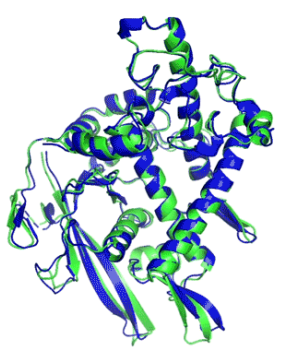
De novo prediction of protein structural dynamics

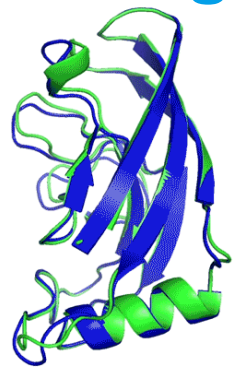
Diego del Alamo, PhD
Principal Scientist, Protein Design & Informatics, GSK

Protein structure prediction is solved, but protein folding is not

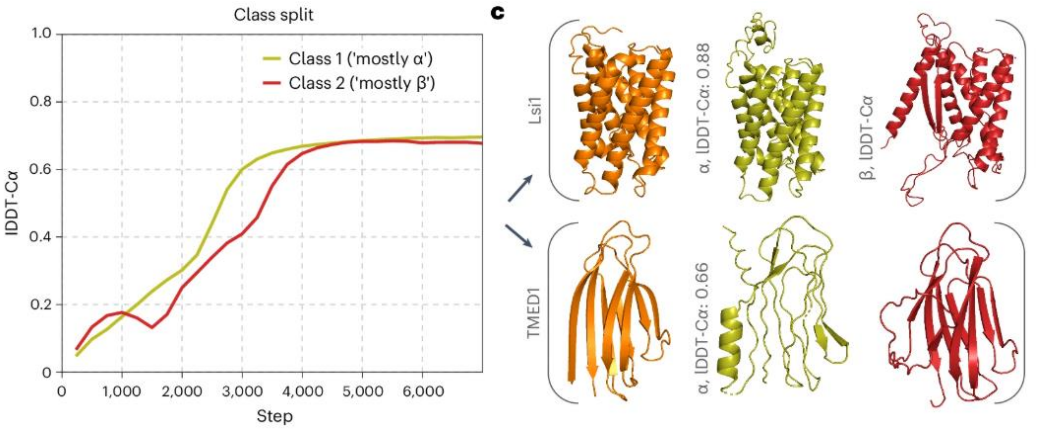
- State-of-the-art AI methods can reliably predict static structures of monomeric proteins to high accuracy **(Simpkin 2023)**
  - Includes *de novo* designed proteins **(Verkuil 2022)**
  - Unseen topologies **(Ahdritz 2024, Frank 2024)**
- These <u>do not extend</u> to modeling their dynamics: how/when proteins move
- Some questions about dynamics can still be answered by vanilla protein structure prediction neural networks
- Newer architectures can answer others



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



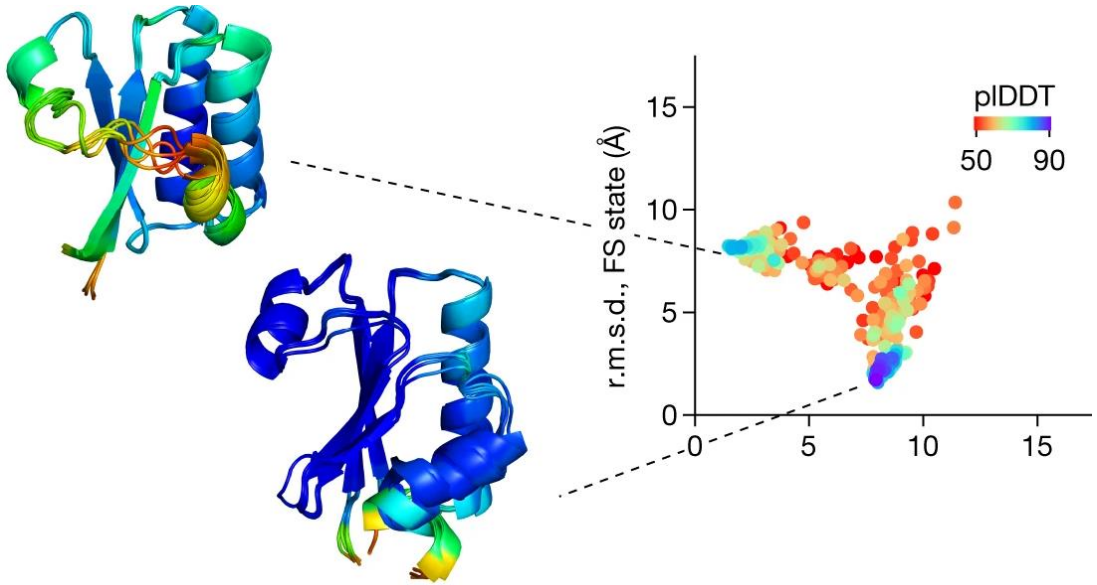
T1049 / 6y4f
93.3 GDT
(adhesin tip)



Ahdritz et al "OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization"Nat Methods 2024
Simpkin et al "Tertiary structure assessment at CASP15" Proteins 2025
Verkuil et al "Language models generalize beyond natural proteins" bioRxiv 2022
DeepMind "AlphaFold" Github 2021

Protein structure prediction

- Current ensemble prediction methods can model:
  - User-specified alternate conformations
  - Some aspects of Brownian motion
- They cannot:
  - Model relative proportions of different populations **(Vani 2023, Riccabona 2024)**
  - Model interconversion dynamics/transition paths **(del Alamo 2023)**
  - Model the effects of mutations, ligands, or modifications on these populations **(Zheng 2025, Ramasamy 2025)**
- Requires lots of tinkering with settings, trial & error



del Alamo et al "Conformational sampling and interconversion using language-based protein folding neural networks" bioRxiv 2023
Ramasamy et al "Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability" bioRxiv 2025
Riccabona et al "Assessing AF2's ability to predict structural ensembles of proteins" Structure 2024
Vani et al "Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE" JCTC 2023
Wayment-Steele et al "Predicting multiple conformations via sequence clustering and AlphaFold2" Nature 2023
Zheng et al "AlphaFold3 in Drug Discovery: A comprehensive assessment of capabilities, limitations, and applications" bioRxiv 2025

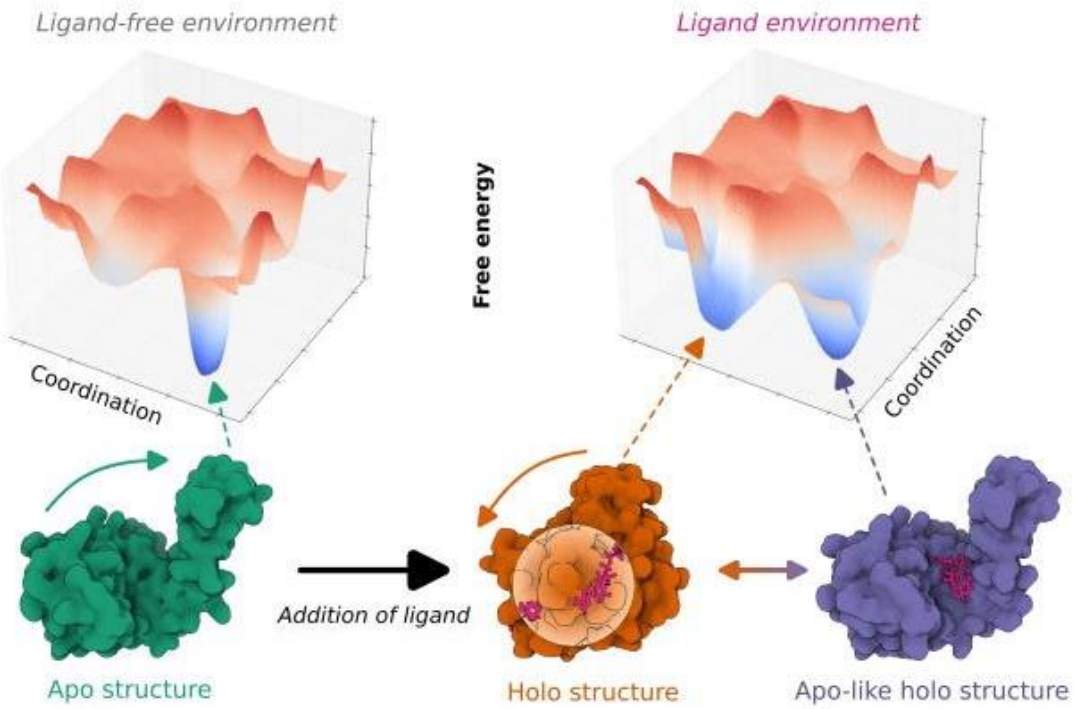
De novo prediction of protein structural dynamics: an outline

- What are protein dynamics?
- What kinds of biological problems are informed by studying dynamics?
- How can vanilla protein folding neural networks be modified to model dynamics?
- What modified and bespoke methods exist to answer these questions?
- Why don’t these methods learn protein dynamics?
- Integration of these methods with molecular dynamics

- **Caveat:** this talk will only cover backbone modeling of well-structured proteins, and will ignore innovations in modeling intrinsically disordered proteins (IDPs), oligonucleotides, or sidechain ensembles

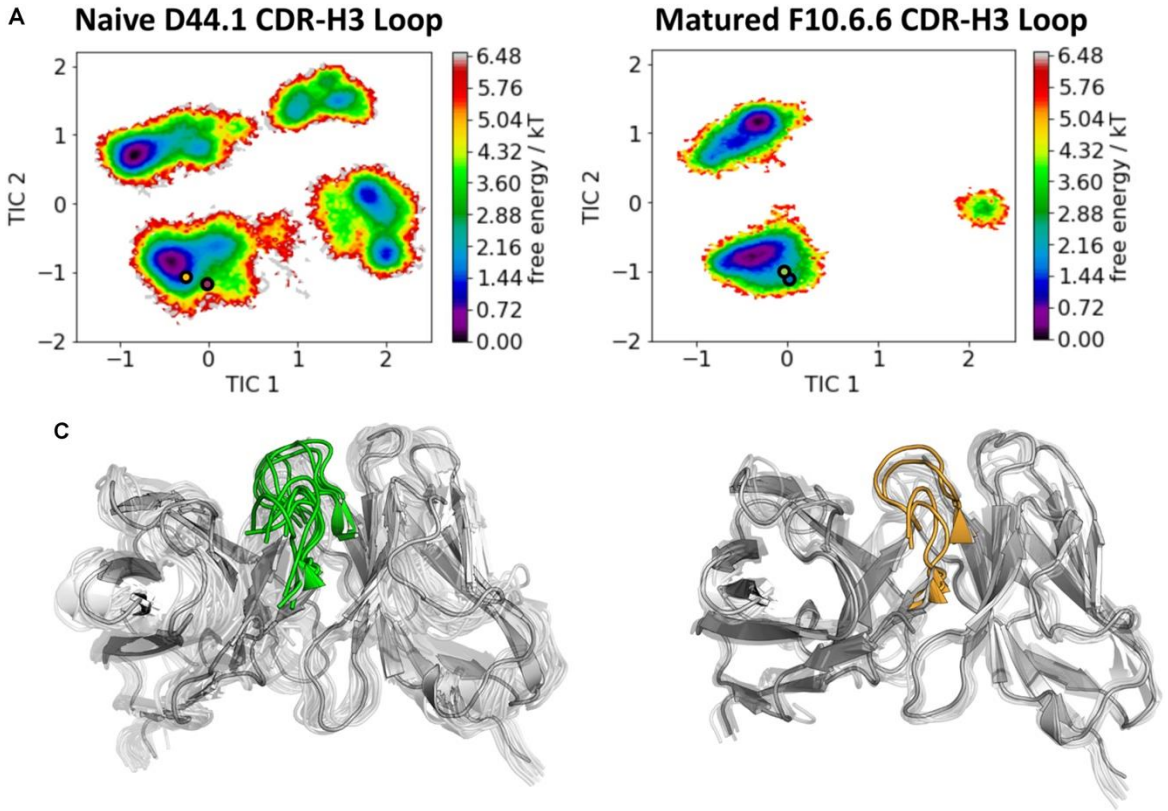
What are protein dynamics, anyway?

1. Breadth of motion, e.g., Brownian motion (microstates)
2. Structures of distinct low-energy states (macrostates)
3. Boltzmann weights between distinct low energy populations
   - Transition paths and energy barriers separa these populations
   - The whole thing is called the free energy landscape
4. Effect of ligands, mutations, covalent modifications, and combinations thereof on the free energy landscape
   - Metals, phosphoryl groups, waters, protons, etc



What protein dynamics tell us about biological mechanisms

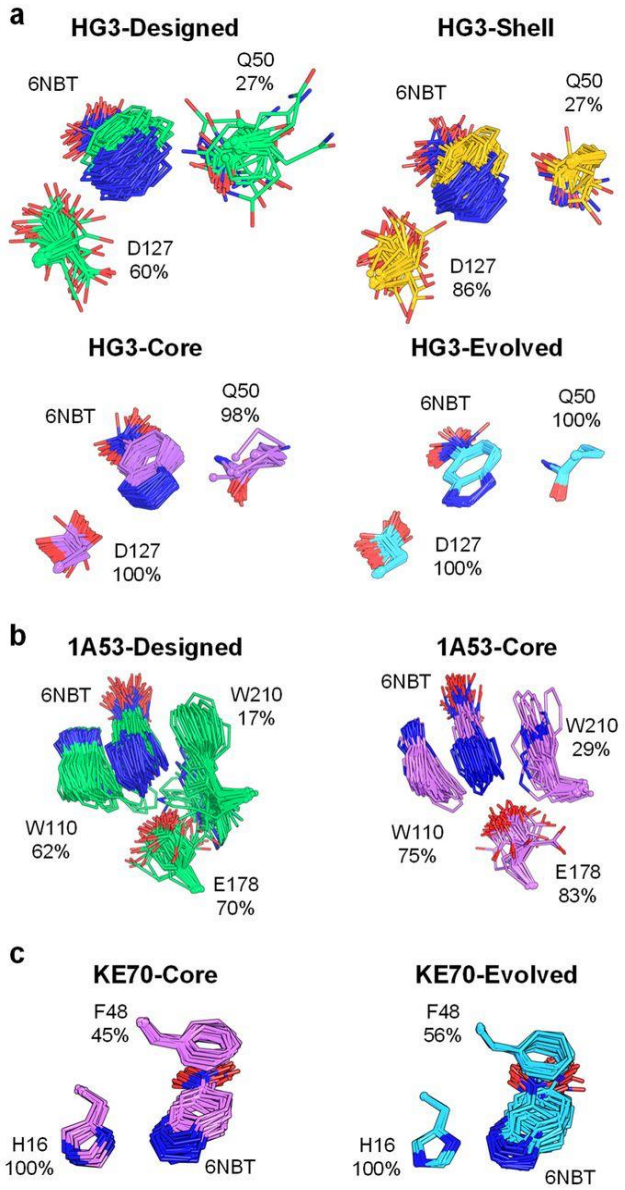
- **Conformational heterogeneity and binding promiscuity (antibodies)**
- Active site preorganization (designed and/or engineered enzymes)
- Cryptic pockets and alternative druggable states (drug targets)
- Molecular basis of genetic diseases (almost all human proteins)
- Evolution of functional novelty (most proteins)
- Sometimes, structural modeling isn't the best way to answer biological questions about protein dynamics



Dishman et al, "Evolution of fold switching in a metamorphic protein" Science 2021
Fernandez-Quintero et al, "Local and Global Rigidification Upon Antibody Affinity Maturation" Frontiers in Immunology 2020
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Tokuriki & Tawfik "Protein dynamism and evolvability" Science 2009
Yee et al "The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics" Mol Cell 2025
Zarifi et al "Distal mutations enhance catalysis in designed enzymes by facilitating substrate binding and product release" bioRxiv 2025

What protein dynamics tell us about biological mechanisms

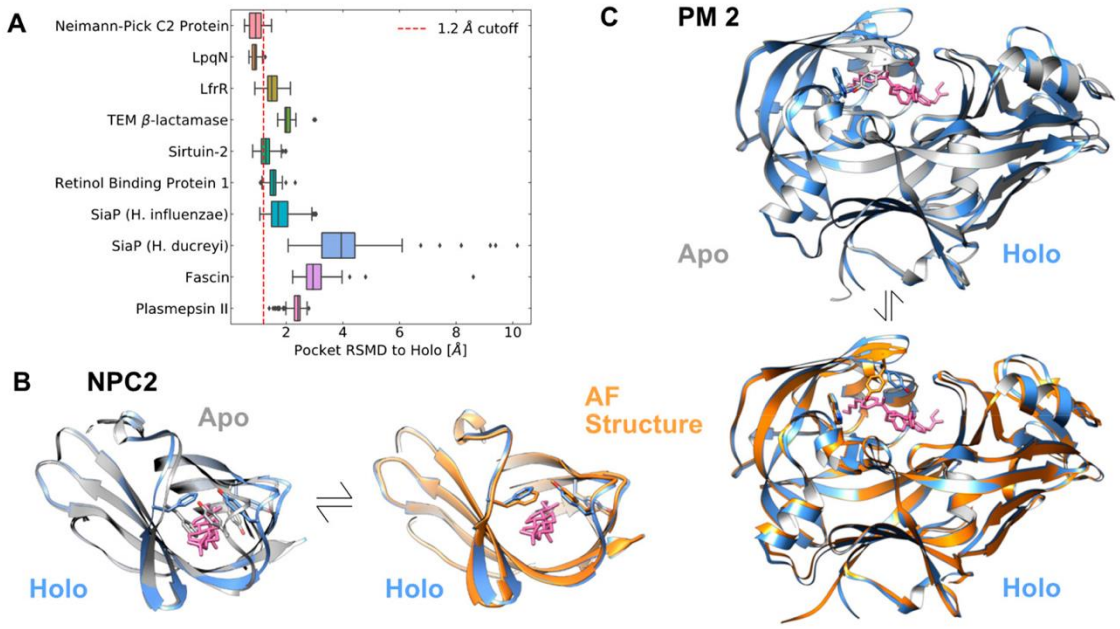
- Conformational heterogeneity and binding promiscuity (antibodies)
- **Active site preorganization (designed and/or engineered enzymes)**
- Cryptic pockets and alternative druggable states (drug targets)
- Molecular basis of genetic diseases (almost all human proteins)
- Evolution of functional novelty (most proteins)
- Sometimes, structural modeling isn't the best way to answer biological questions about protein dynamics



Dishman et al, "Evolution of fold switching in a metamorphic protein" Science 2021
Fernandez-Quintero et al, "Local and Global Rigidification Upon Antibody Affinity Maturation" Frontiers in Immunology 2020
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Tokuriki & Tawfik "Protein dynamism and evolvability" Science 2009
Yee et al "The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics" Mol Cell 2025
Zarifi et al "Distal mutations enhance catalysis in designed enzymes by facilitating substrate binding and product release" bioRxiv 2025

What protein dynamics tell us about biological mechanisms

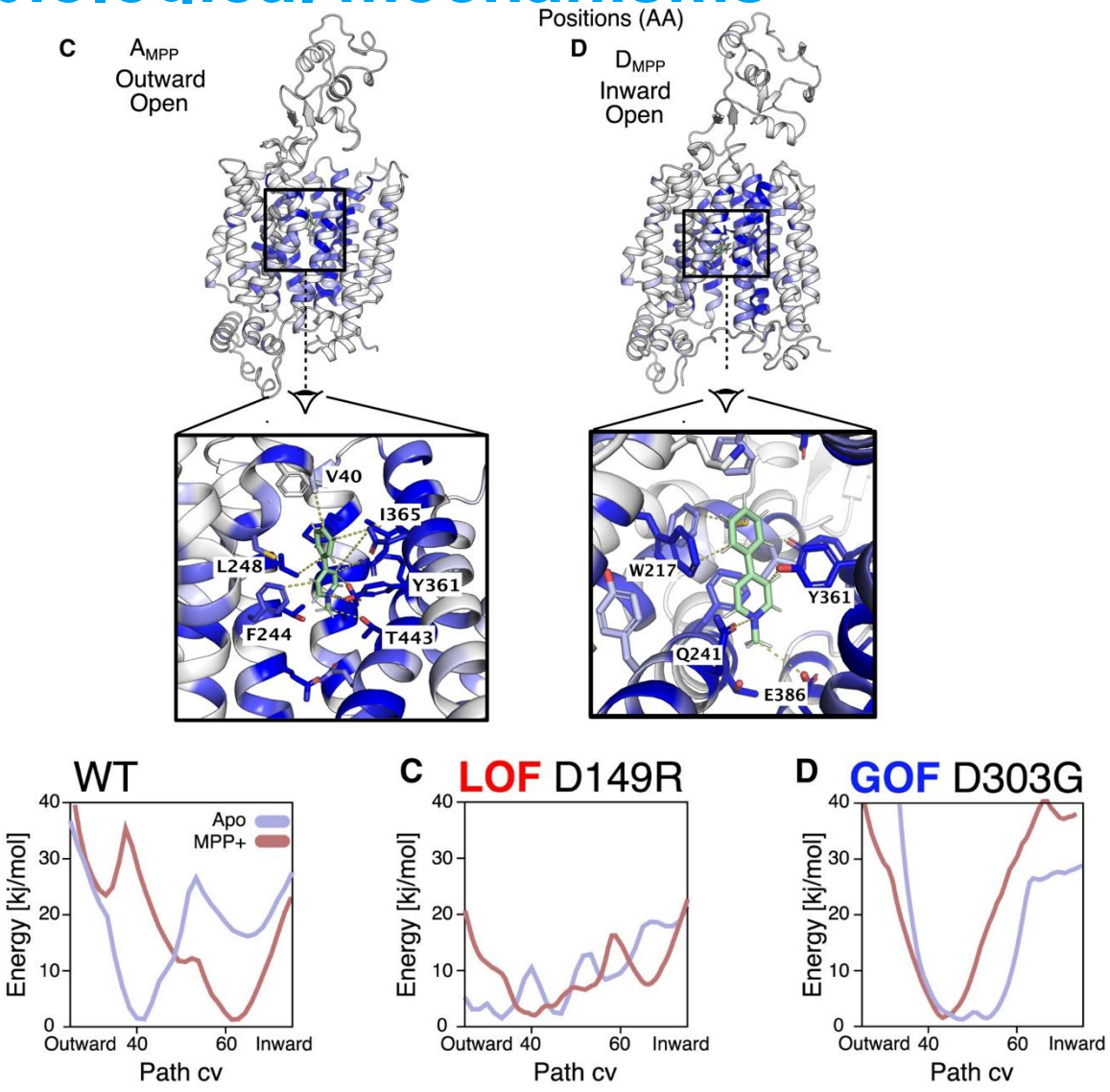
- Conformational heterogeneity and binding promiscuity (antibodies)
- Active site preorganization (designed and/or engineered enzymes)
- **Cryptic pockets and alternative druggable states (drug targets)**
- Molecular basis of genetic diseases (almost all human proteins)
- Evolution of functional novelty (most proteins)
- Sometimes, structural modeling isn't the best way to answer biological questions about protein dynamics



Dishman et al, "Evolution of fold switching in a metamorphic protein" Science 2021
Fernandez-Quintero et al, "Local and Global Rigidification Upon Antibody Affinity Maturation" Frontiers in Immunology 2020
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Tokuriki & Tawfik "Protein dynamism and evolvability" Science 2009
Yee et al "The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics" Mol Cell 2025
Zarifi et al "Distal mutations enhance catalysis in designed enzymes by facilitating substrate binding and product release" bioRxiv 2025

What protein dynamics tell us about biological mechanisms

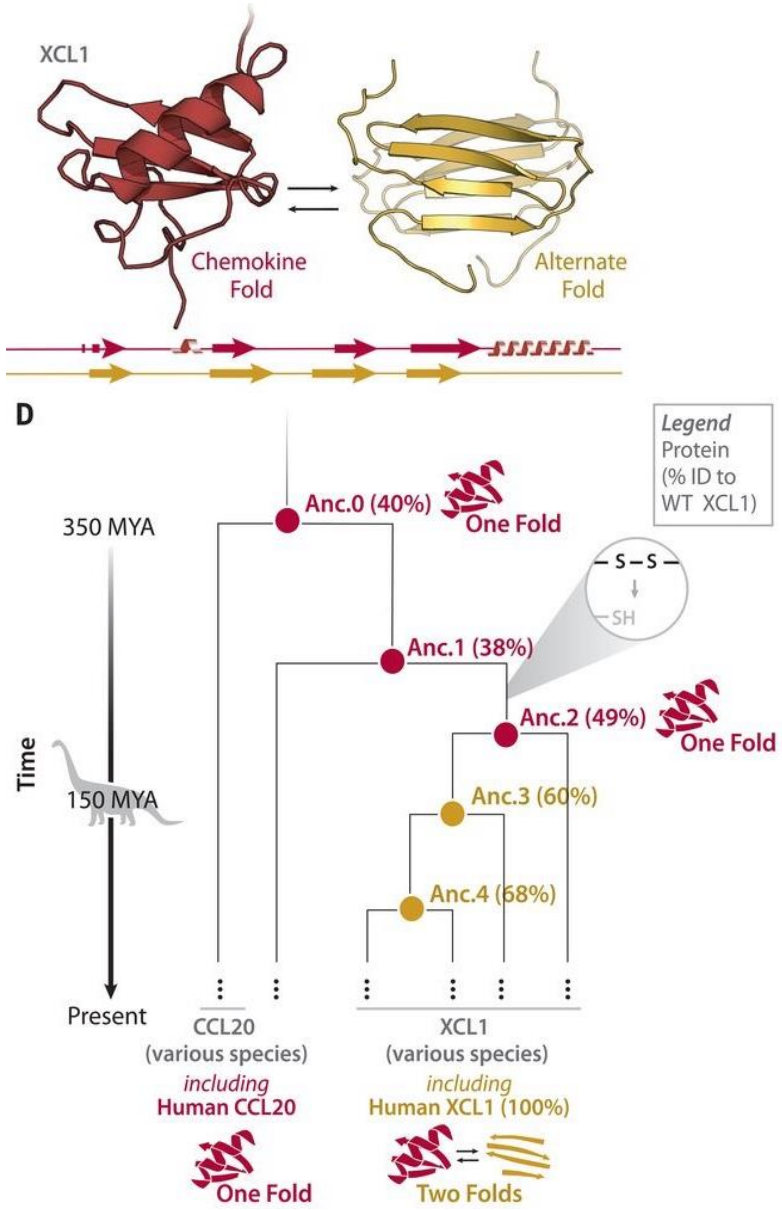
- Conformational heterogeneity and binding promiscuity (antibodies)
- Active site preorganization (designed and/or engineered enzymes)
- Cryptic pockets and alternative druggable states (drug targets)
- **Molecular basis of genetic diseases (almost all human proteins)**
- Evolution of functional novelty (most proteins)
- Sometimes, structural modeling isn't the best way to answer biological questions about protein dynamics



Dishman et al, "Evolution of fold switching in a metamorphic protein" Science 2021
Fernandez-Quintero et al, "Local and Global Rigidification Upon Antibody Affinity Maturation" Frontiers in Immunology 2020
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Tokuriki & Tawfik "Protein dynamism and evolvability" Science 2009
Yee et al "The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics" Mol Cell 2025
Zarifi et al "Distal mutations enhance catalysis in designed enzymes by facilitating substrate binding and product release" bioRxiv 2025

What protein dynamics tell us about biological mechanisms

- Conformational heterogeneity and binding promiscuity (antibodies)
- Active site preorganization (designed and/or engineered enzymes)
- Cryptic pockets and alternative druggable states (drug targets)
- Molecular basis of genetic diseases (almost all human proteins)
- **Evolution of functional novelty (most proteins)**
- Sometimes, structural modeling isn't the best way to answer biological questions about protein dynamics



Dishman et al, "Evolution of fold switching in a metamorphic protein" Science 2021
Fernandez-Quintero et al, "Local and Global Rigidification Upon Antibody Affinity Maturation" Frontiers in Immunology 2020
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Tokuriki & Tawfik "Protein dynamism and evolvability" Science 2009
Yee et al "The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics" Mol Cell 2025
Zarifi et al "Distal mutations enhance catalysis in designed enzymes by facilitating substrate binding and product release" bioRxiv 2025

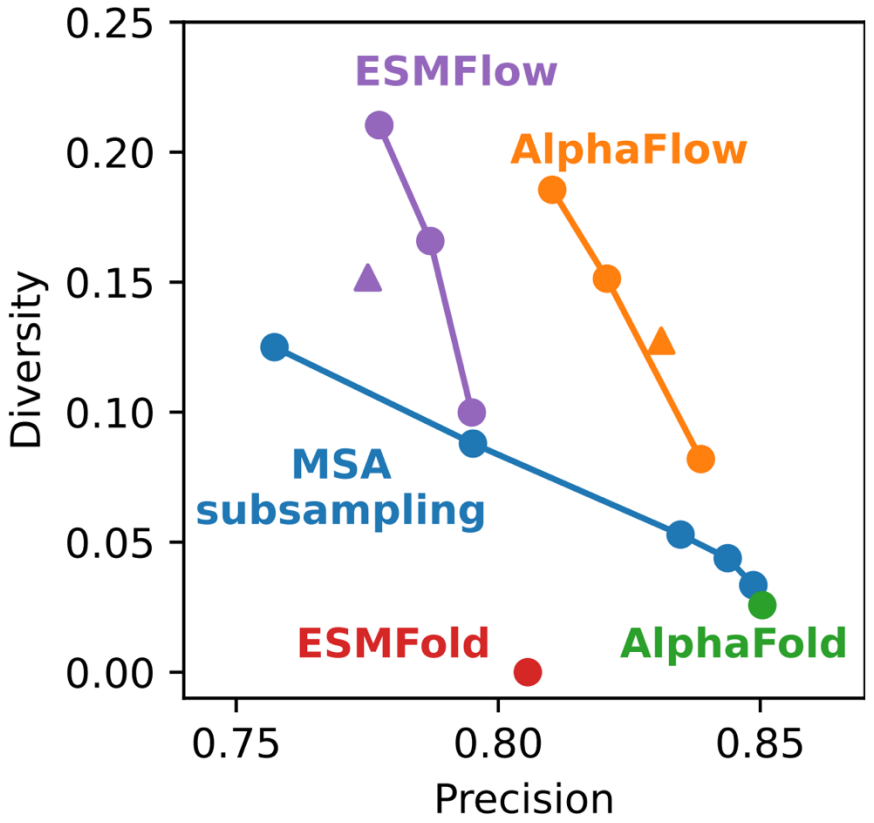
What protein dynamics tell us about biological mechanisms

- Conformational heterogeneity and binding promiscuity (antibodies)
- Active site preorganization (designed and/or engineered enzymes)
- Cryptic pockets and alternative druggable states (drug targets)
- Molecular basis of genetic diseases (almost all human proteins)
- Evolution of functional novelty (most proteins)
- ***Sometimes, structural modeling isn't the best way to answer biological questions about protein dynamics***

Dishman et al, "Evolution of fold switching in a metamorphic protein" Science 2021
Fernandez-Quintero et al, "Local and Global Rigidification Upon Antibody Affinity Maturation" Frontiers in Immunology 2020
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Tokuriki & Tawfik "Protein dynamism and evolvability" Science 2009
Yee et al "The full spectrum of SLC22 OCT1 mutations illuminates the bridge between drug transporter biophysics and pharmacogenomics" Mol Cell 2025
Zarifi et al "Distal mutations enhance catalysis in designed enzymes by facilitating substrate binding and product release" bioRxiv 2025

The state of the art

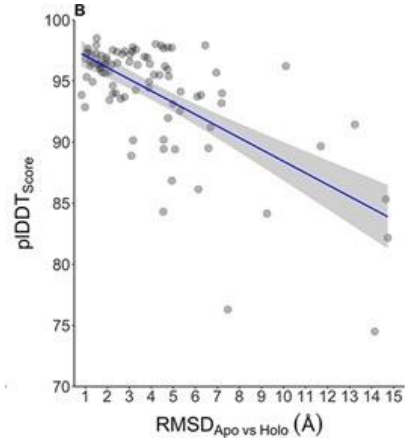
- Modeling alternative conformations:
  - Vanilla AlphaFold2 with subsampled or modified MSAs, custom template databases, etc
  - Fine-tuned AlphaFold2 (AlphaFlow, DiG, etc), ESMFold (ESMFlow), or RosettaFold2 neural networks
  - Trained from scratch (Str2str, aSAMt, Cfold)
- Modeling entire free energy landscapes
  - Distributional Graphormer (DiG), aSAMt, P2DFlow
  - BioEmu: Prediction of free energy landscapes of monomeric proteins to ~1 kcal/mol
- Precision/diversity trade-off: one highly accurate model vs lots of potentially inaccurate models
- MD is the gold standard, but is excrutiatingly expensive



Jing et al "AlphaFold meets Flow Matching for generating protein ensembles" ICLR 2024

Protein folding NNs already “know” dynamics

- Error/confidence metrics show some correlation with mobility/conformational flexibility in MD simulations **(Gavaldo-Garcia 2025, Jussupow 2023, Saldaño 2022)**
- Contacts for non-sampled states show up in distograms **(Jumper 2021)**
- (Note that error metrics in AF2 & RosettaFold2 are highly correlated; **Bennett 2023**)



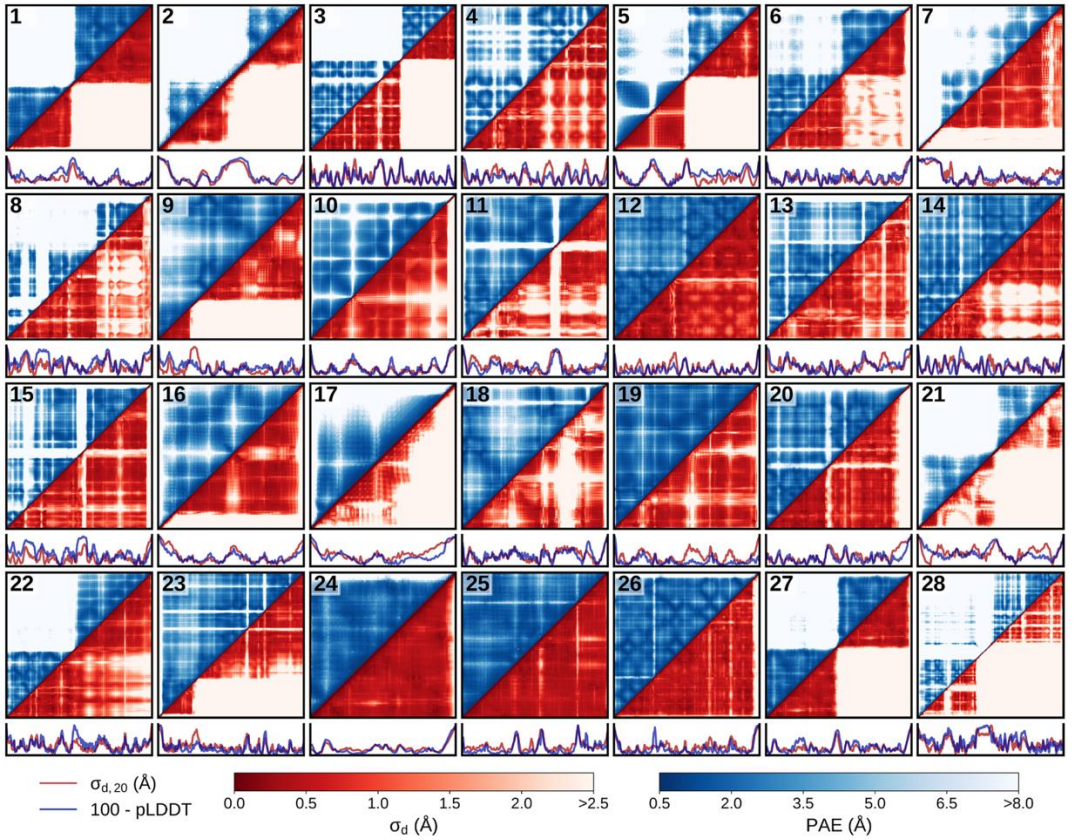
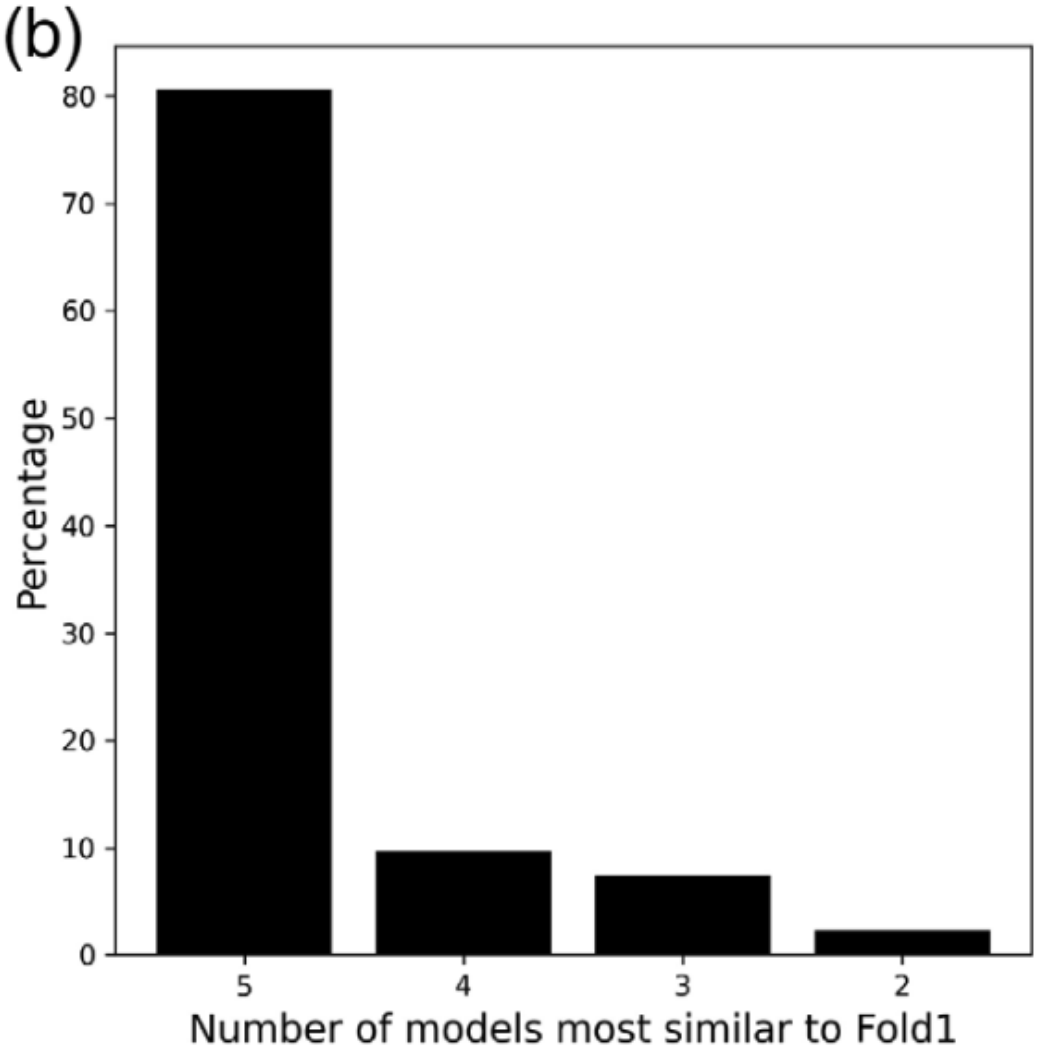


Figure 2. Overview of AlphaFold scores and protein dynamics metrics for all studied systems. A fully symmetric matrix indicates a perfect correlation between the AlphaFold and MD data. The figure is divided into 28 subfigures for each of the proteins shown in Figure 1. The top matrices show a comparison between (symmetrized) PAE matrices (blue) against the standard deviation of all C_α distances σ_d (red). The PAE scores range between 0.5 and 8.0, while the threshold for σ_d is set to <2.5 Å to enable better comparisons between the different systems. The bottom graphs compare the pLDDT scores (blue) and average standard deviation of the Cα-distances to the 20 closest amino acids $\sigma_{d,20}$ (red). See Figures S1–S28 for a detailed analysis of the individual systems.

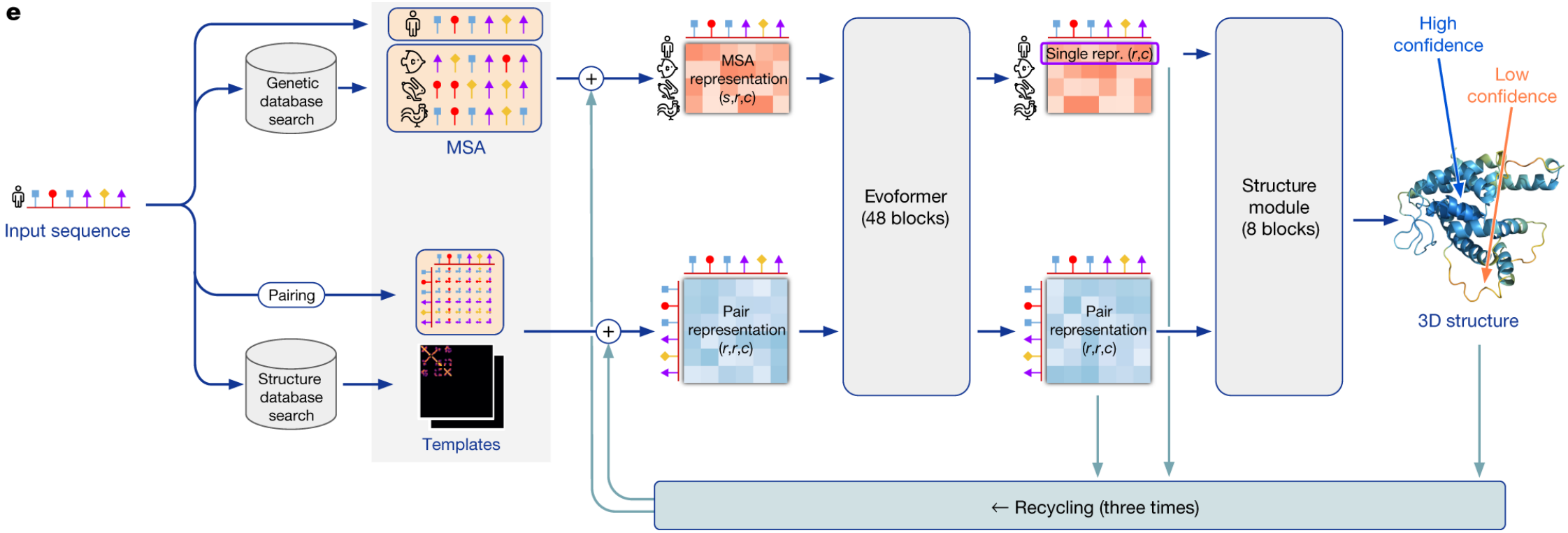
Bennett et al “Improving de novo protein binder design with deep learning” Nature Comms 2023
Gavalda-Garcia et al “Gradations in protein dynamics captured by experimental NMR are not well represented by AlphaFold2 models and other computational metrics” JMB 2025
Jussupow & Kaila “Effective Molecular Dynamics from Neural Network-Based Structure Prediction Models” JCTC 2023
Saldaño et al “Impact of protein conformational diversity on AlphaFold predictions” Bioinformatics 2022

Protein folding NNs tend to sample the same state repeatedly



Chakravarty & Porter "AlphaFold2 fails to predict protein fold switching" Protein Science 2022

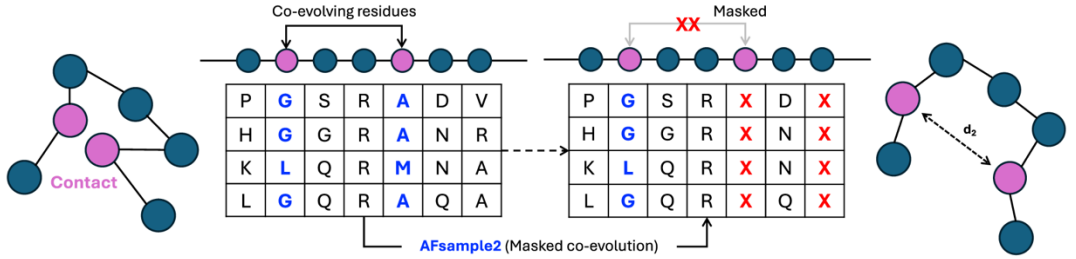
Review of AF2 architecture



Jumper et al "Highly accurate protein structure prediction with AlphaFold" Nature 2021

Predicting alternate conformational states with vanilla AlphaFold2

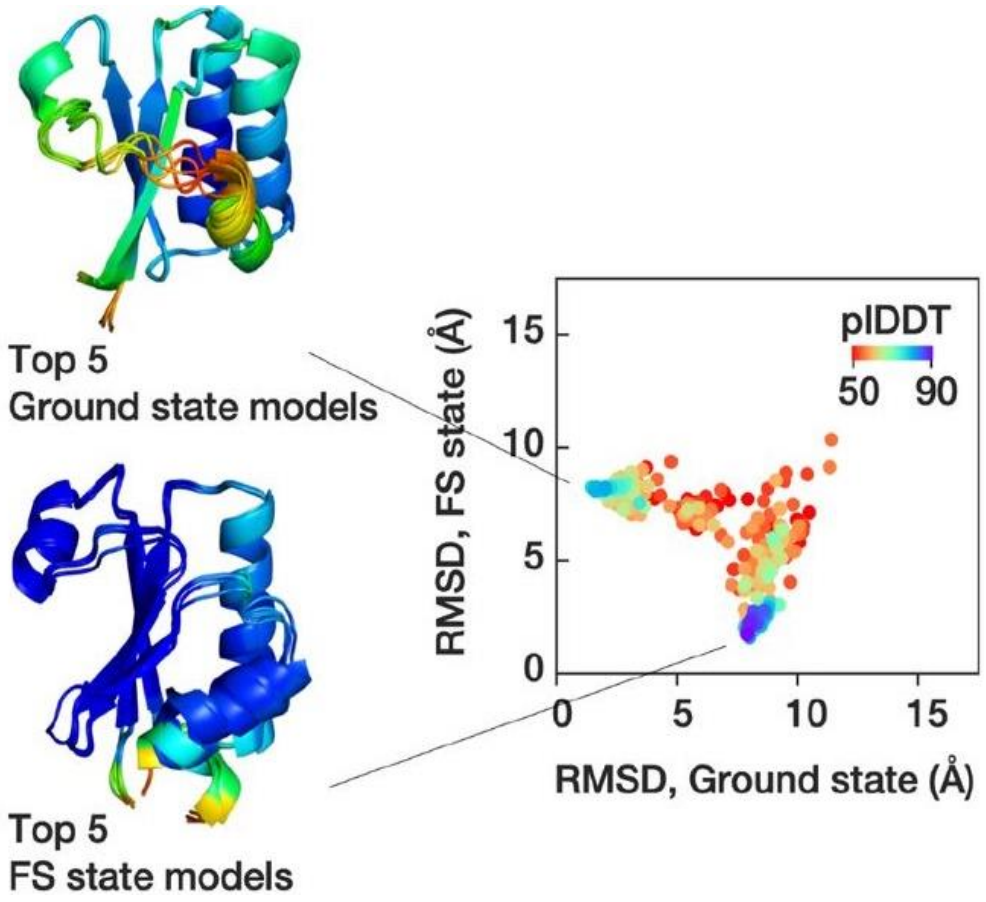
- Sequence-based methods that work
  - **MSA masking** (Stein 2022, Kalakoti 2025)
  - MSA sourcing (da Silva 2024, Faezov 2023)
  - MSA clustering (Wayment-Steele 2023)
  - MSA subsampling (del Alamo 2022)
  - Truncation of autoinhibitory domains (del Alamo 2022a)
  - Inclusion of conformationally selective binding partners (Cummins 2022)
- Sequence-based augmentations that don't work:
  - Point mutations, including those that work *in vitro* (Ramsamy 2025, Vani 2023, Zheng 2025; sometimes many mutations work)



da Silva et al, "High-throughput prediction of protein conformational distributions with subsampled AlphaFold2" Nature Comms 2024
Cummins et al, "AlphaFold accurately predicts distinct conformations based on the oligomeric state of a de novo designed protein" Prot Sci 2022
del Alamo et al, "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
del Alamo et al, "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter" PNAS 2022a
Kalakoti et al, "AFsample2 predicts multiple conformations and ensembles with AlphaFold2" Sci Rep 2025
Ramasamy et al "Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability"bioRxiv 2025
Stein et al , "SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2" Plos Comp Biol 2022
Wayment-Steele et al, "Predicting multiple conformations via sequence clustering and AlphaFold2" Nature 2023
Vani et al "Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE" JCTC 2023
Zheng et al "AlphaFold3 in Drug Discovery: A comprehensive assessment of capabilities, limitations, and applications" bioRxiv 2025

Predicting alternate conformational states with vanilla AlphaFold2

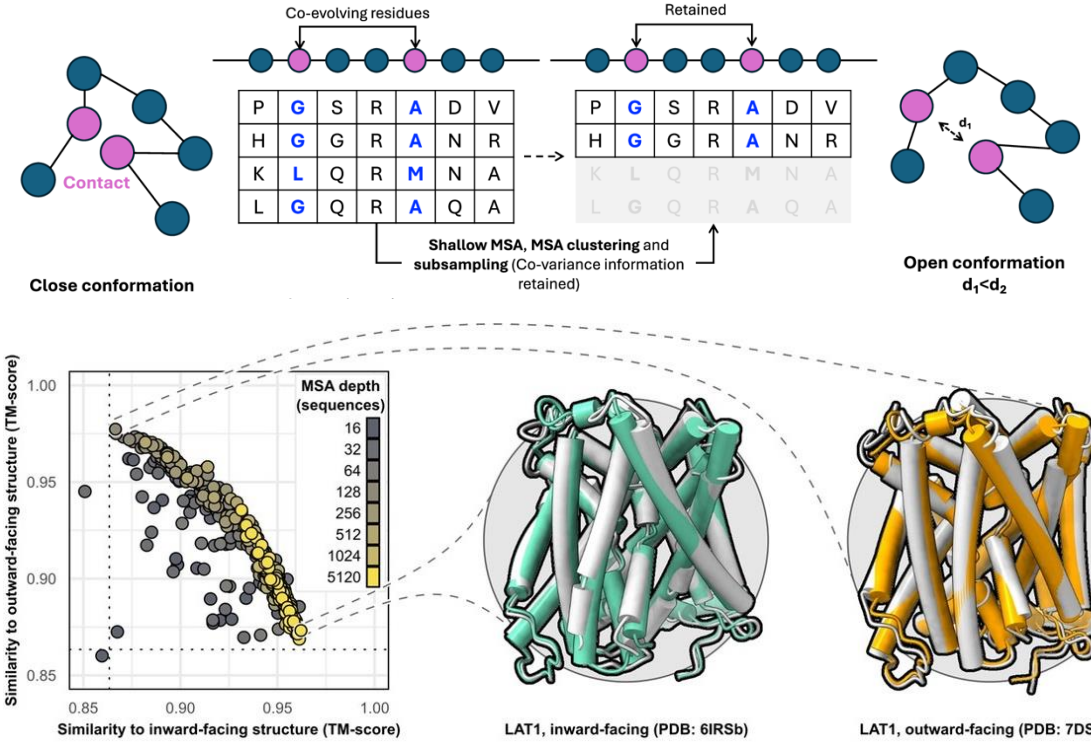
- Sequence-based methods that work
  - MSA masking **(Stein 2022, Kalakoti 2025)**
  - MSA sourcing **(da Silva 2024, Faezov 2023)**
  - **MSA clustering (Wayment-Steele 2023)**
  - MSA subsampling **(del Alamo 2022)**
  - Truncation of autoinhibitory domains **(del Alamo 2022a)**
  - Inclusion of conformationally selective binding partners **(Cummins 2022)**
- Sequence-based augmentations that don't work:
  - Point mutations, including those that work *in vitro* **(Ramsamy 2025, Vani 2023, Zheng 2025**; sometimes many mutations work)



da Silva et al, "High-throughput prediction of protein conformational distributions with subsampled AlphaFold2" Nature Comms 2024
Cummins et al, "AlphaFold accurately predicts distinct conformations based on the oligomeric state of a de novo designed protein" Prot Sci 2022
del Alamo et al, "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
del Alamo et al, "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter" PNAS 2022a
Kalakoti et al, "AFsample2 predicts multiple conformations and ensembles with AlphaFold2" Sci Rep 2025
Ramasamy et al "Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability"bioRxiv 2025
Stein et al , "SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2" Plos Comp Biol 2022
Wayment-Steele et al, "Predicting multiple conformations via sequence clustering and AlphaFold2" Nature 2023
Vani et al "Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE" JCTC 2023
Zheng et al "AlphaFold3 in Drug Discovery: A comprehensive assessment of capabilities, limitations, and applications" bioRxiv 2025

Predicting alternate conformational states with vanilla AlphaFold2

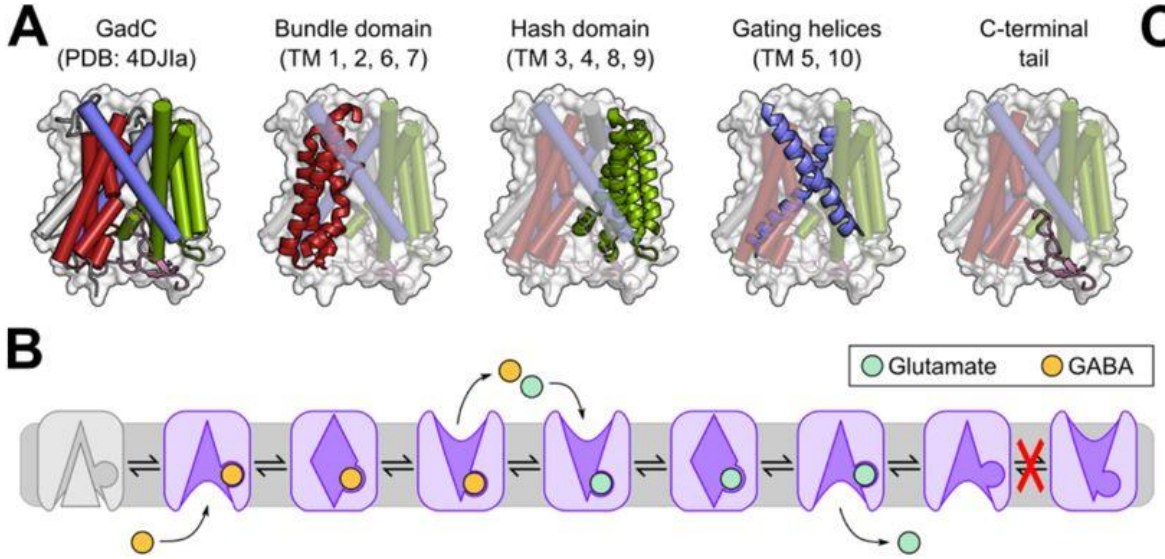
- Sequence-based methods that work
  - MSA masking **(Stein 2022, Kalakoti 2025)**
  - MSA sourcing **(da Silva 2024, Faezov 2023)**
  - MSA clustering **(Wayment-Steele 2023)**
  - **MSA subsampling (del Alamo 2022)**
  - Truncation of autoinhibitory domains **(del Alamo 2022a)**
  - Inclusion of conformationally selective binding partners **(Cummins 2022)**
- Sequence-based augmentations that don't work:
  - Point mutations, including those that work *in vitro* **(Ramsamy 2025, Vani 2023, Zheng 2025**; sometimes many mutations work)



da Silva et al, "High-throughput prediction of protein conformational distributions with subsampled AlphaFold2" Nature Comms 2024
Cummins et al, "AlphaFold accurately predicts distinct conformations based on the oligomeric state of a de novo designed protein" Prot Sci 2022
del Alamo et al, "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
del Alamo et al, "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter" PNAS 2022a
Kalakoti et al, "AFsample2 predicts multiple conformations and ensembles with AlphaFold2" Sci Rep 2025
Ramasamy et al "Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability"bioRxiv 2025
Stein et al , "SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2" Plos Comp Biol 2022
Wayment-Steele et al, "Predicting multiple conformations via sequence clustering and AlphaFold2" Nature 2023
Vani et al "Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE" JCTC 2023
Zheng et al "AlphaFold3 in Drug Discovery: A comprehensive assessment of capabilities, limitations, and applications" bioRxiv 2025

Predicting alternate conformational states with vanilla AlphaFold2

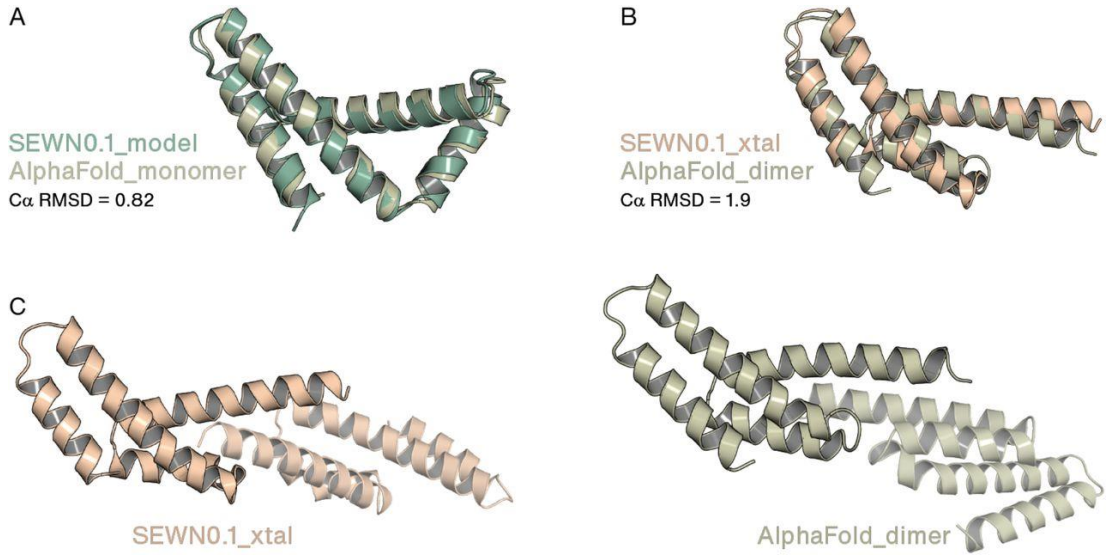
- Sequence-based methods that work
  - MSA masking **(Stein 2022, Kalakoti 2025)**
  - MSA sourcing **(da Silva 2024, Faezov 2023)**
  - MSA clustering **(Wayment-Steele 2023)**
  - MSA subsampling **(del Alamo 2022)**
  - **Truncation of autoinhibitory domains (del Alamo 2022a)**
  - Inclusion of conformationally selective binding partners **(Cummins 2022)**
- Sequence-based augmentations that don't work:
  - Point mutations, including those that work *in vitro* **(Ramsamy 2025, Vani 2023, Zheng 2025**; sometimes many mutations work)



da Silva et al, "High-throughput prediction of protein conformational distributions with subsampled AlphaFold2" Nature Comms 2024
Cummins et al, "AlphaFold accurately predicts distinct conformations based on the oligomeric state of a de novo designed protein" Prot Sci 2022
del Alamo et al, "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
del Alamo et al, "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter" PNAS 2022a
Kalakoti et al, "AFsample2 predicts multiple conformations and ensembles with AlphaFold2" Sci Rep 2025
Ramasamy et al "Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability"bioRxiv 2025
Stein et al , "SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2" Plos Comp Biol 2022
Wayment-Steele et al, "Predicting multiple conformations via sequence clustering and AlphaFold2" Nature 2023
Vani et al "Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE" JCTC 2023
Zheng et al "AlphaFold3 in Drug Discovery: A comprehensive assessment of capabilities, limitations, and applications" bioRxiv 2025

Predicting alternate conformational states with vanilla AlphaFold2

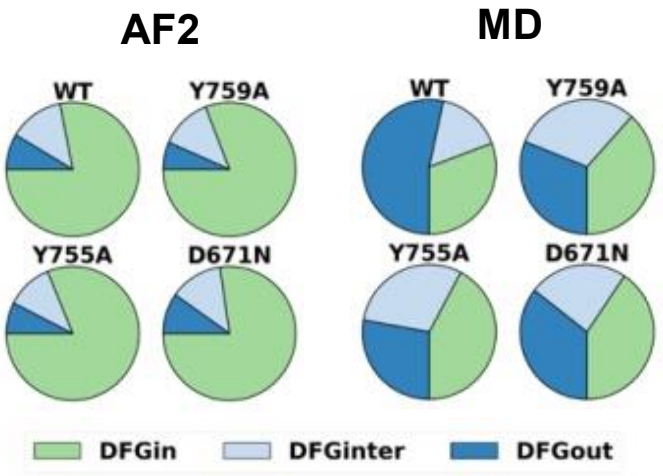
- Sequence-based methods that work
  - MSA masking **(Stein 2022, Kalakoti 2025)**
  - MSA sourcing **(da Silva 2024, Faezov 2023)**
  - MSA clustering **(Wayment-Steele 2023)**
  - MSA subsampling **(del Alamo 2022)**
  - Truncation of autoinhibitory domains **(del Alamo 2022a)**
  - **Inclusion of conformationally selective binding partners (Cummins 2022)**
- Sequence-based augmentations that don't work:
  - Point mutations, including those that work *in vitro* **(Ramsamy 2025, Vani 2023, Zheng 2025**; sometimes many mutations work)

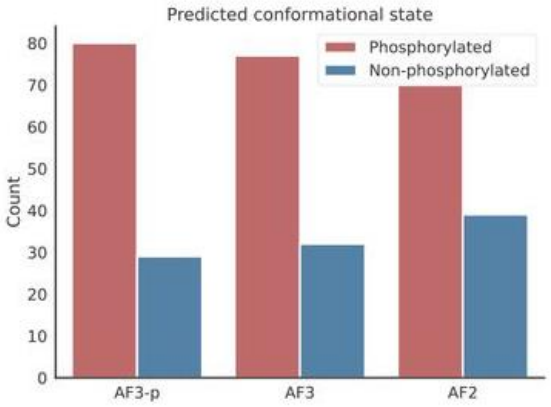


da Silva et al, "High-throughput prediction of protein conformational distributions with subsampled AlphaFold2" Nature Comms 2024
Cummins et al, "AlphaFold accurately predicts distinct conformations based on the oligomeric state of a de novo designed protein" Prot Sci 2022
del Alamo et al, "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
del Alamo et al, "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter" PNAS 2022a
Kalakoti et al, "AFsample2 predicts multiple conformations and ensembles with AlphaFold2" Sci Rep 2025
Ramasamy et al "Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability"bioRxiv 2025
Stein et al , "SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2" Plos Comp Biol 2022
Wayment-Steele et al, "Predicting multiple conformations via sequence clustering and AlphaFold2" Nature 2023
Vani et al "Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE" JCTC 2023
Zheng et al "AlphaFold3 in Drug Discovery: A comprehensive assessment of capabilities, limitations, and applications" bioRxiv 2025

Predicting alternate conformational states with vanilla AlphaFold2

- Sequence-based methods that work
  - MSA masking **(Stein 2022, Kalakoti 2025)**
  - MSA sourcing **(da Silva 2024, Faezov 2023)**
  - MSA clustering **(Wayment-Steele 2023)**
  - MSA subsampling **(del Alamo 2022)**
  - Truncation of autoinhibitory domains **(del Alamo 2022a)**
  - Inclusion of conformationally selective binding partners **(Cummins 2022)**
- **Sequence-based augmentations that don't work:**
  - **Point mutations, including those that work *in vitro* (Ramsamy 2025, Vani 2023, Zheng 2025; sometimes many mutations work)**

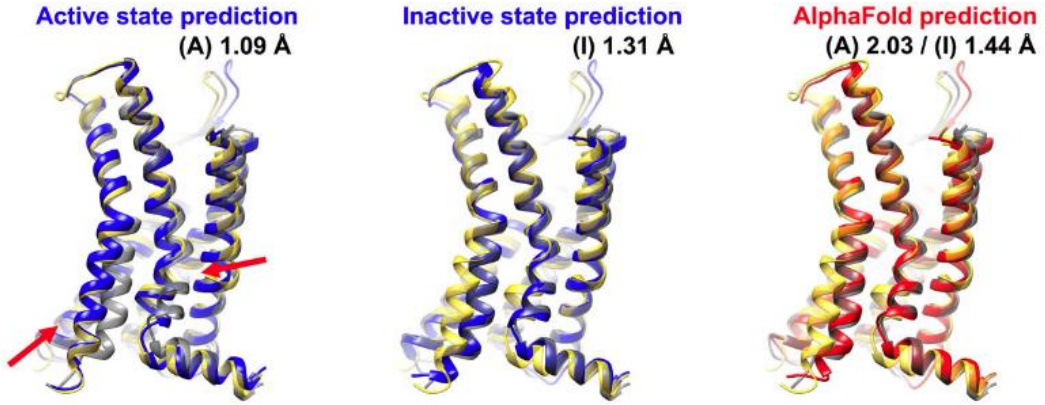


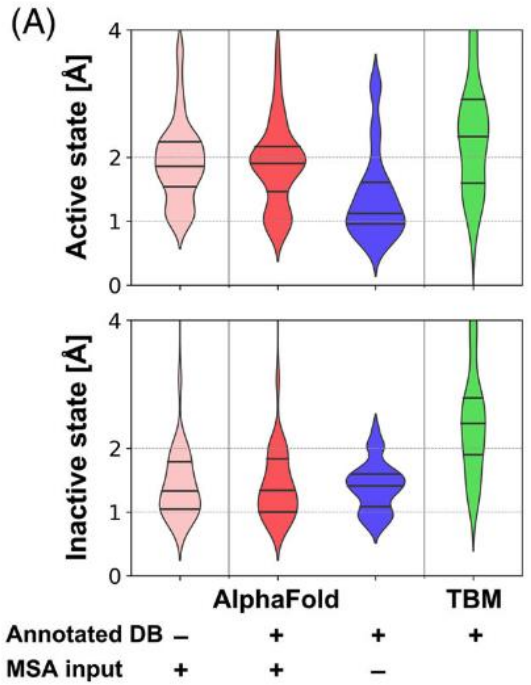


da Silva et al, "High-throughput prediction of protein conformational distributions with subsampled AlphaFold2" Nature Comms 2024
Cummins et al, "AlphaFold accurately predicts distinct conformations based on the oligomeric state of a de novo designed protein" Prot Sci 2022
del Alamo et al, "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
del Alamo et al, "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter" PNAS 2022a
Kalakoti et al, "AFsample2 predicts multiple conformations and ensembles with AlphaFold2" Sci Rep 2025
Ramasamy et al "Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability"bioRxiv 2025
Stein et al , "SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2" Plos Comp Biol 2022
Wayment-Steele et al, "Predicting multiple conformations via sequence clustering and AlphaFold2" Nature 2023
Vani et al "Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE" JCTC 2023
Zheng et al "AlphaFold3 in Drug Discovery: A comprehensive assessment of capabilities, limitations, and applications" bioRxiv 2025

Predicting alternate conformational states with vanilla AlphaFold2

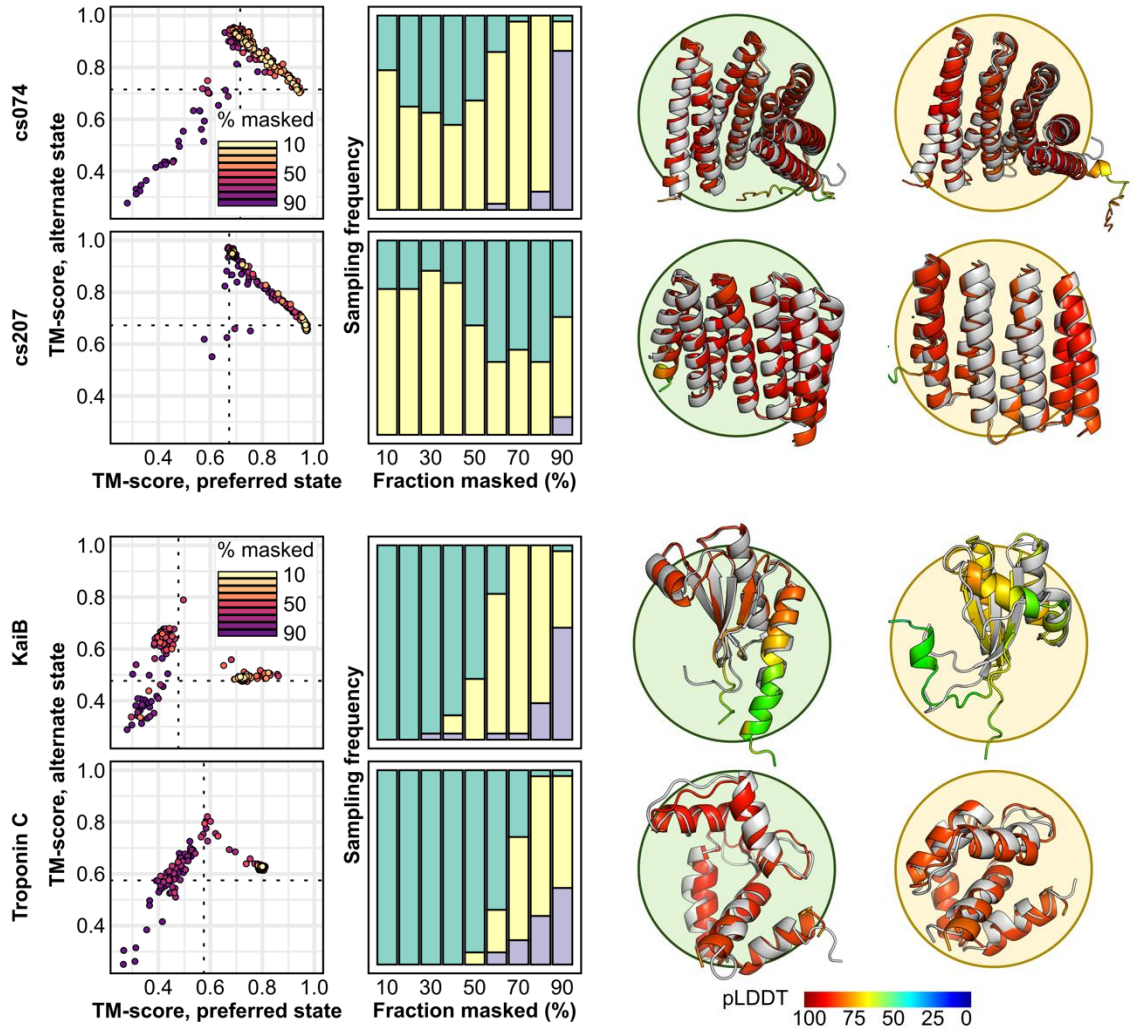
- Structure-based approaches that work:
  - **Template curation** (del Alamo 2022, Faezov 2023, Heo 2022)
  - **Caveat:** only two AF2 monomer models can use templates (all multimer models can use templates)





del Alamo et al, "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
Faezov & Dunbrack "AlphaFold2 models of the active form of all 437 catalytically competent human protein kinase domains" bioRxiv 2023
Heo & Feig "Multi-state modeling of G-protein coupled receptors at experimental accuracy" Proteins 2022

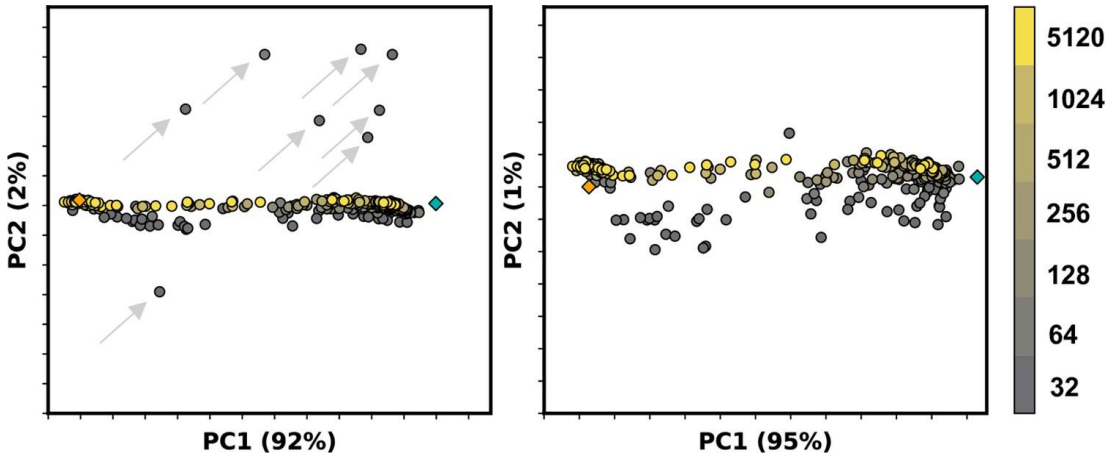
Strategies that work with other NNs

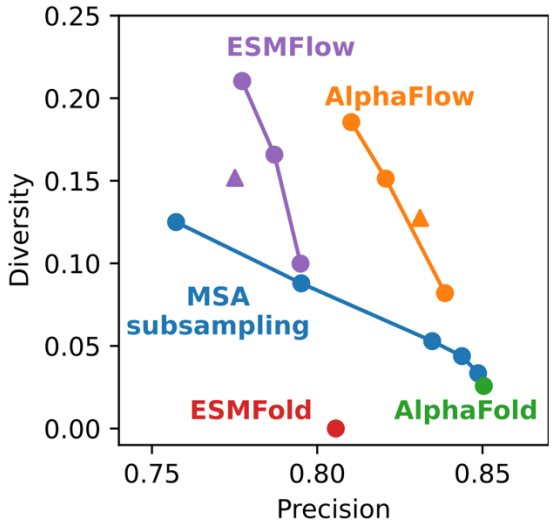


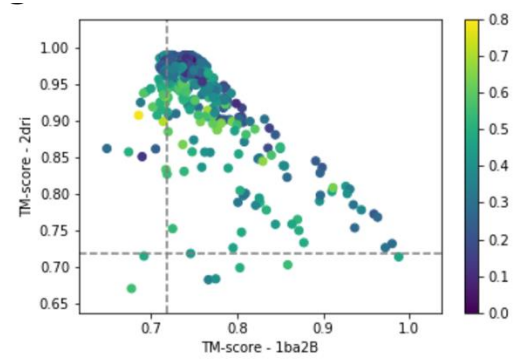
del Alamo et al, "Conformational sampling and interpolation using language-based protein folding neural networks" bioRxiv 2023

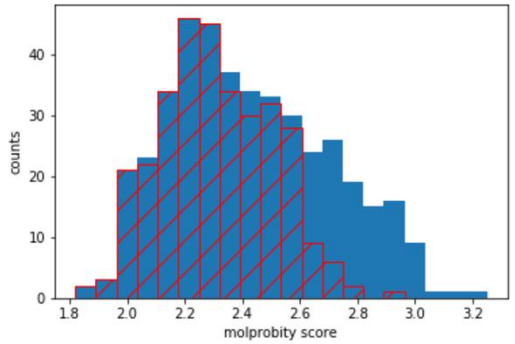
How to filter bad models

- Iterative PCA: Look for ensembles where predicted motions of models are consistent with one another
- Model quality check: Use MolProbity score to check for residue-scale errors
- Manually

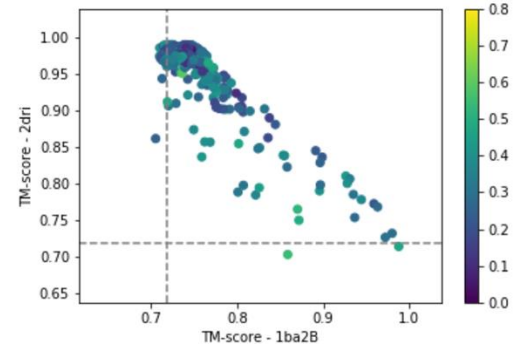




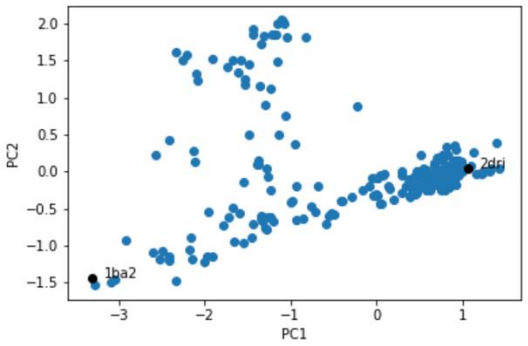




E



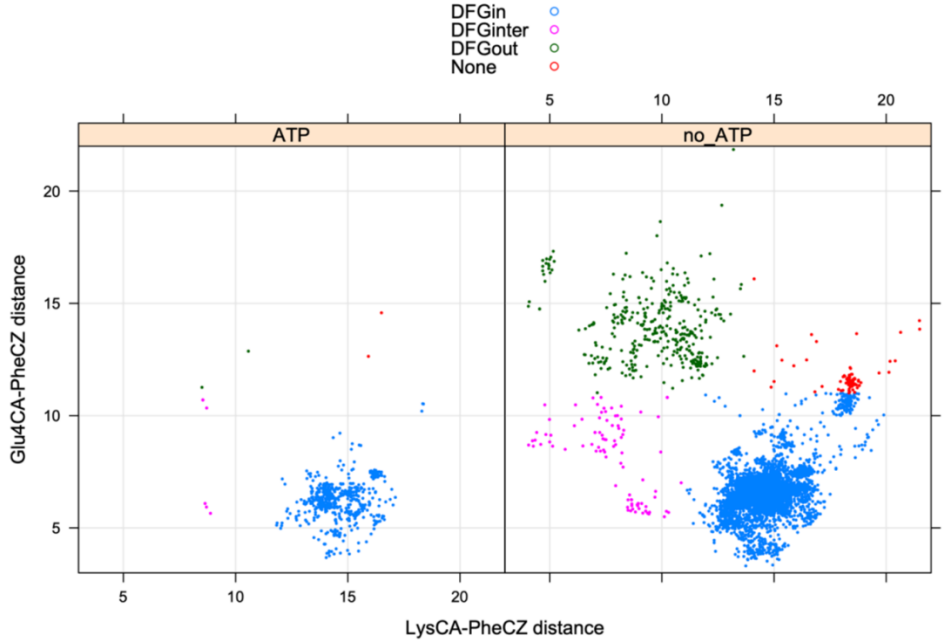
F

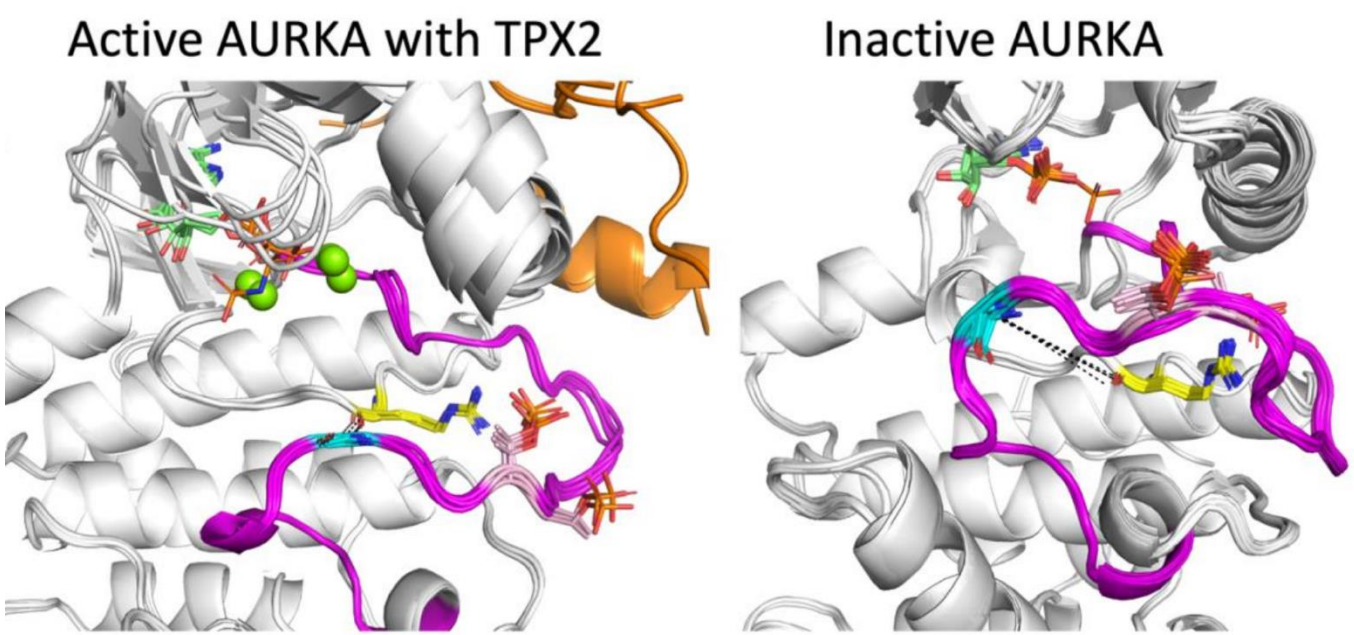


del Alamo et al "Sampling alternative conformational states of transporters and receptors with AlphaFold2" eLife 2022
Jing et al "AlphaFold meets Flow Matching for generating protein ensembles" ICLR 2024
Stein & Mchaourab "SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2" Plos Comp Biol 2022

Example: prediction of all 437 kinases in the human proteome

- Faezov & Dunbrack predicted the active states of all kinases using MSA sourcing, MSA subsampling, and custom template DBs
- *“No one set of inputs (MSA source, MSA depth, template database) produces active models of all 437 catalytic targets.”*
- For two cases, AF2 models of active kinases were required as templates





Faezov & Dunbrack “AlphaFold2 models of the active form of all 437 catalytically competent human protein kinase domains” bioRxiv 2023

Example: modeling alternative conformations of a transporter

Truncate sequence
Cut autoinhibitory domain

Setup template DB
HH-suite

Reduce MSA depth
max_msa_clusters = 8
max_extra_msa = 16

Generate models
monomer_ptm NN 1: 25 models
monomer_ptm NN 2: 25 models

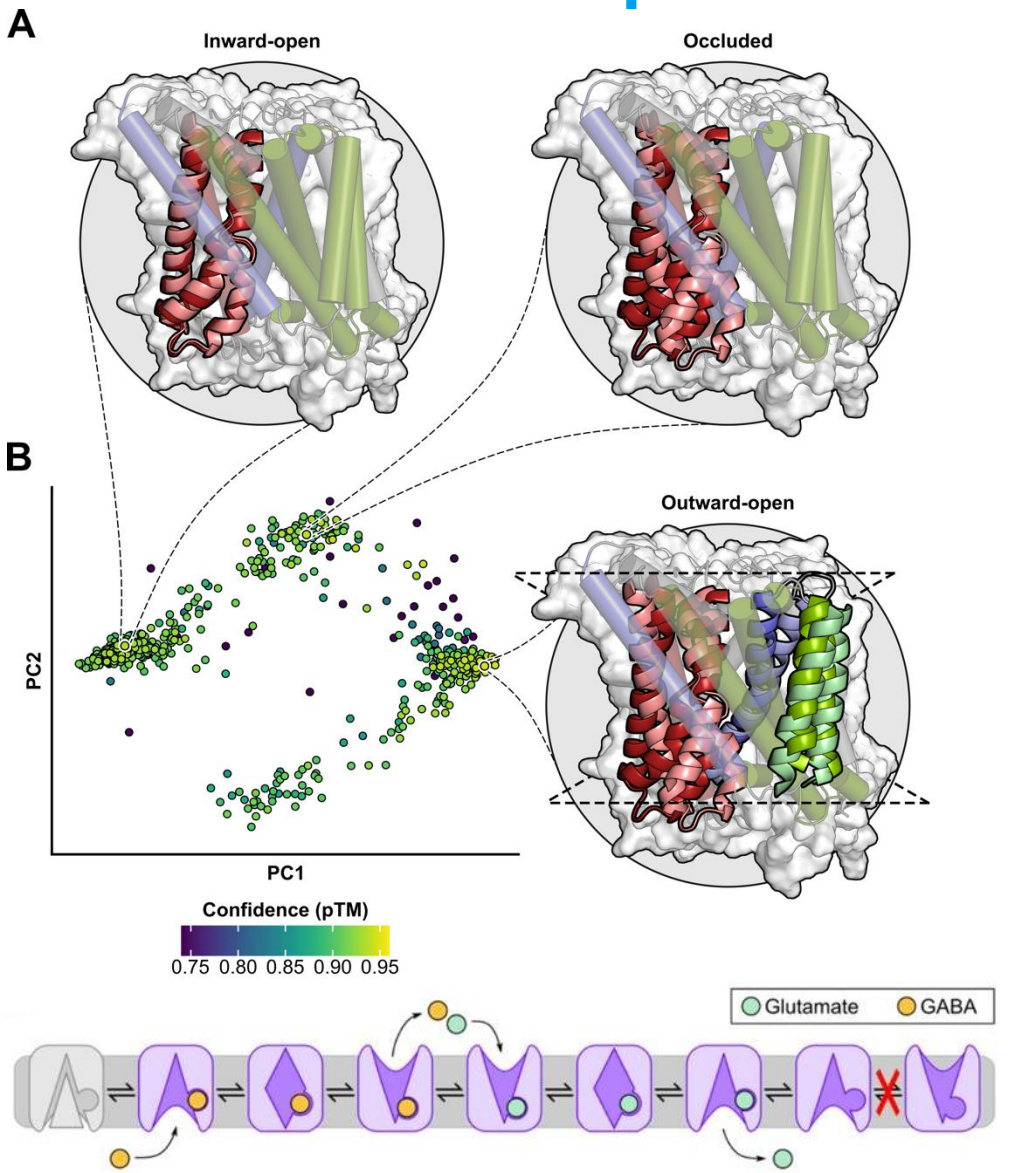
Increment MSA depth
max_msa_cluster += 1 (until >16)
max_extra_msa += 2 (until >32)

Align all models
TM-Align

Dimensionality reduction
PCA using helical CA atoms
SciKit-Learn

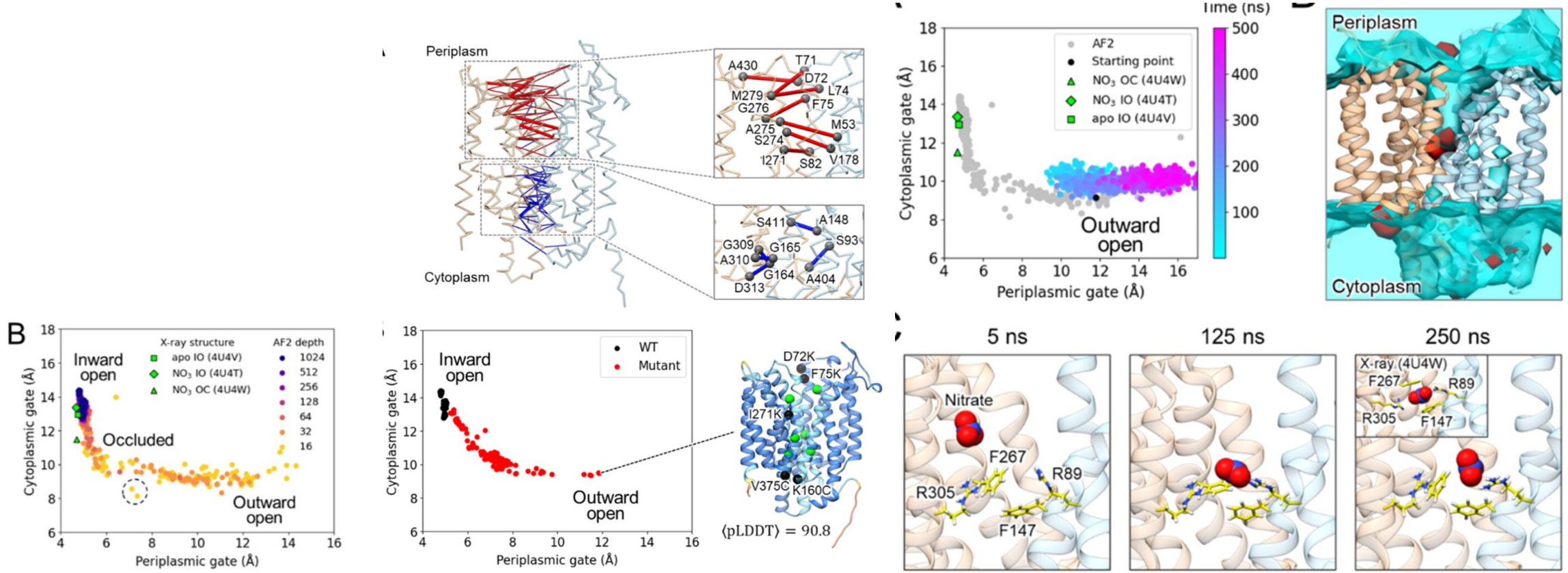
Remove PCA outliers

Select representatives & design experiments
Models with best pTM in each cluster



del Alamo et al, "Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter" PNAS 2022a

Example: modeling alternate conformation of the transporter NarK



Ohnuki & Okazaki, "Integration of alphafold with molecular dynamics for efficient conformational sampling of transporter protein nark" JPC B 2024

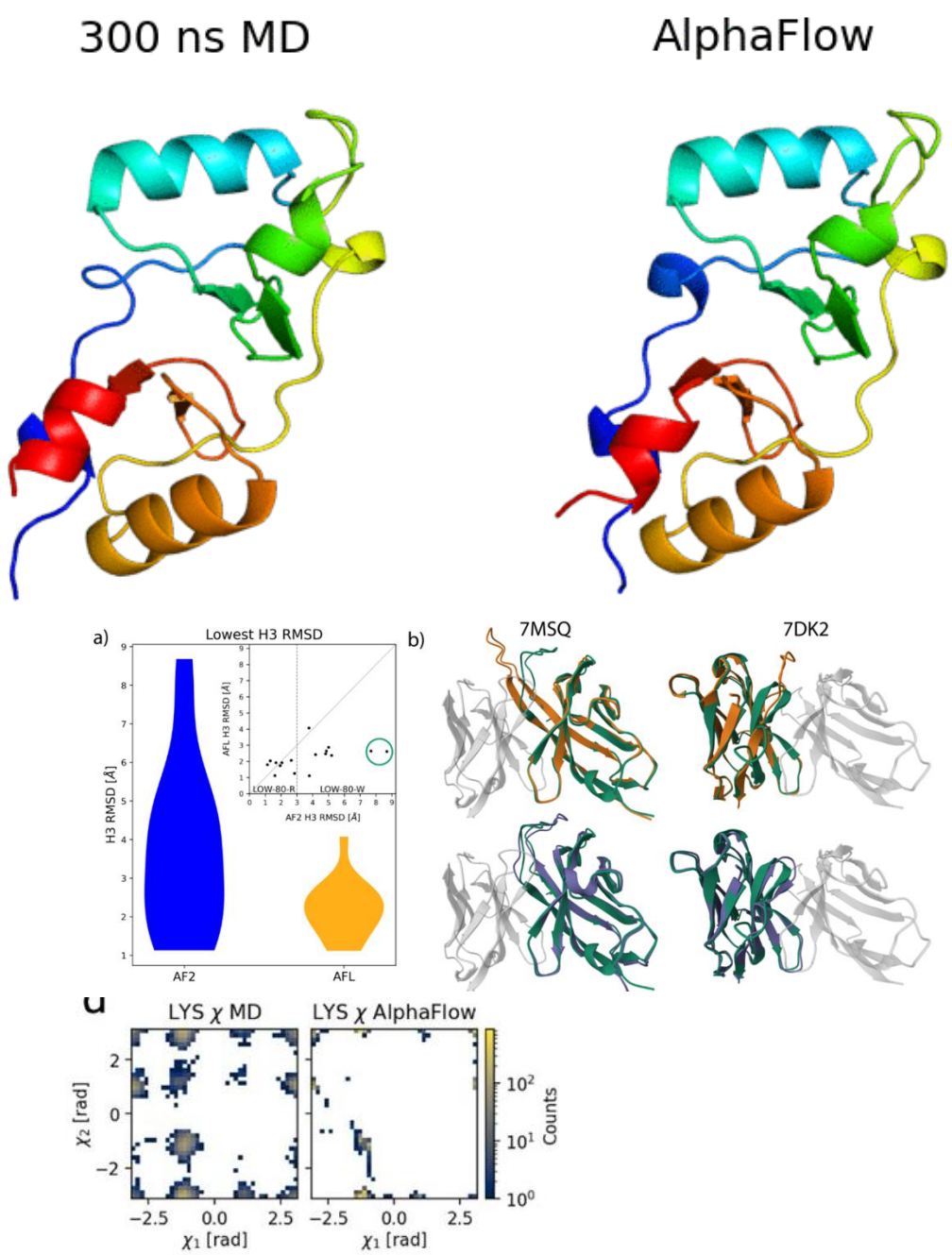
Custom neural networks for obtaining alternate states

- What makes these special?
  - Custom training data (AlphaFlow, DiG, aSAMt) such as family-specific data **(Xu 2025, Mansoor 2024)**
  - Curated or reweighted training sets (Cfold, UFConf)
  - Distinct architectures and training schedules (AlphaFlow, BioEmu, DiG, RF-VAE)
  - Noised inputs (Str2str, diffusion- and flow matching-based methods)
- In general, these are:
  - Worse at single-state modeling than dedicated protein folding NNs
  - A bit better at recovering multiple relevant states (although not always)
  - Much better at modeling “Brownian motion”

Mansoor et al “Protein Ensemble Generation Through Variational Autoencoder Latent Space Sampling” JCTC 2024
Xu et al “Generating Multistate Conformations of P-type ATPases with a Conditional Diffusion Model” JCIM 2024

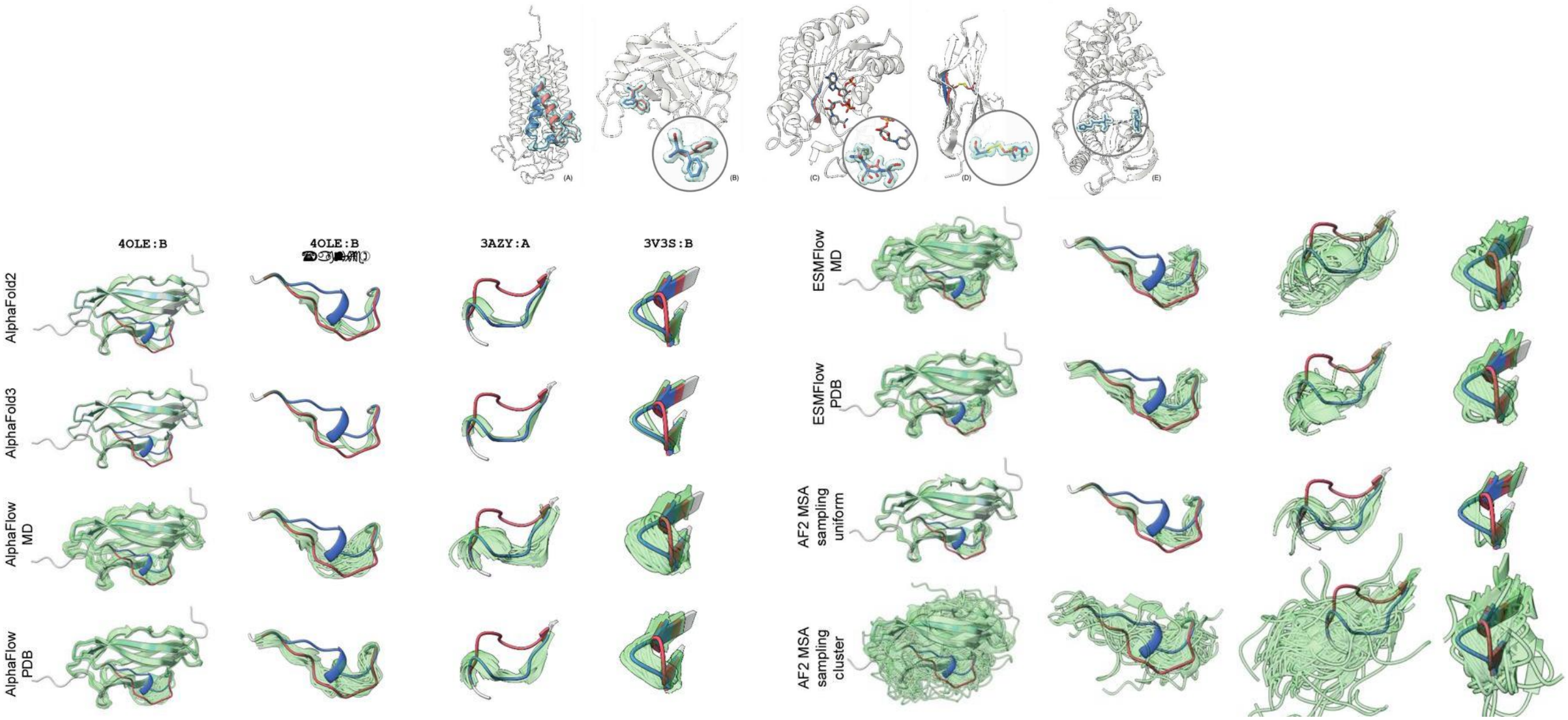
AlphaFlow and ESMFlow

- Fine-tuned AlphaFold2/ESMFold for prediction of different conformations – trained on short (many <1 μs MD simulations; **Jing 2024**)
- Anecdotally, doesn't work for all proteins, such as fold-switching proteins
- Still requires MSA for AlphaFold
- In a separate benchmark, AlphaFlow but not AF2 always generated antibody CDRH3 conformations within 3.0 Å RMSD of ground truth (note: monomer only; **Giulini 2025**)
- Limited side chain sampling **(Janson 2025)**



Giulini et al "Improved structural modelling of antibodies and their complexes with clustered diffusion ensembles" bioRxiv 2025
Janson et al "Deep generative modeling of temperature-dependent structural ensembles of proteins" bioRxiv 2025
Jing et al "AlphaFold meets Flow Matching for generating protein ensembles" ICLR 2024

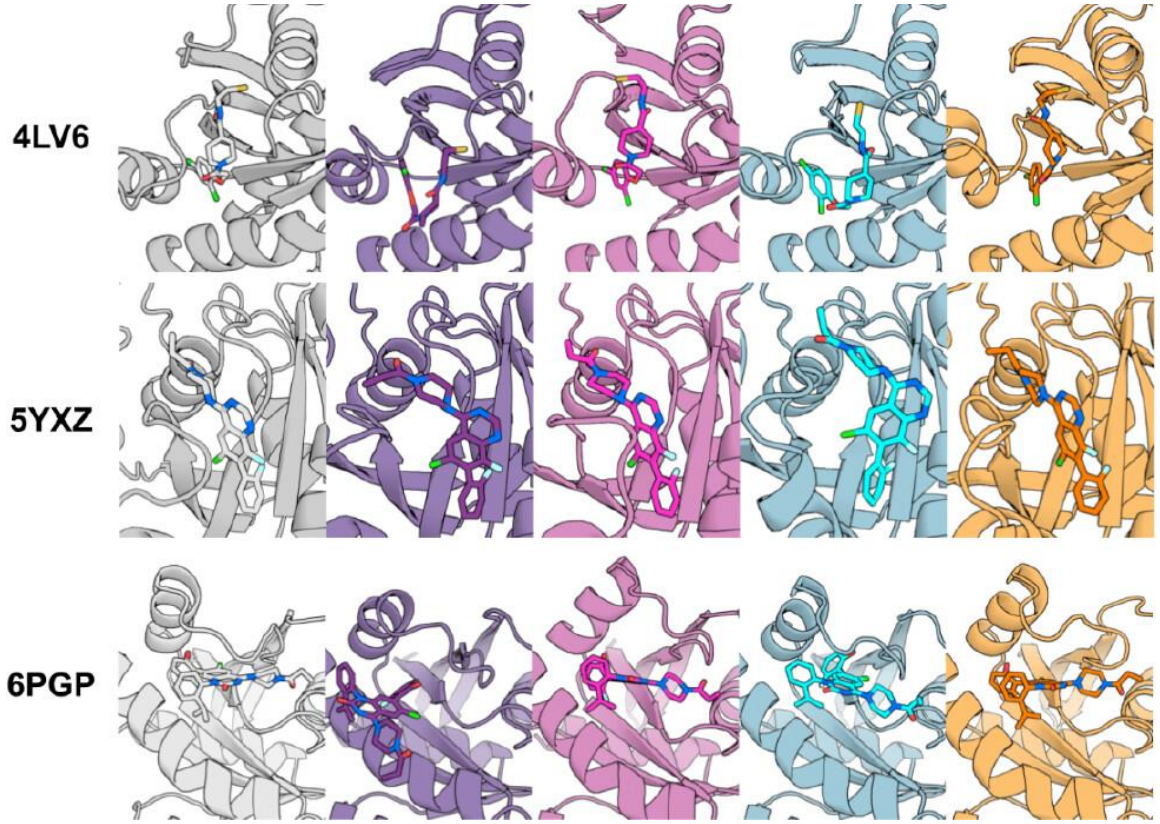
Examples of these methods failing to predict alternate states

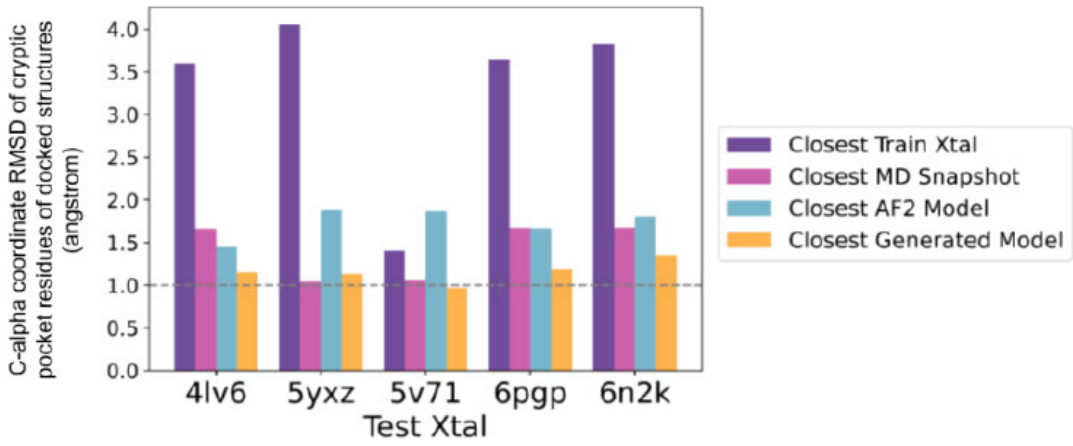


Rosenberg et al "Seeing Double: Molecular dynamics simulations reveal the stability of certain alternate protein conformations in crystal structures" bioRxiv 2024

Fine-tuned RosettaFold+VAE for exploring Ras conformations

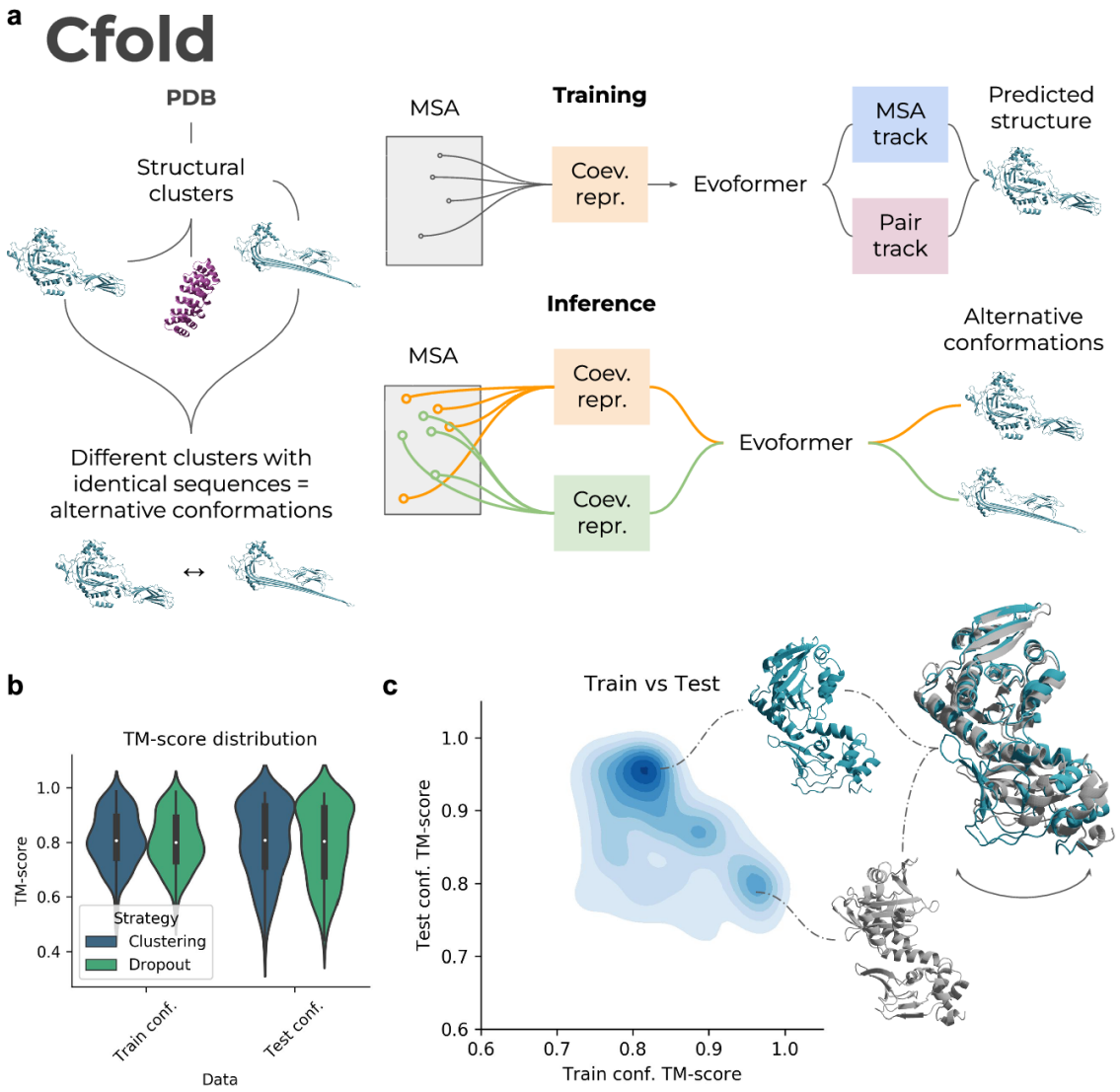
- Variational autoencoder added to RosettaFold, fine-tuned on many Ras structures and MD snapshots **(Mansoor 2024)**
- Tested on held-out crystal structures

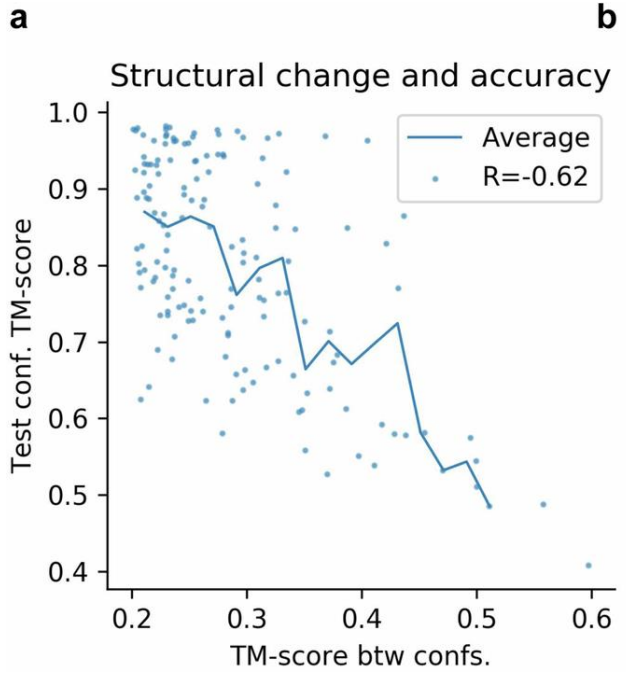


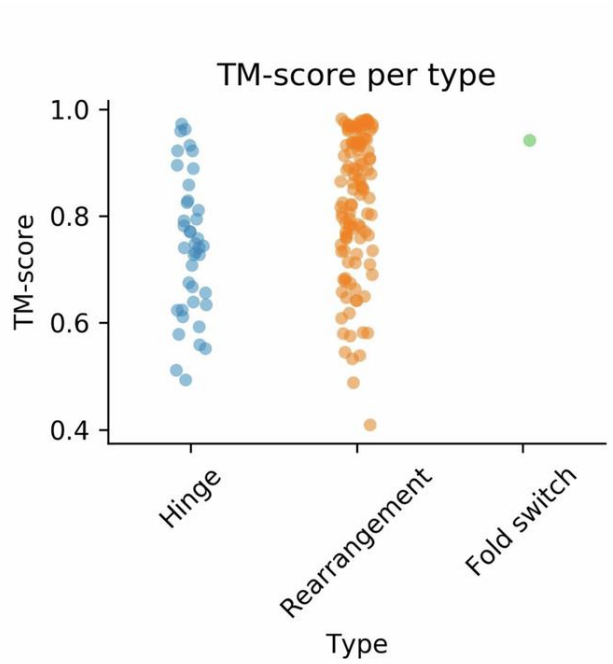


Mansoor et al "Protein Ensemble Generation Through Variational Autoencoder Latent Space Sampling" JCTC 2024

Cfold: retrained AF2-like NN with a bespoke training set



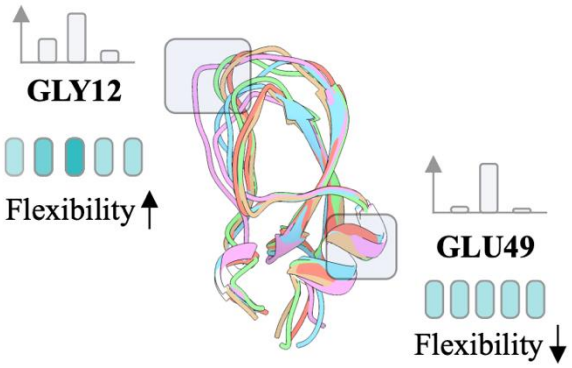




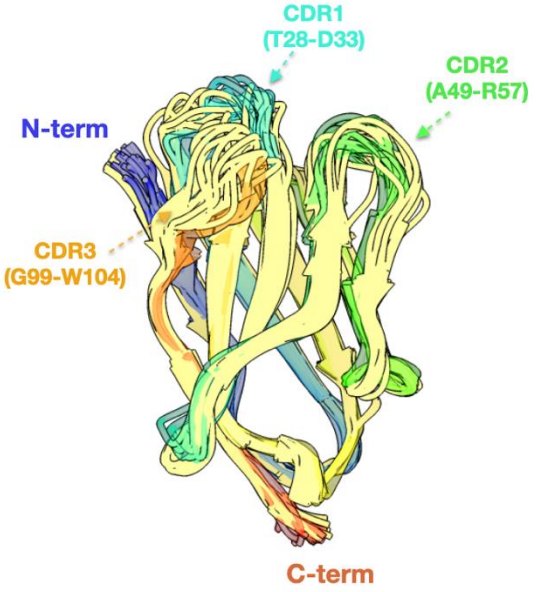
Bryant & Noé "Structure prediction of alternative protein conformations" Nature Comms 2024

ESMDiff – Fine-tuned ESM3 for tokenized structure generation

- Relies on structure tokenization, a recent paradigm for encoding structure as a discrete vocabulary via a "codebook" (example: Foldseek)
- Fine-tuned from ESM3, a structure-aware language model
- Promising early results comparable to AlphaFlow



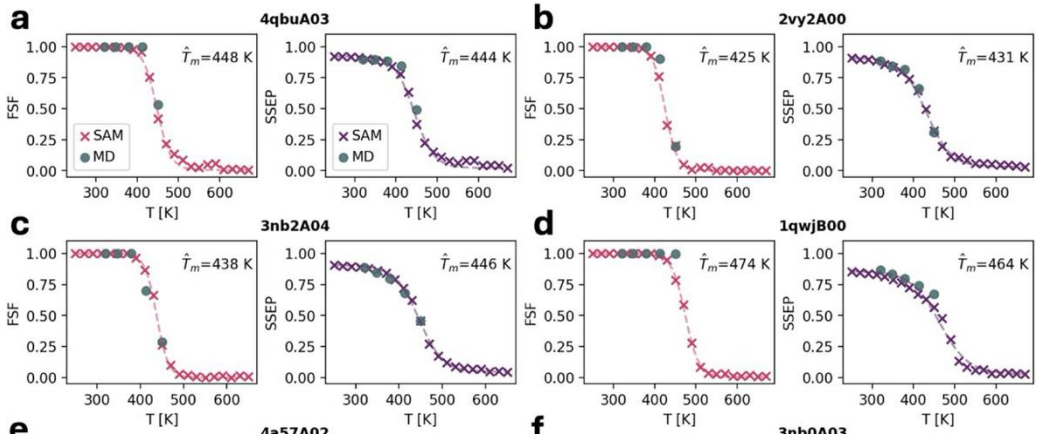
Method		Apo/holo			Fold-switch		
		ResFlex r (gl.)	**ResFlex r (pt.)**	**TM-ens**	**ResFlex r (gl.)**	**ResFlex r (pt.)**	**TM-ens**
MSA-based	MSA-Subs.	0.398	0.404 / 0.371	0.856 / 0.894	0.350	0.320 / 0.303	0.714 / 0.765
	AlphaFlow	**0.455**	**0.527 / 0.527**	**0.864 / 0.893**	0.385	**0.384 / 0.376**	**0.730 / 0.788**
Seq-based	Eigenfold	0.126	0.407 / 0.401	0.830 / 0.870	0.225	0.279 / 0.255	0.614 / 0.653
	Str2Str (PF)	0.174	0.326 / 0.307	0.731 / 0.728	0.161	0.246 / 0.233	0.615 / 0.644
	Str2Str (SDE)	0.148	0.349 / 0.340	0.659 / 0.681	0.111	0.224 / 0.220	0.521 / 0.545
	ESMFlow	0.416	0.496 / 0.522	0.856 / 0.893	0.269	0.345 / 0.329	0.700 / 0.755
SLM	S-T5	0.097	0.144 / 0.166	0.726 / 0.787	0.313	0.135 / 0.099	0.437 / 0.392
	S-GPT	0.112	0.134 / 0.112	0.571 / 0.562	0.207	0.075 / 0.078	0.349 / 0.300
	ESM3 (zero shot)	0.312	0.473 / 0.466	0.839 / 0.876	0.388	0.323 / 0.320	0.627 / 0.717
	ESMDiff (ID)	0.424	0.502 / 0.517	0.851 / 0.883	0.391	0.328 / 0.346	0.660 / 0.720
	ESMDiff (DDPM)	0.420	0.489 / 0.515	0.838 / 0.877	**0.402**	0.341 / 0.288	0.626 / 0.685

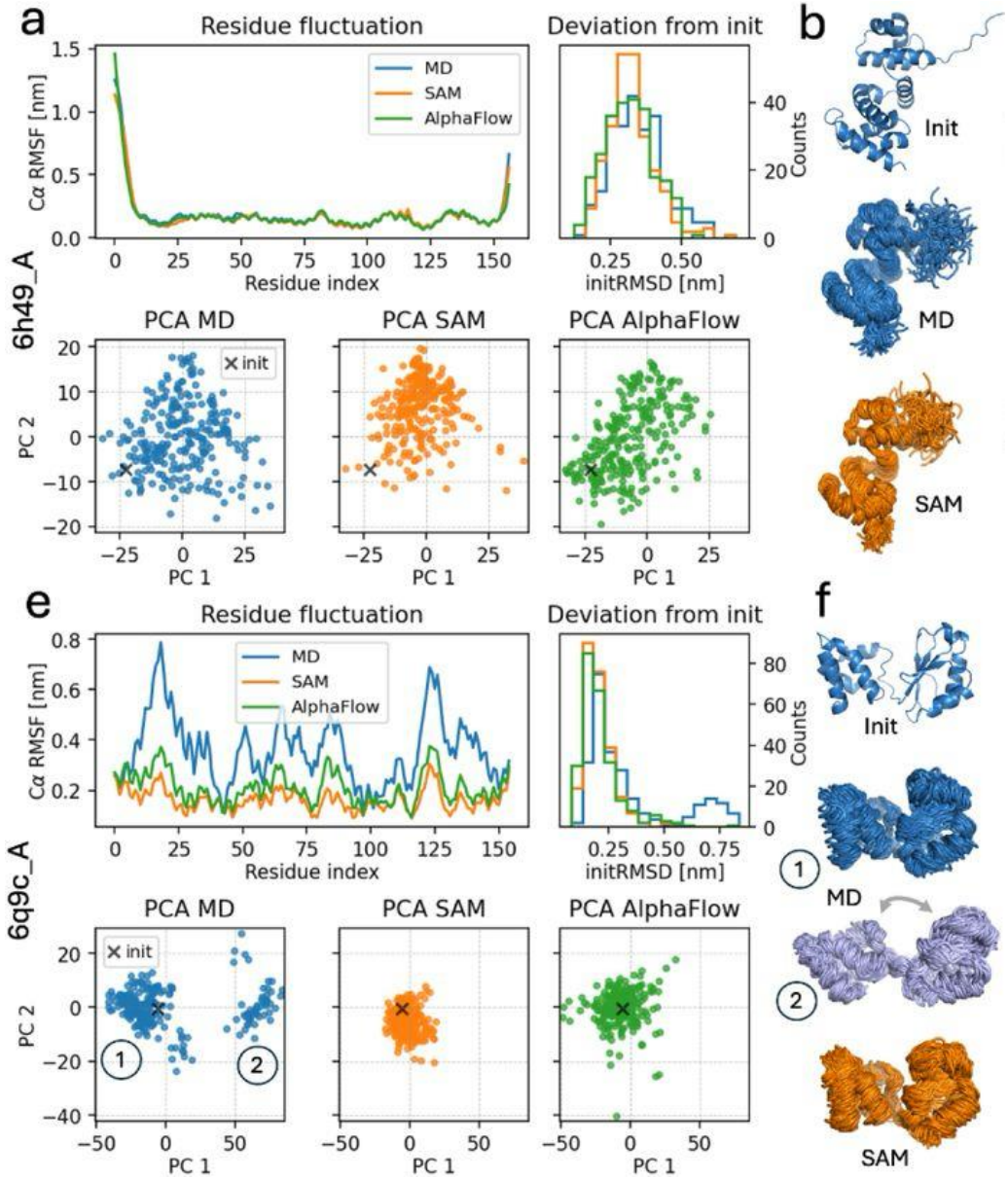


Lu et al "Structure Language Models for Protein Conformation Generation" ICLR 2025

aSAMt

- Also trained on bulk MD simulations
- Has a tunable “temperature” parameter that predicts unfolded state when raised
- Some evidence that it generalizes dG, and captures differences between closely related proteins
- However, not always to sample distinct, dissimilar conformations

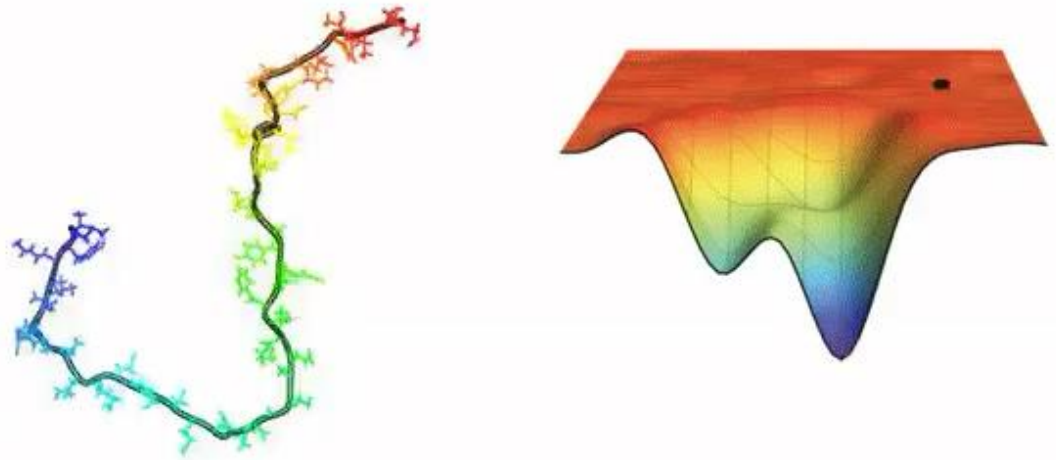


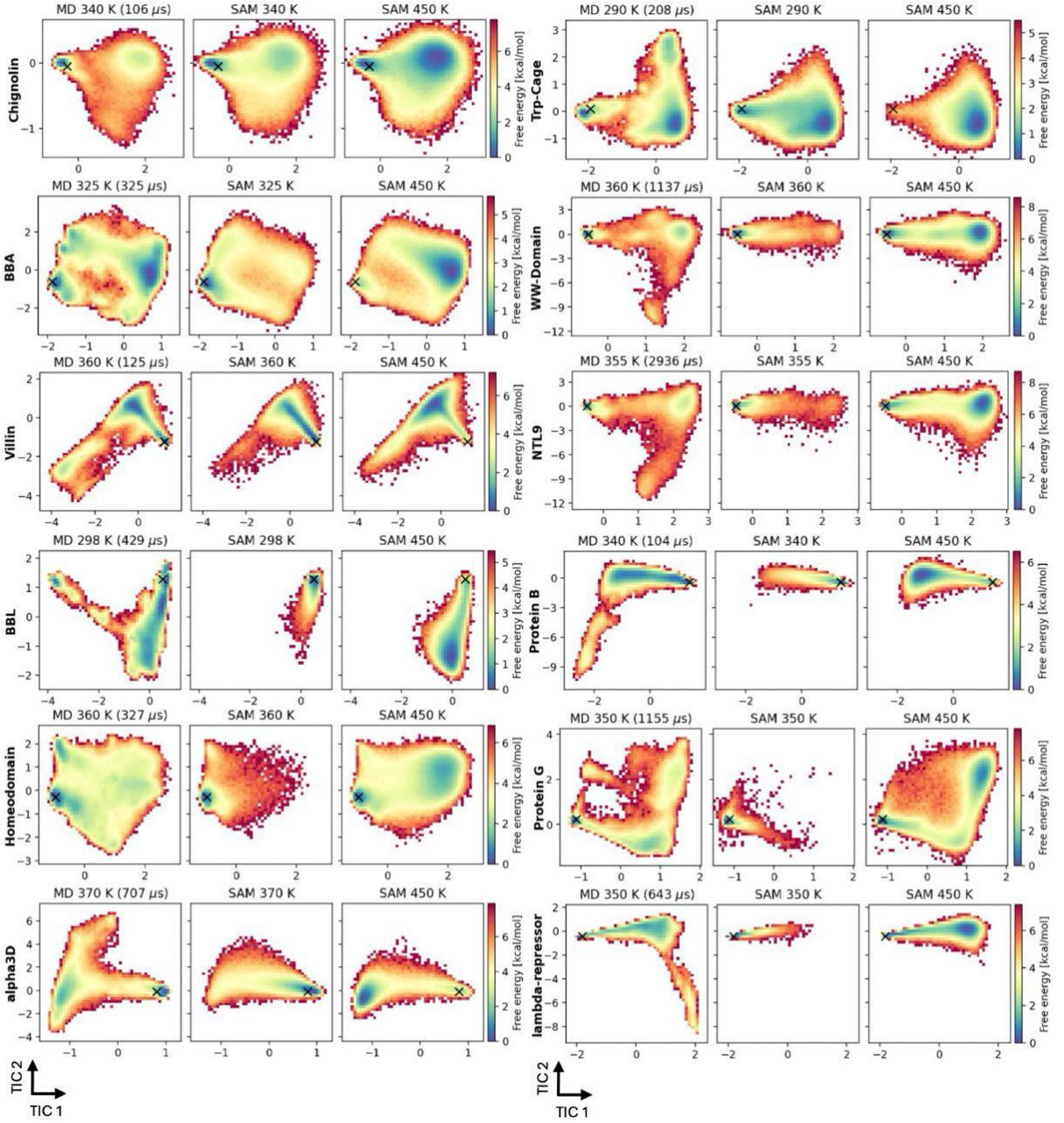


Janson et al “Deep generative modeling of temperature-dependent structural ensembles of proteins” bioRxiv 2025

aSAMt (continued)

- Partial generalization to slow dynamics in fast-folding proteins



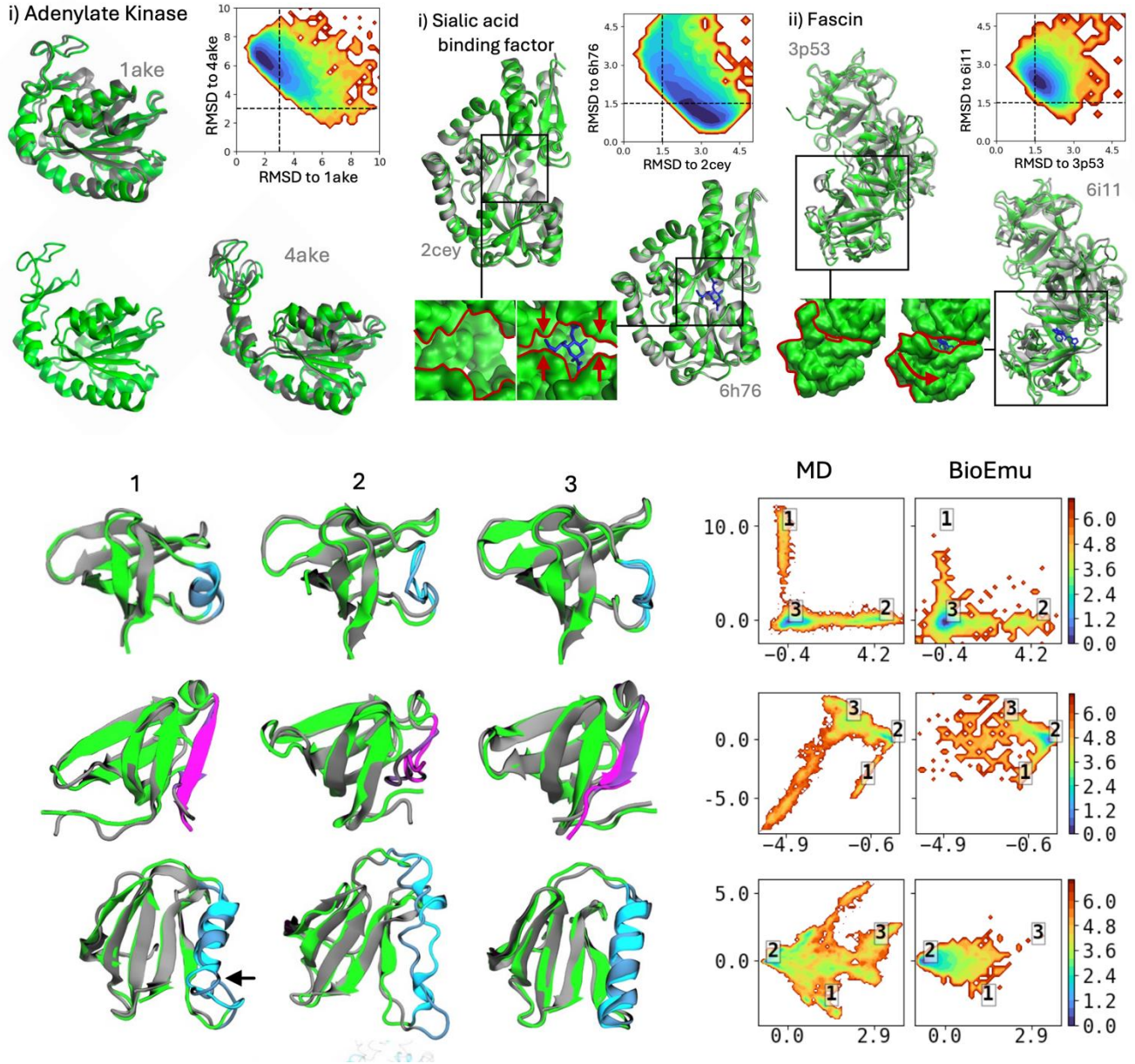


GIF source unknown

Janson et al "Deep generative modeling of temperature-dependent structural ensembles of proteins" bioRxiv 2025

BioEmu

- Among the most promising of the bespoke NNs for conformational modeling
- Independent outputs can sample distinct conformations; but not tested on large proteins
- Is able to model energy landscapes of simple proteins to ~1 kcal/mol @ 300K
- How? Bespoke fine-tuning method trained on 200 milliseconds of MD + 750,000 stability datapoints
- Caveats:
  - Folding/unfolding data used for training happen at different timescales than many conformational interconversion processes

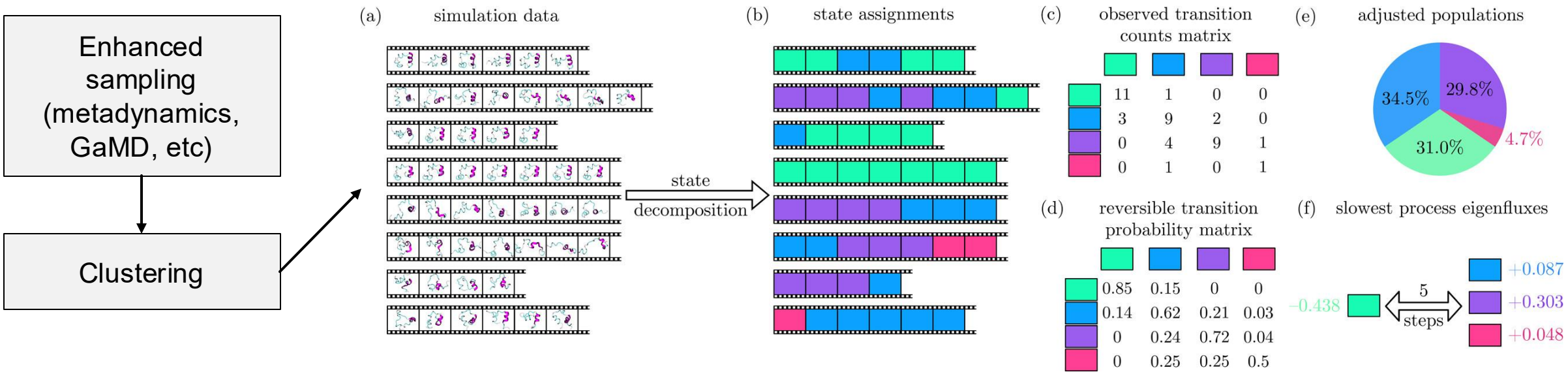


Lewis et al "Scalable emulation of protein equilibrium ensembles with generative deep learning" bioRxiv 2024

Why don’t protein folding neural networks predict dynamics?

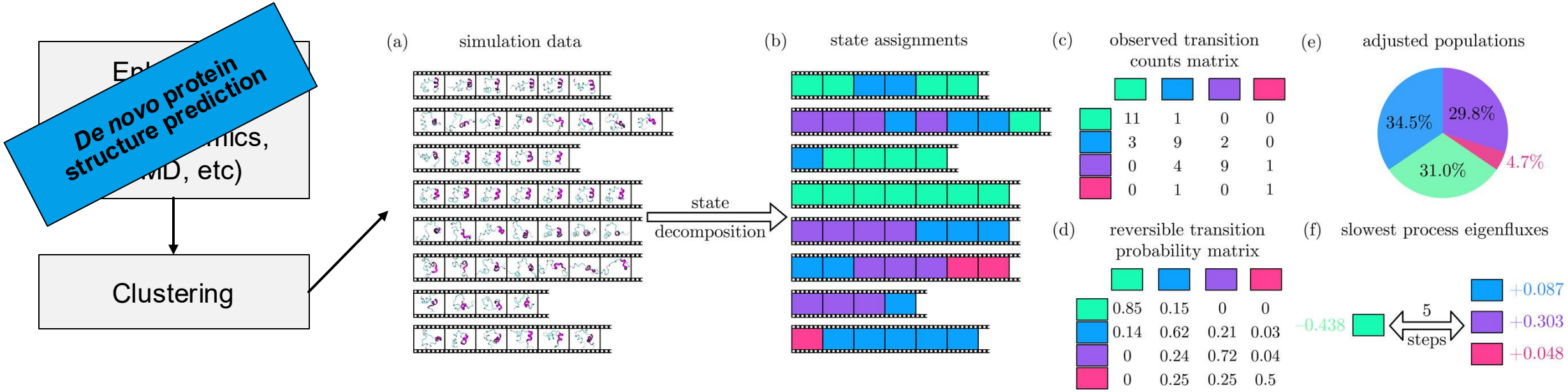
- The problem is difficult:
  - Prediction of protein dynamics is fundamentally different from prediction of protein structure
  - Structure prediction training objective is clear and well-suited to learning from high-quality PDB data **(Bronstein 2024)**
  - Dynamics are less conserved than structure **(Tokuriki 2009)**
- The training data are ill-suited:
  - Dynamics data are ambiguous, sensitive to temperature and (for MD) starting conditions
  - Experimental dynamics data are ambiguous, low-resolution, coarse (SAXS, NMR, FRET)
  - MD simulations are sparse, incomplete, potentially inaccurate
  - Combinations data from different sources introduces noise

An example MD workflow for free energy calculations



Husic & Pande, "Markov State Models: From an Art to a Science" JACS 2018
Noé & Rosta, "Markov Models of Molecular Kinetics" JCP 2019

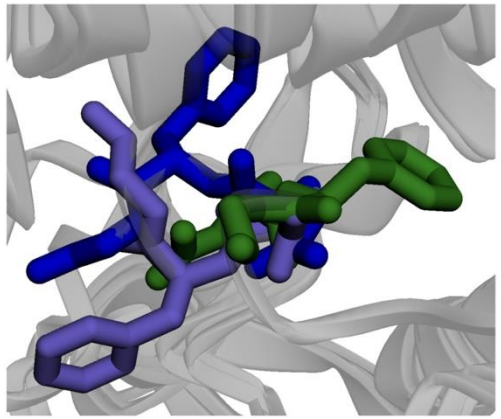
An example MD workflow for free energy calculations

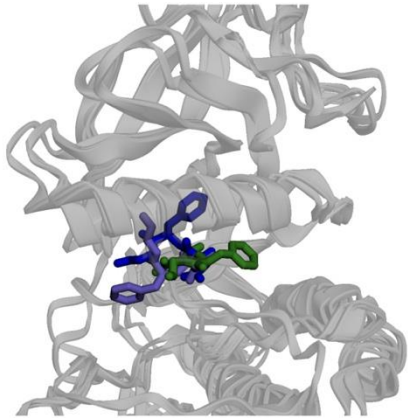


Husic & Pande, "Markov State Models: From an Art to a Science" JACS 2018
Noé & Rosta, "Markov Models of Molecular Kinetics" JCP 2019

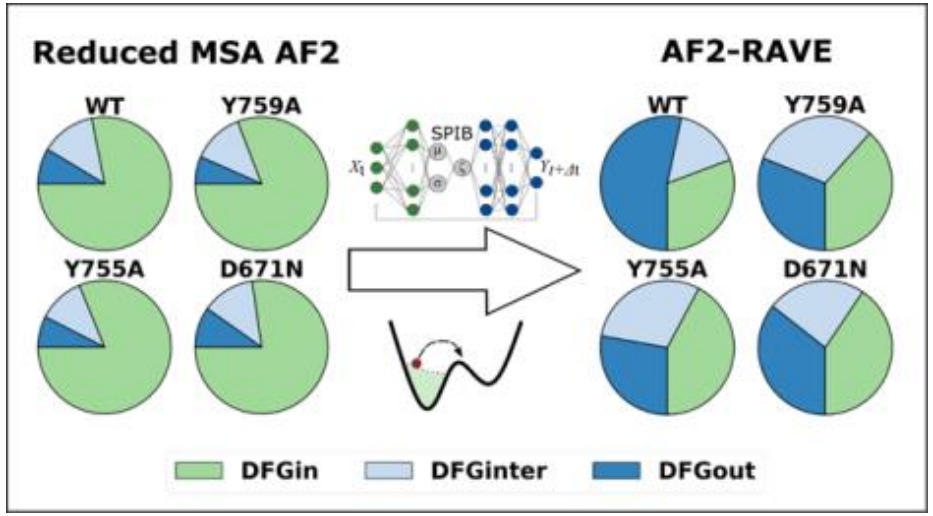
Replacing enhanced sampling with ensemble prediction

- Several examples involving AlphaFold2 + RAVE **(Vani 2023, Vani 2023a)**
  - **RAVE is a protocol for converting MD trajectories and snapshots into Boltzmann-weighted state assignments**
  - Kinases with/without ligand **(Gu 2024)**, GPCRs **(Vani 2023)**, SARS-CoV-2 RBD **(Teng 2025)**
- Collective variables from short MD trajectories of AF2 models match those of much longer simulations (>100 μs; **Teng 2025 & Vats 2025**)
- Can lead to the observation of rare events, such as cryptic pocket opening **(Meller 2023, Vats 2025)**, conformational interconversion **(Bhakat 2025)**





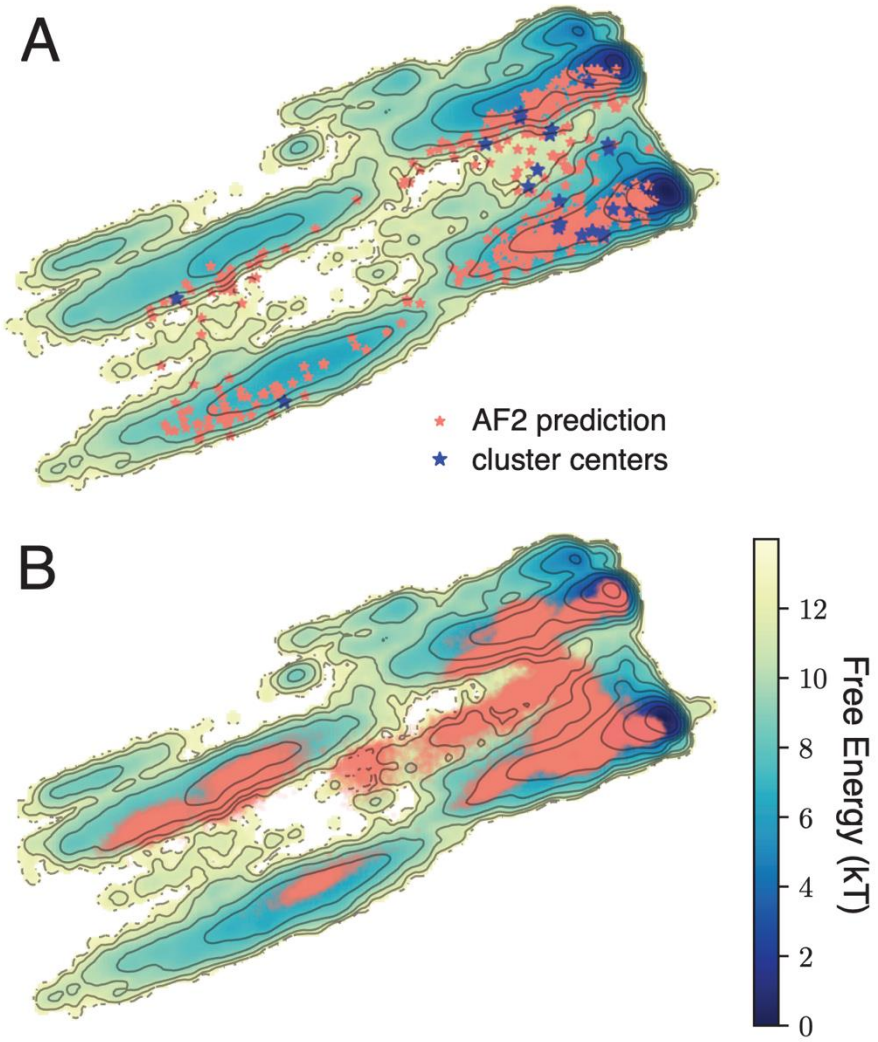
DFG-IN DFG-INTER DFG-OUT



Bhakat et al, "Generalizable Protein Dynamics in Serine-Threonine Kinases: Physics is the key" bioRxiv 2025
Gu et al, "Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE" eLife 2024
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Teng et al, "AlphaFold2-RAVE: Protein Ensemble Generation with Physics-Based Sampling" ChemRxiv 2025
Vani et al, "Exploring kinase Asp-Phe-Gly (DFG) loop conformational stability with AlphaFold2-RAVE" JCIM 2023
Vani et al, "AlphaFold2-RAVE: From Sequence to Boltzmann Ranking" JCTC 2023
Vats et al, "AlphaFold-SFA: Accelerated sampling of cryptic pocket opening, protein-ligand binding and allostery by AlphaFold, slow feature analysis and metadynamics" Plos One 2025

Replacing enhanced sampling with ensemble prediction

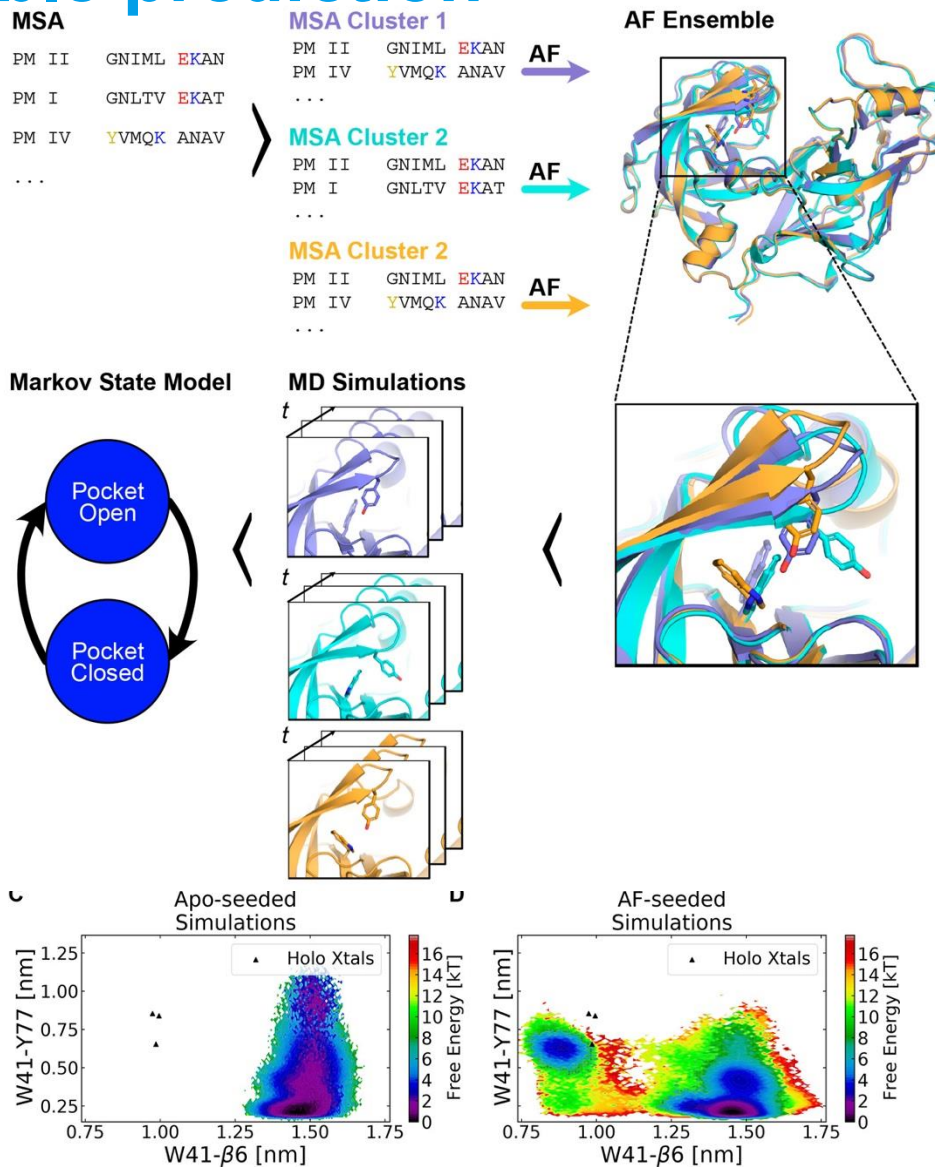
- Several examples involving AlphaFold2 + RAVE **(Vani 2023, Vani 2023a)**
  - RAVE is a protocol for converting MD trajectories and snapshots into Boltzmann-weighted state assignments
  - Kinases with/without ligand **(Gu 2024)**, GPCRs **(Vani 2023)**, SARS-CoV-2 RBD **(Teng 2025)**
- **Collective variables from short MD trajectories of AF2 models match those of much longer simulations (>100 μs; Teng 2025 & Vats 2025)**
- Can lead to the observation of rare events, such as cryptic pocket opening **(Meller 2023, Vats 2025)**, conformational interconversion **(Bhakat 2025)**



Bhakat et al, "Generalizable Protein Dynamics in Serine-Threonine Kinases: Physics is the key" bioRxiv 2025
Gu et al, "Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE" eLife 2024
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Teng et al, "AlphaFold2-RAVE: Protein Ensemble Generation with Physics-Based Sampling" ChemRxiv 2025
Vani et al, "Exploring kinase Asp-Phe-Gly (DFG) loop conformational stability with AlphaFold2-RAVE" JCIM 2023
Vani et al, "AlphaFold2-RAVE: From Sequence to Boltzmann Ranking" JCTC 2023
Vats et al, "AlphaFold-SFA: Accelerated sampling of cryptic pocket opening, protein-ligand binding and allostery by AlphaFold, slow feature analysis and metadynamics" Plos One 2025

Replacing enhanced sampling with ensemble prediction

- Several examples involving AlphaFold2 + RAVE **(Vani 2023, Vani 2023a)**
  - RAVE is a protocol for converting MD trajectories and snapshots into Boltzmann-weighted state assignments
  - Kinases with/without ligand **(Gu 2024)**, GPCRs **(Vani 2023)**, SARS-CoV-2 RBD **(Teng 2025)**
- Collective variables from short MD trajectories of AF2 models match those of much longer simulations (>100 μs; **Teng 2025 & Vats 2025**)
- **Can lead to the observation of rare events, such as cryptic pocket opening (Meller 2023, Vats 2025), conformational interconversion (Bhakat 2025)**



Bhakat et al, "Generalizable Protein Dynamics in Serine-Threonine Kinases: Physics is the key" bioRxiv 2025
Gu et al, "Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-RAVE" eLife 2024
Meller et al, "Accelerating Cryptic Pocket Discovery Using AlphaFold" JCTC 2023
Teng et al, "AlphaFold2-RAVE: Protein Ensemble Generation with Physics-Based Sampling" ChemRxiv 2025
Vani et al, "Exploring kinase Asp-Phe-Gly (DFG) loop conformational stability with AlphaFold2-RAVE" JCIM 2023
Vani et al, "AlphaFold2-RAVE: From Sequence to Boltzmann Ranking" JCTC 2023
Vats et al, "AlphaFold-SFA: Accelerated sampling of cryptic pocket opening, protein-ligand binding and allostery by AlphaFold, slow feature analysis and metadynamics" Plos One 2025

Conclusions

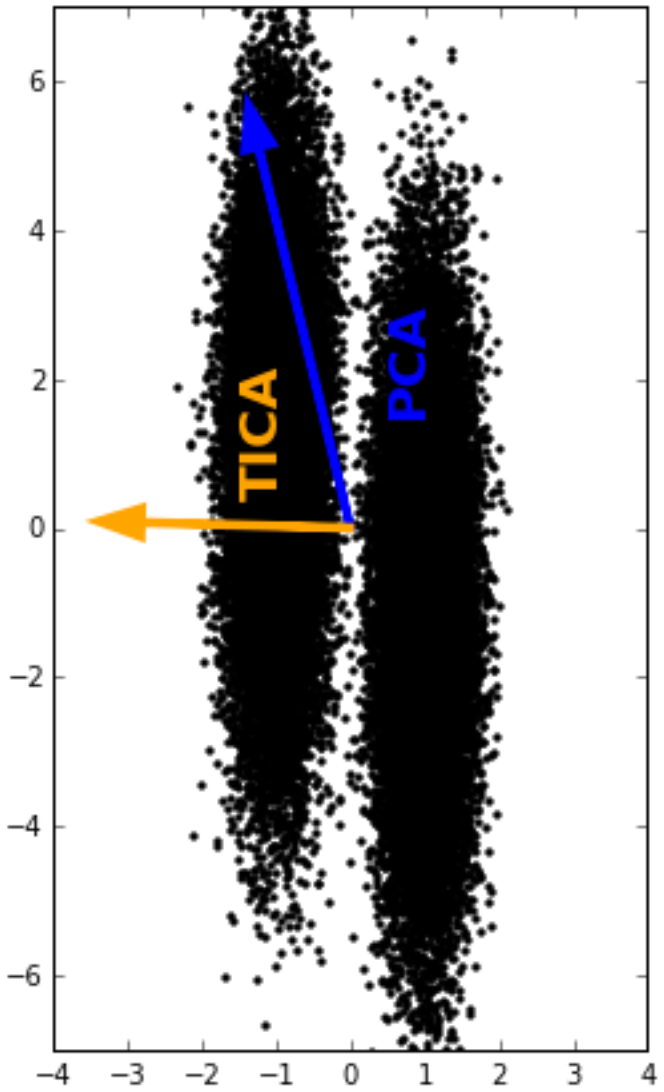
- Protein folding neural networks can be re-used for some dynamics prediction problems
- Other problems cannot be addressed by today's suite of tools
- New benchmarks and test cases being released all the time to get a sense of which tools have which strengths
- A separate branch of research is focusing on accelerating MD simulations to take huge time steps (not discussed here)

- **Availability**
  - Most MSA- and template-based approaches available on google colab via **ColabFold**
  - Most methods available on GitHub: **AlphaFlow**, **Rosetta-VAE**, **AFSample2**, **Cfold**, **aSAMt**, **BioEmu**; **ESMDiff** available on GitHub for non-commercial use

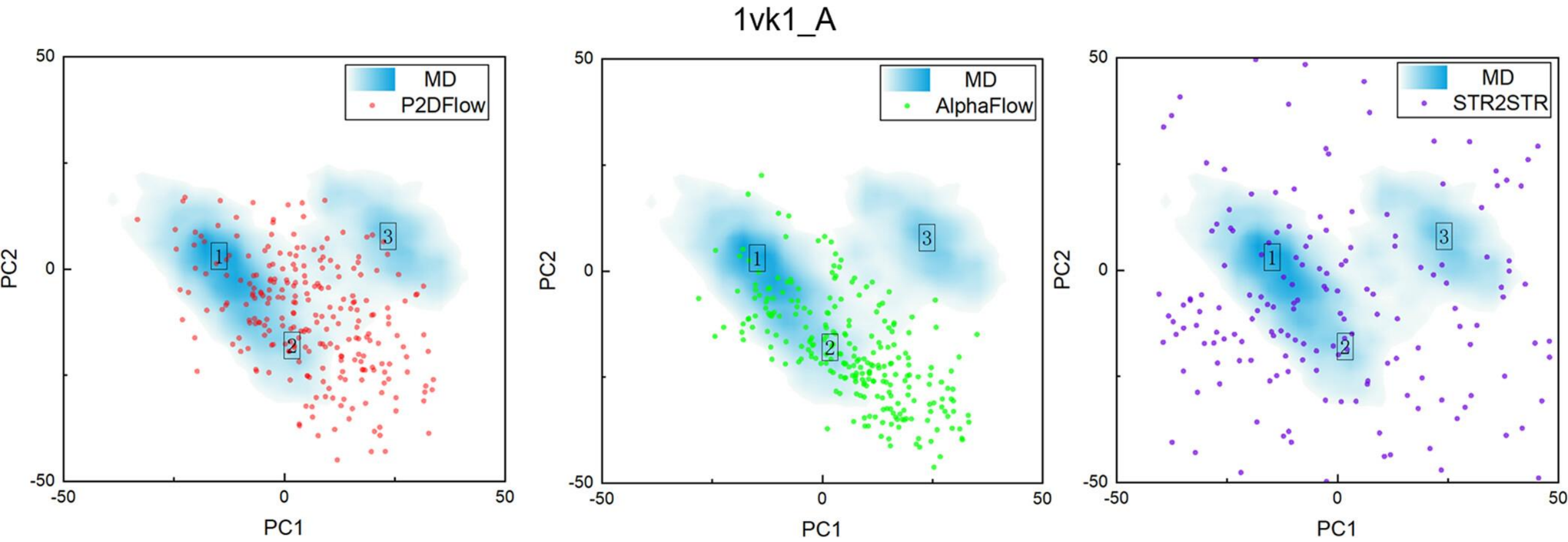
- **Additional resources**
  - "Modeling Boltzmann-weighted structural ensembles of proteins using artificial intelligence–based methods" by Aranganathan et al 2025
  - "Prediction of structural variation" by Kalakoti et al 2025

SUPPLEMENTAL SLIDES

What is TICA?



From Frank Noé's blog - http://docs.markovmodel.org/lecture_tica.html



Jin et al “P2DFlow: A Protein Ensemble Generative Model with SE(3) Flow Matching” JCTC 2025