluded)	4.1	**1.727**	**1.677**	**0.979**	2.929	2.904	0.942	3.440	3.420	0.921	2.3	0.010
Pgp (collapsed)	4.4	**1.309**	**1.261**	**0.988**	3.425	3.412	0.917	4.568	4.554	0.864	2.5	0.010
Pma1 (active)	3.25	**2.046**	**1.776**	**0.973**	2.987	2.752	0.935	6.628	6.389	0.769	91.5	0.889
Pma1 (inhibited)	3.52	**1.999**	**1.590**	**0.979**	6.140	5.829	0.794	8.017	7.776	0.723	72.8	0.721
CYP (open)	6.5	**4.167**	**3.946**	**0.957**	8.532	8.439	0.788	6.490	6.361	0.890	0.0	0.000
CYP (closed)	4.4	**2.004**	**1.552**	**0.990**	8.784	8.667	0.743	3.585	3.391	0.946	18.9	0.102
YbbAP (bound)	3.66	**1.320**	**0.678**	**0.997**	3.623	3.339	0.928	3.749	3.480	0.922	81.7	0.801
YbbAP (unbound)	4.05	**2.454**	**2.039**	**0.974**	7.842	7.654	0.776	4.022	3.744	0.913	55.4	0.548

cryoboltz.cs.princeton.edu