

Kozakov/Vajda and ClusPro team in CASP/CAPRI experiment

Dima Kozakov

Laufer Center for Physical and Quantitative Biology

Stony Brook University

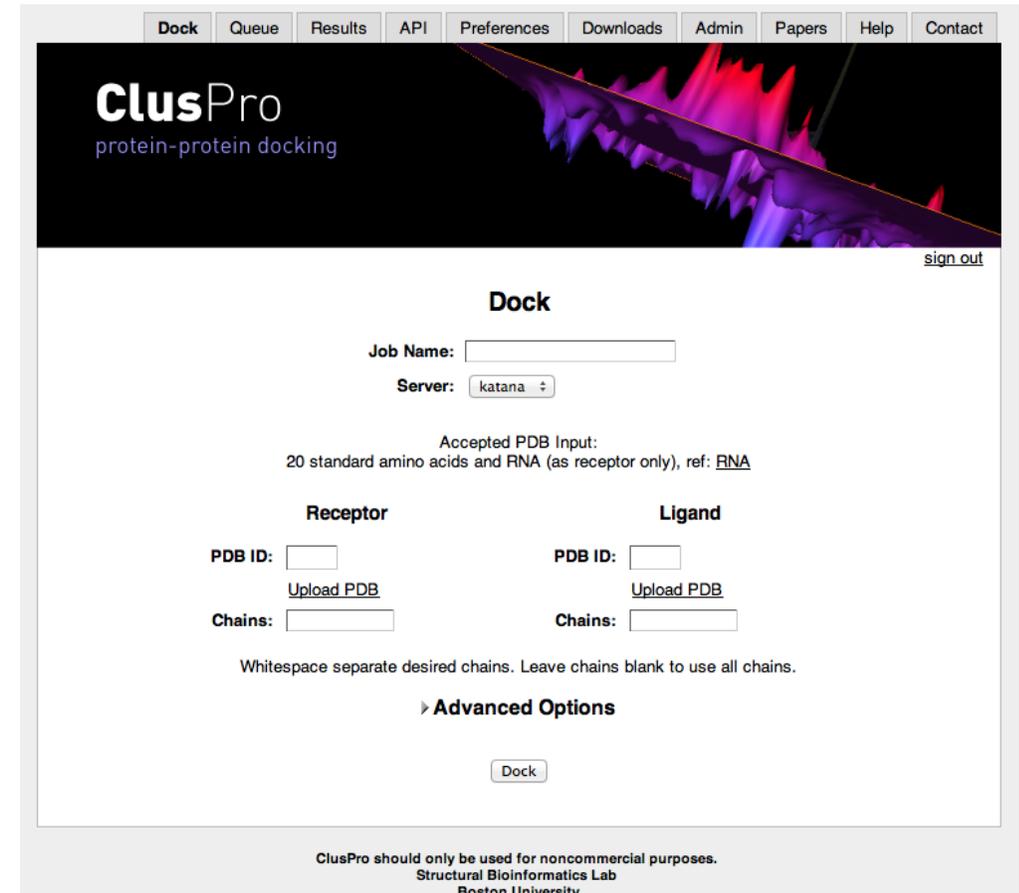


A bit of history

Sampling DOT+ FTDOCK + Scoring (Energy+Cluster size)

Sampling PIPER (multiple weights) + Scoring (Energy+Cluster size)

.....



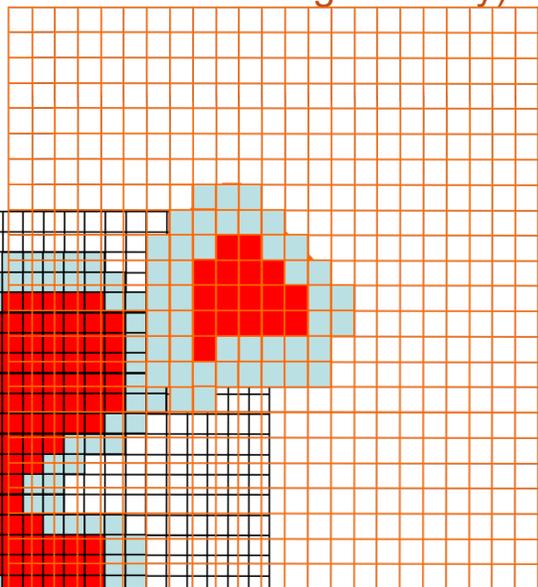
The screenshot shows the ClusPro web interface for protein-protein docking. At the top, there is a navigation menu with links for Dock, Queue, Results, API, Preferences, Downloads, Admin, Papers, Help, and Contact. Below the menu is a header banner with the ClusPro logo and the text "protein-protein docking". A "sign out" link is visible in the top right corner of the banner. The main content area is titled "Dock" and contains the following fields and options:

- Job Name:** A text input field.
- Server:** A dropdown menu currently set to "katana".
- Accepted PDB Input:** 20 standard amino acids and RNA (as receptor only), ref: [RNA](#)
- Receptor:**
 - PDB ID:** A text input field.
 - Upload PDB:** A link to upload a PDB file.
 - Chains:** A text input field.
- Ligand:**
 - PDB ID:** A text input field.
 - Upload PDB:** A link to upload a PDB file.
 - Chains:** A text input field.

Below the input fields, there is a note: "Whitespace separate desired chains. Leave chains blank to use all chains." and a link for "Advanced Options". At the bottom of the form is a "Dock" button. The footer of the page contains the text: "ClusPro should only be used for noncommercial purposes. Structural Bioinformatics Lab Boston University".

Systematic FFT based sampling

Ligand representation (e.g. charge density)



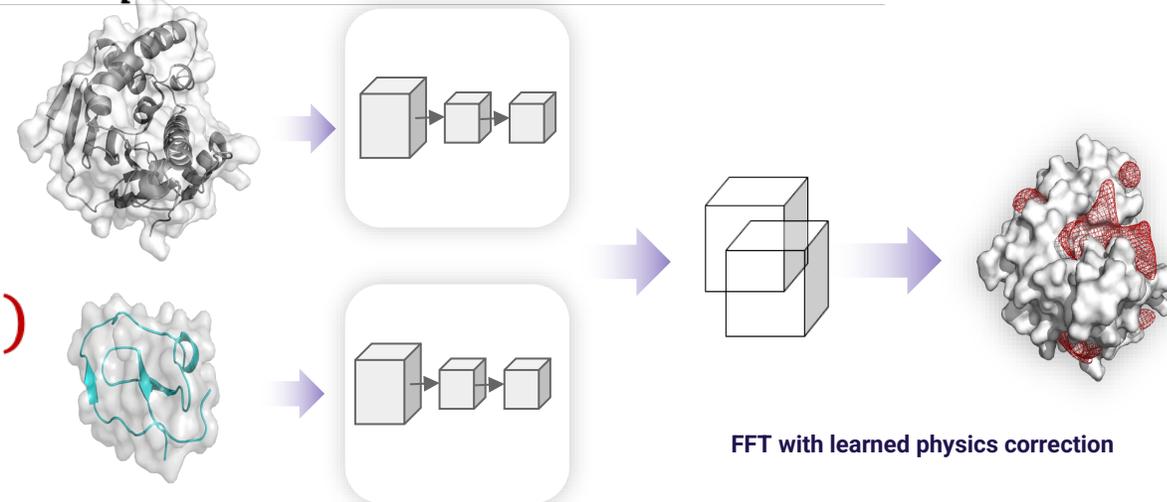
Interaction energy as a sum of FFT convolutions

$$E(\bar{t}, r) = \sum_n \sum_{\bar{x}} R_p(\bar{x}) L_{pr}(\bar{t} - \bar{x})$$

$$\begin{aligned} E_{full} &= E_{vdw} + E_{elec} + E_{pair} \\ E_{vdw} &= E_{rep} + E_{attr} \\ E_{elec} &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \frac{q_i q_j}{\left(r_{ij}^2 + D^2 \exp\left(-\frac{r_{ij}^2}{4D^2}\right)\right)^{1/2}} \\ E_{pair} &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \varepsilon(i, j) = \begin{cases} 0, & r_{ij} > D \\ \varepsilon_{ij}, & d < r_{ij} < D \end{cases} \\ \varepsilon_{ij} &= \sum_t \varepsilon_{ti} \lambda_t \varepsilon_{tj} \end{aligned}$$

$$E(\bar{t}, r) = IFT \left[\sum_p FT^* \{R_p(\bar{x})\} FT \{L_{pr}(\bar{x})\} \right]$$

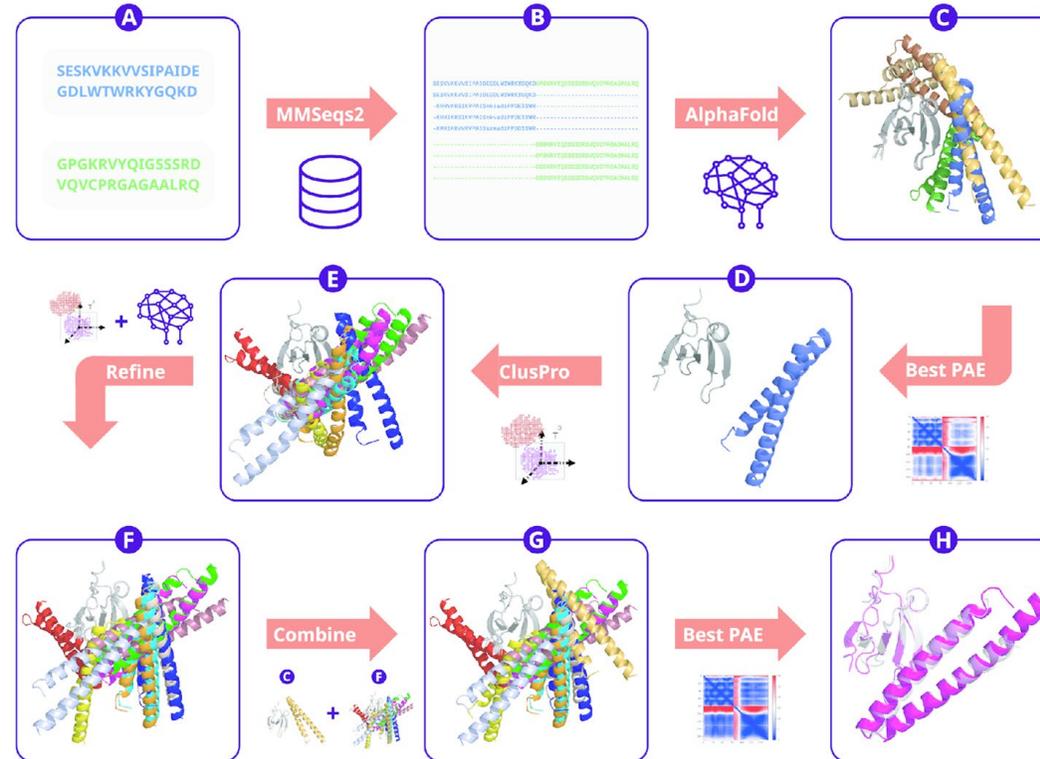
Receptor representation $O(N^6) \rightarrow O(N^3 \ln N^3)$
(e.g. electrostatic potential)



Ideas –

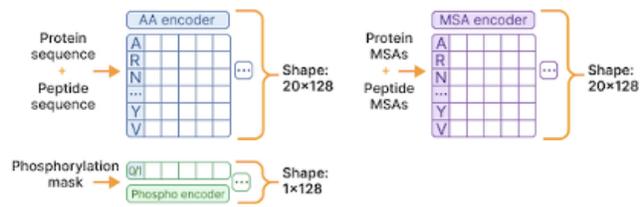
Can we sample more if the model is not confident-

How do we combine the samples from different approaches?



Pytorch-AF – Customizable Alphafold-style architecture- Local refinement

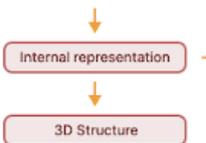
5.1 Input encoding



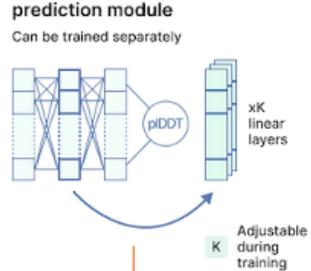
5.2 The Evoformer



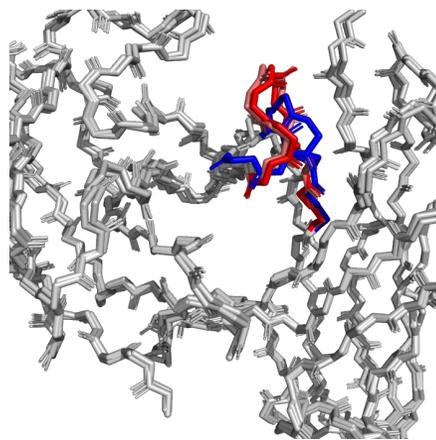
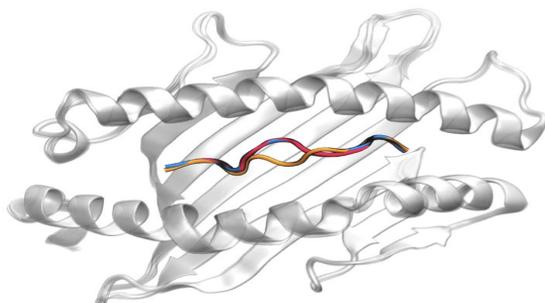
5.3 The structure module



5.4 The confidence prediction module

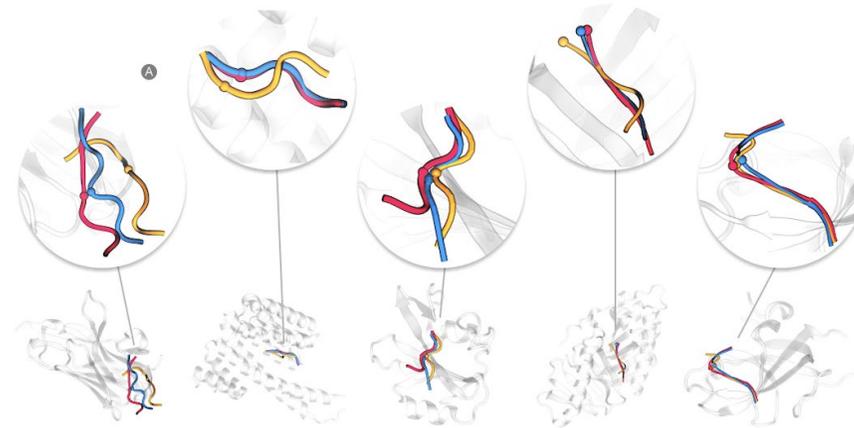
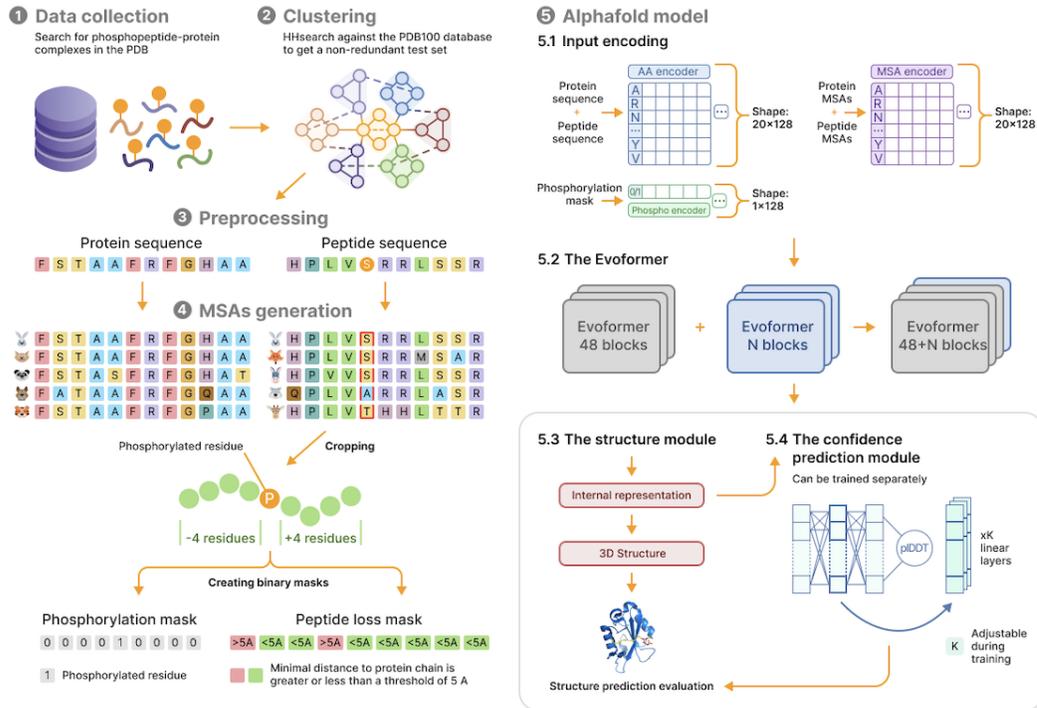


Structure prediction evaluation



Jumper et. al. 2021; Glukhov et. al 2024 Biophysical journal;
Glukhov et. 2024; Averkava et. al in preparation

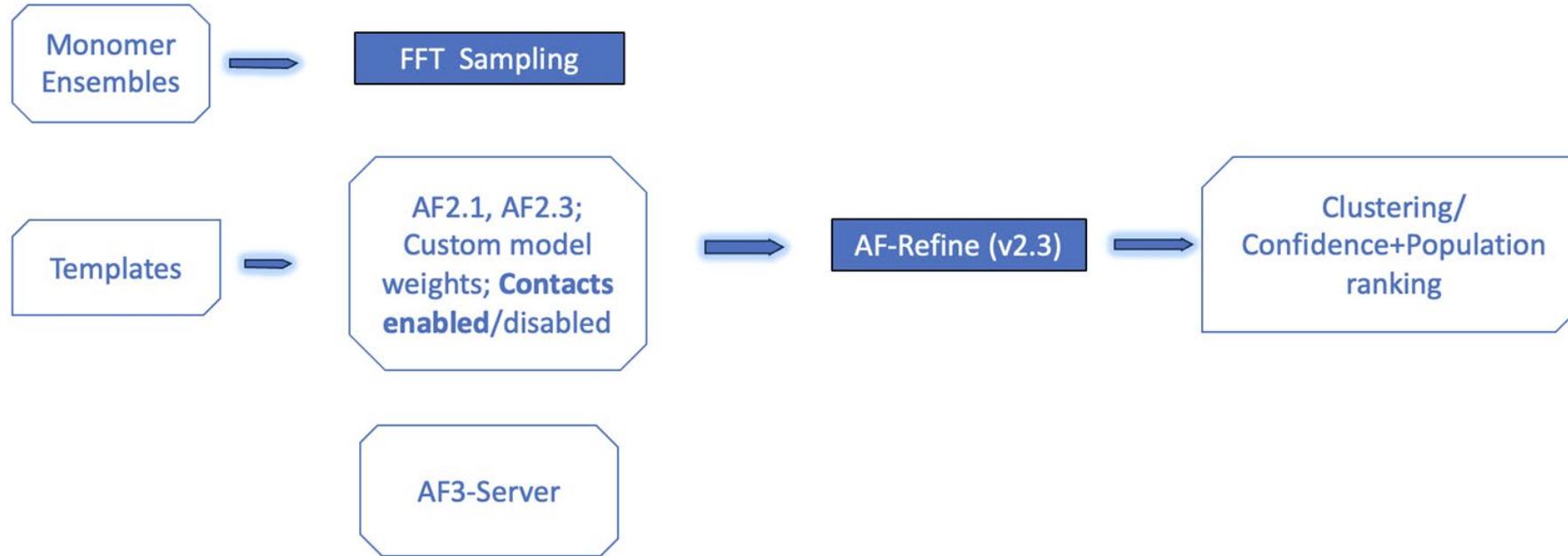
Global models Peptide and Phosphorylated interactions



Additional recent complexes model trained on recent Data

Glukhov et. al, Biorxiv 2024; Averkava et.al in preparation

The Pipeline

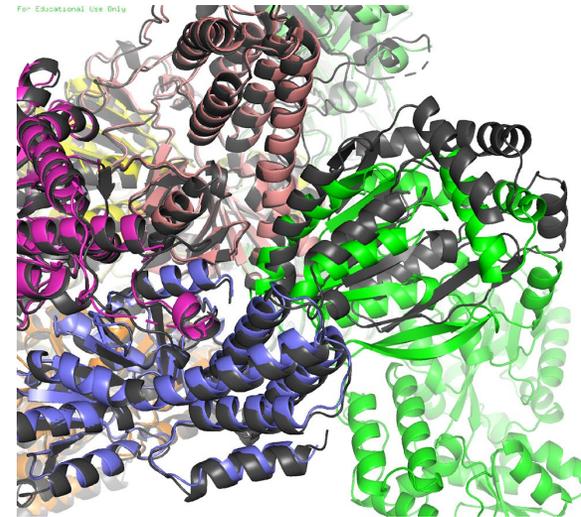


What we did as human group – keep sampling more (bigger ensemble of monomers for FFT , more seeds for AF models)

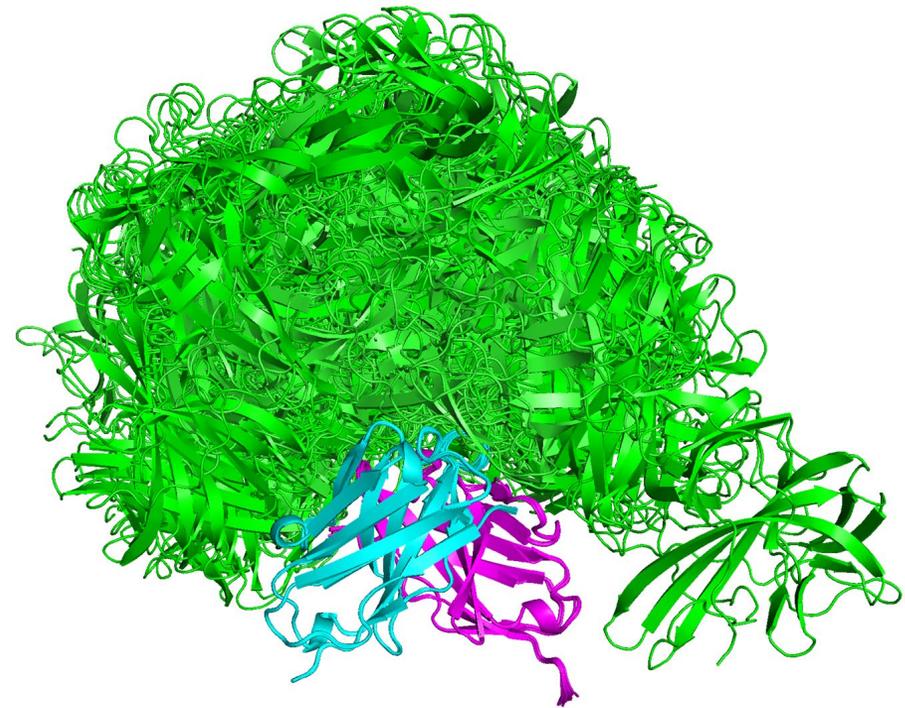
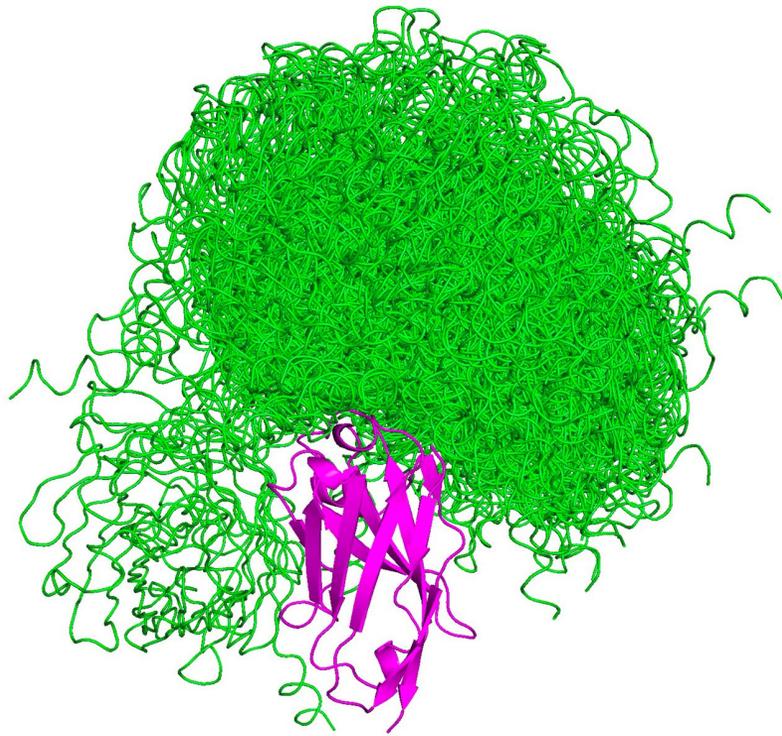
AcP10_eIF2B_complex, A2B2C2D2E2F2



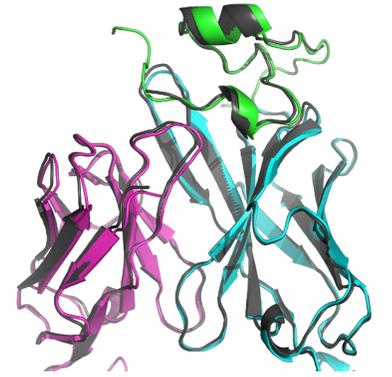
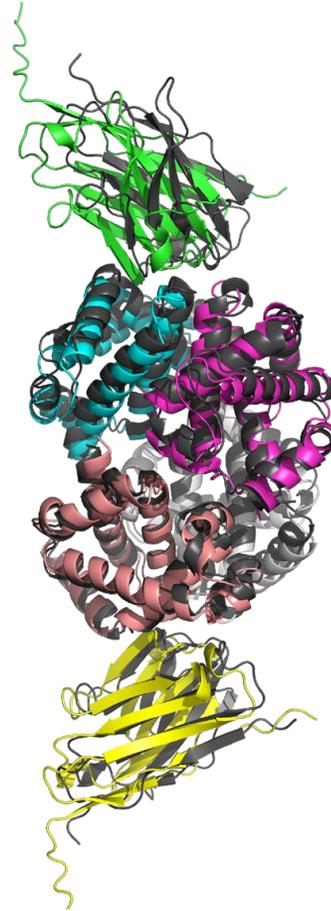
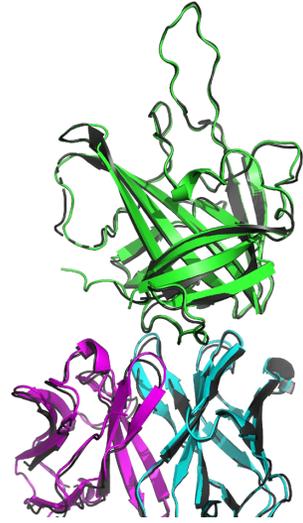
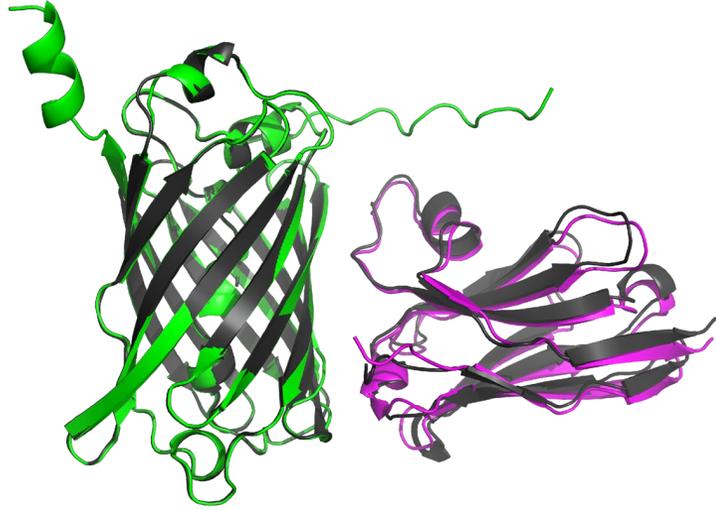
Green: AcP10 protein
Others: eIF2B complex



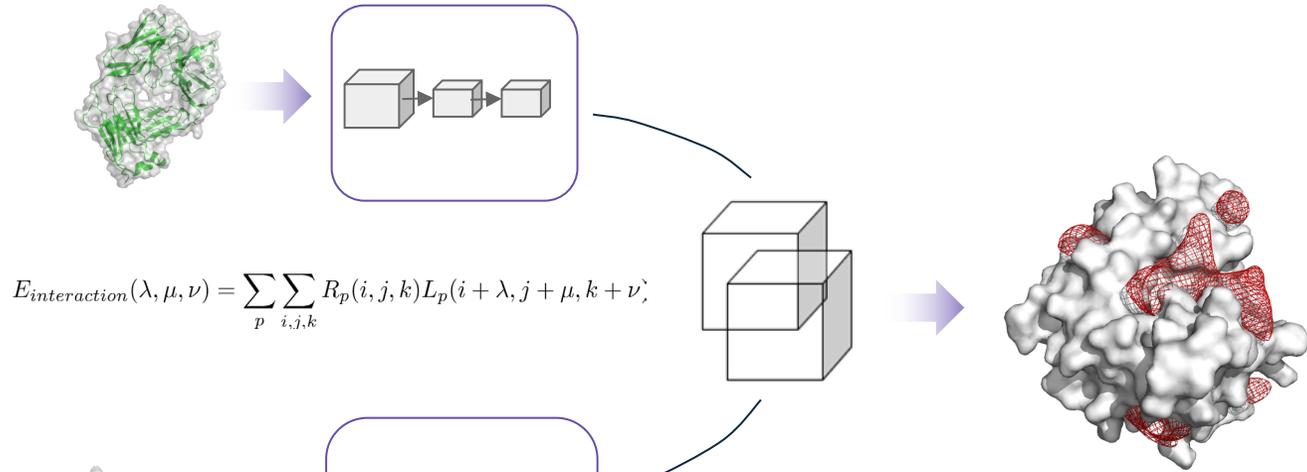
In antibody we do the focused sampling in CDRS (special potential). Monomer Ensembles from AF models



Antibody/Nanobody

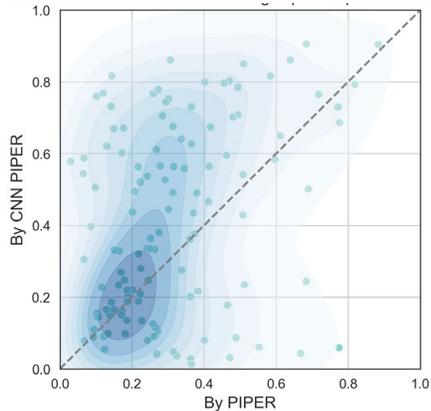


Improved antibody modeling

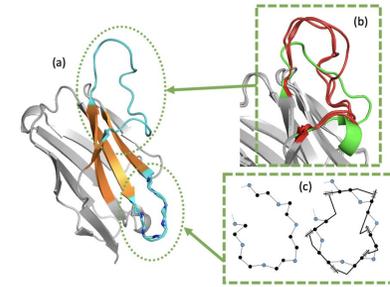
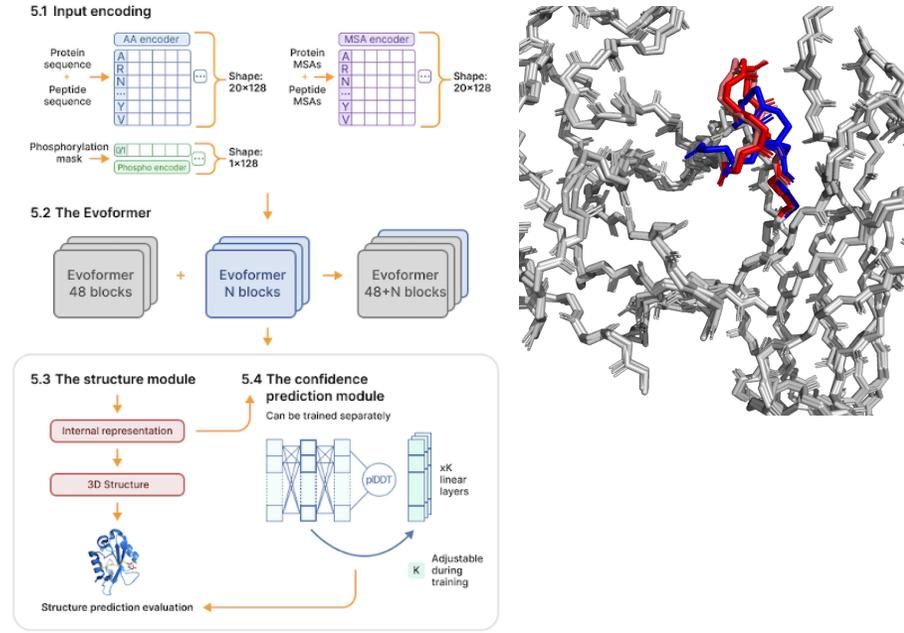


E_{vdw} E_{coul} E_{solv} E_{hb}

FFT physics plus deep learning correction – Improving energy function.

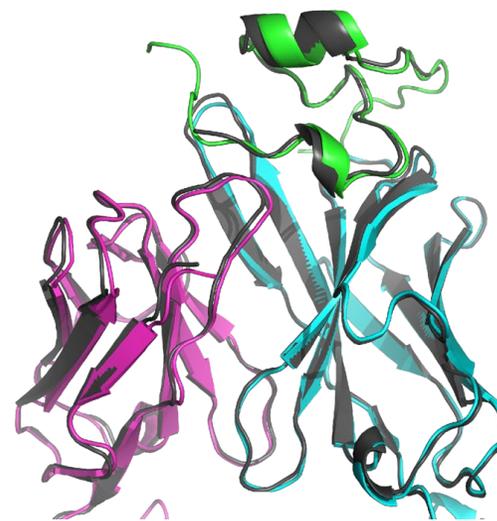
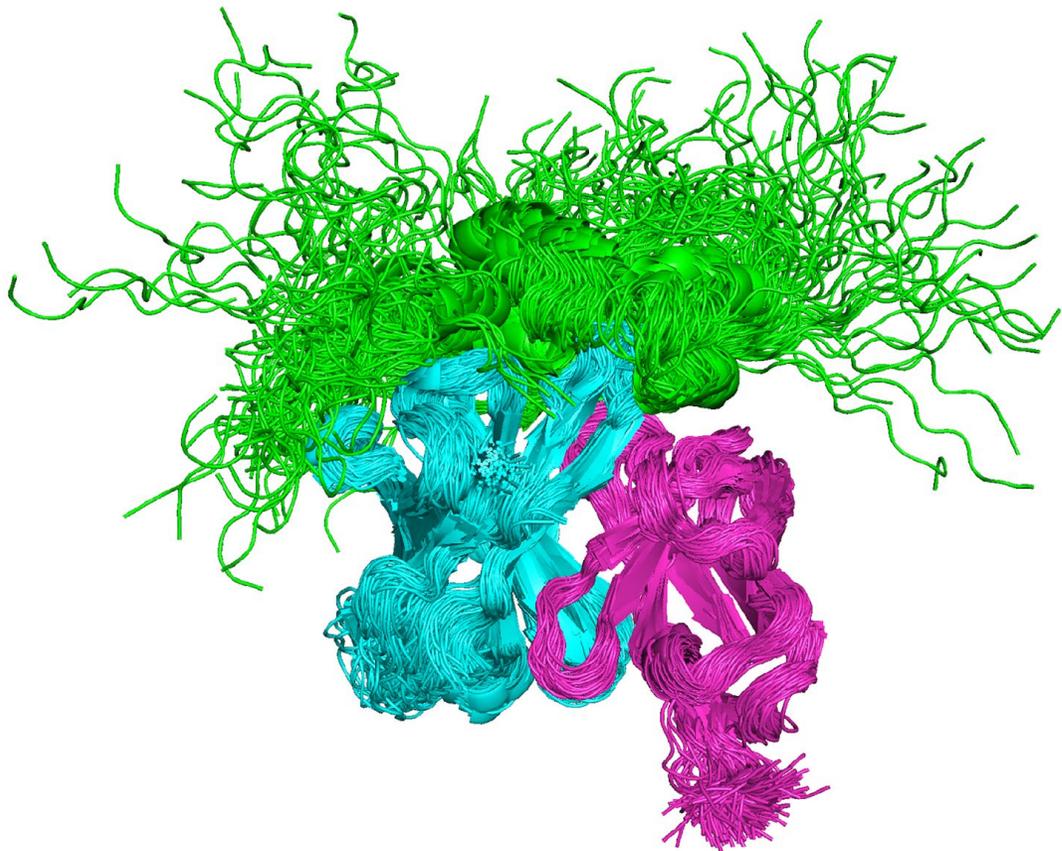


Finetuned model for antibody prediction



Diffusion model on thoracic varieties for antibody prediction

Human group (MD Sampling starting from ClusPro followed by AF - Refinement)



DockQ – (0.69 - >0.71)

ClusPro 3 beta

Submit a new job

Please input either a PDB ID or select a file for upload.

Title: Name of your job (optional)

Ensemble Mode

Rec ensemble: no file selected

Lig ensemble: no file selected

Fasta Mode

Temp input file: no file selected

Model type: 

Fasta inputs:

 Please enter protein fasta chain id and fasta seq, one per line A:FASTA

Chain mapping: Mapping from fasta to pdb, A,B in fasta X,Y in pdb A:X B:Y

Interchain contacts: Interchain contacts ABC C

Acknowledgements

Stony Brook University:

- **Ryota Ashizawa**
- Dzmitry Padhorny
- Ernest Glukhov
- Veranika Averkava
- Yimin Zhu
- Ernest Glukhov
- George Jones
- Dasha Stepanenko
- Dr. Dzmitry Padhorny
- Prof Carlos Simmerling

Boston University:

- Omeir Khan
- Masha Lazou
- Ayse Bekar
- Prof. Sandor Vajda

Lviv Insitute for Condensed Matter Physics:

- Dr. Taras Patsahan
- Dmytro Kalytin
- Bohdan Mahometa

Funding NIH, NSF

**DOE INCITE
Leadership award**