



OpenComplex-2: A Pure Data-Driven System for Biomolecular Conformation Prediction

BEIJING ACADEMY OF ARTIFICIAL INTELLIGENCE



Qiwei YE, Tech Lead
2024.12

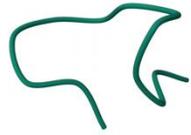




OpenComplex -2 : From Structure to Ensemble



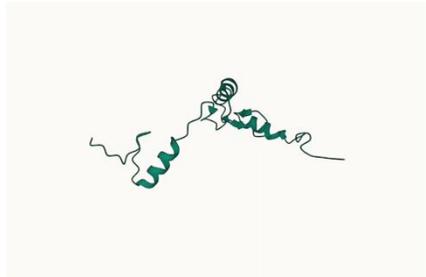
(a) OpenComplex Sampling



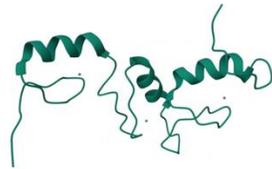
(b) NMR Result
(pdbid: 9bvs)



(c) AlphaFold3 Sampling



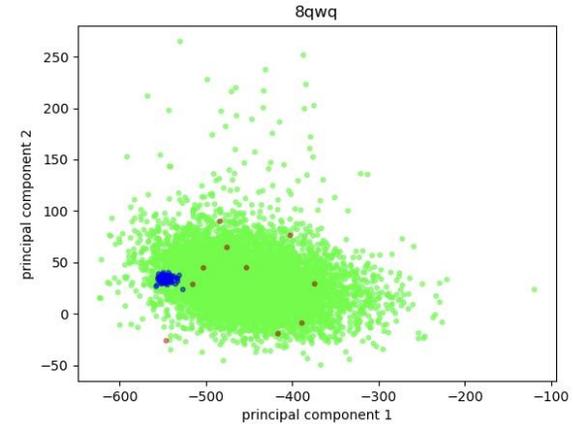
(a) OpenComplex Sampling



(b) NMR Result
(pdbid: 8tho)



(c) AlphaFold3 Sampling



PCA Visualization of multiple samples ;

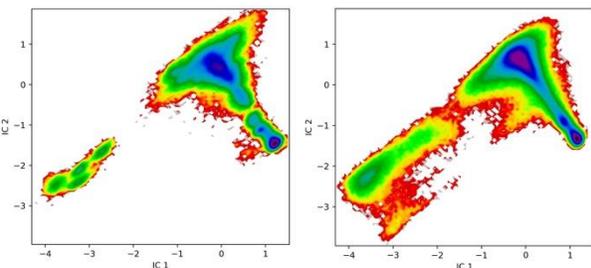
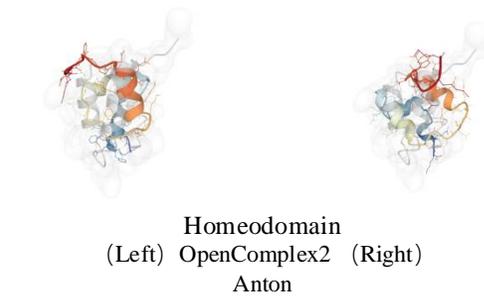
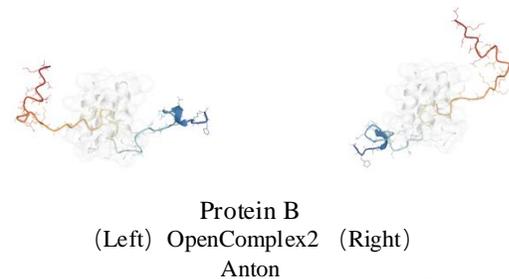
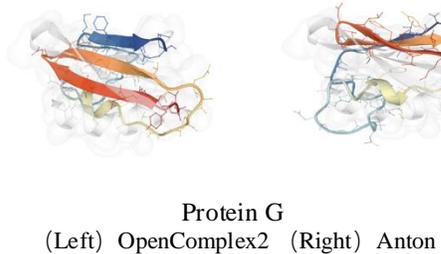
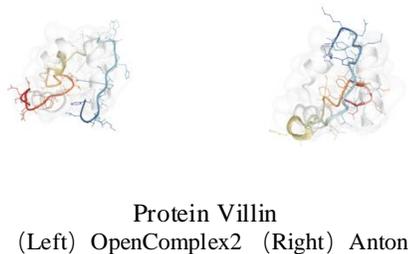
AlphaFold3 tends to predict a stable conformation that cannot cover **NMR** results, while **OpenComplex** sampling can provide stable coverage

AlphaFold3 (100 samples)	OpenComplex2 (100 samples)	OpenComplex2 (1000 samples)
1.9%	31.3%	98.4%

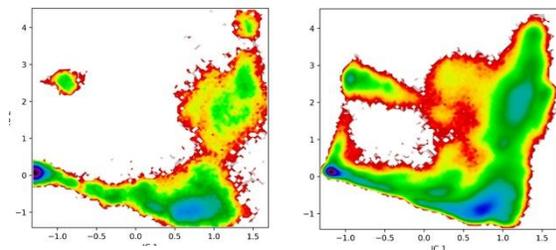
The Successful Coverage Rate over different sampling



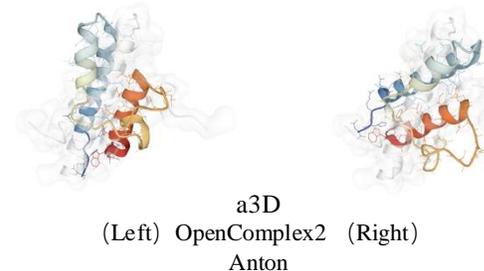
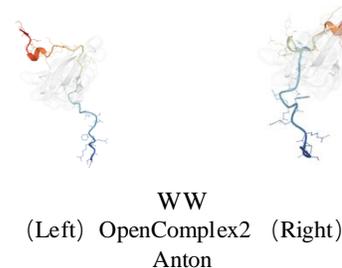
OpenComplex -2 : Pathway Prediction



TICA Map of Protein Villin under 1M samples
(Left) OpenComplex2 (Right) Anton



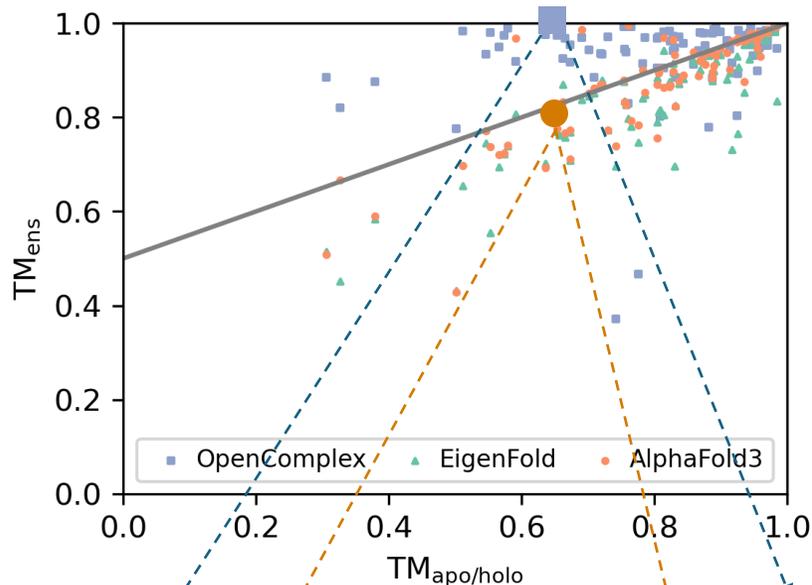
TICA Map of Protein G under 1M samples
(Left) OpenComplex2 (Right) Anton



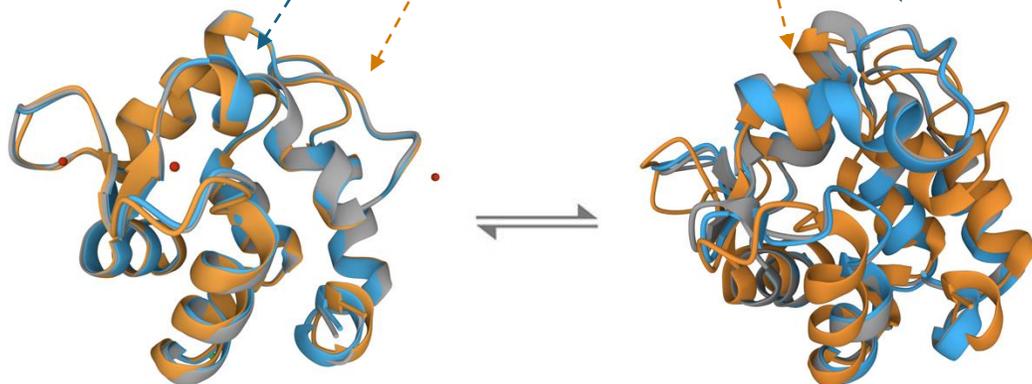


OpenComplex-2 : Multi-Conformation prediction

➤ OpenComplex captures ligand-bound states



- Induced-fit docking phenomenon: the interaction between a protein and a rigid binding partner induces a **conformational change** in the protein
- These conformational changes promote key regulatory mechanisms, including signal transduction, cooperativity, etc
- AlphaFold3 does not reproduce conformational diversity



1RRO (Holo)

TM score: 0.99 vs 0.98

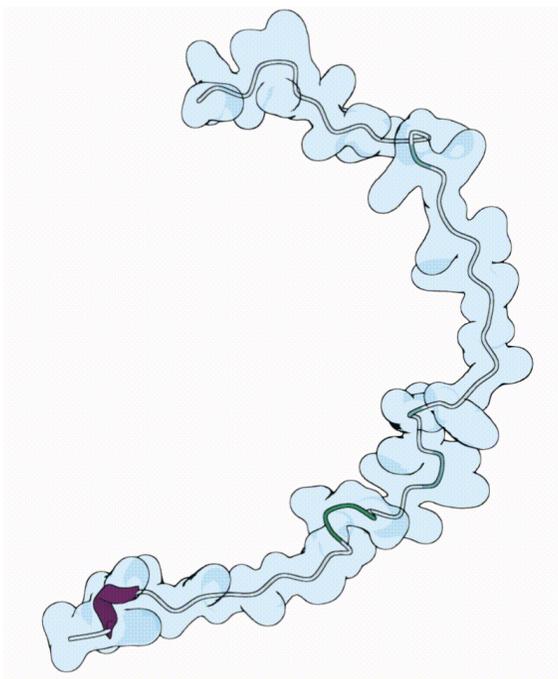
2NLN (Apo)

TM score: 0.98 vs 0.66

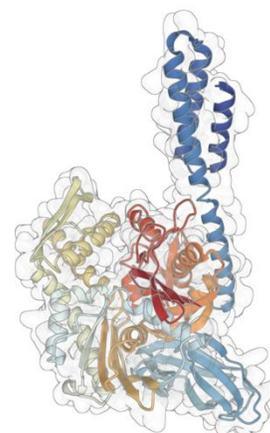
- Experimental
- OC2-predicted
- AF3-predicted



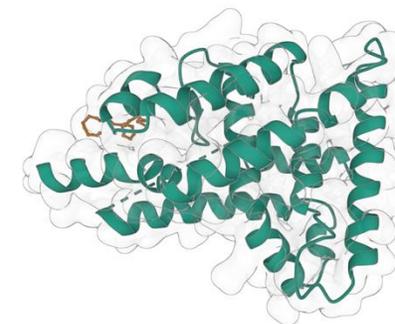
OpenComplex-2: Dynamic Simulation (In-house)



Crambin protein folding simulation:
Simulated the process of protein
gradually folding from an unfolded state
into a stable conformation



AMP-activated protein kinase
conformational change simulation: In
the absence of ADP, AMPK switches
between open and closed states



Retinoid X Receptor (RXR) ligand
binding simulation: After ligand binding,
the RXR protein undergoes a
conformational change.

OpenComplex -2 : A Pure Data-Driven System for Biomolecular Conformation Prediction

Key takeaway:

- A **homology-aware complete-graph representation**
 - encodes comprehensive structural and chemical information between all atom pairs
- **Graphfusion:**
 - FloydAttention for efficient information propagation
 - Pure Diffusion without auxiliary losses

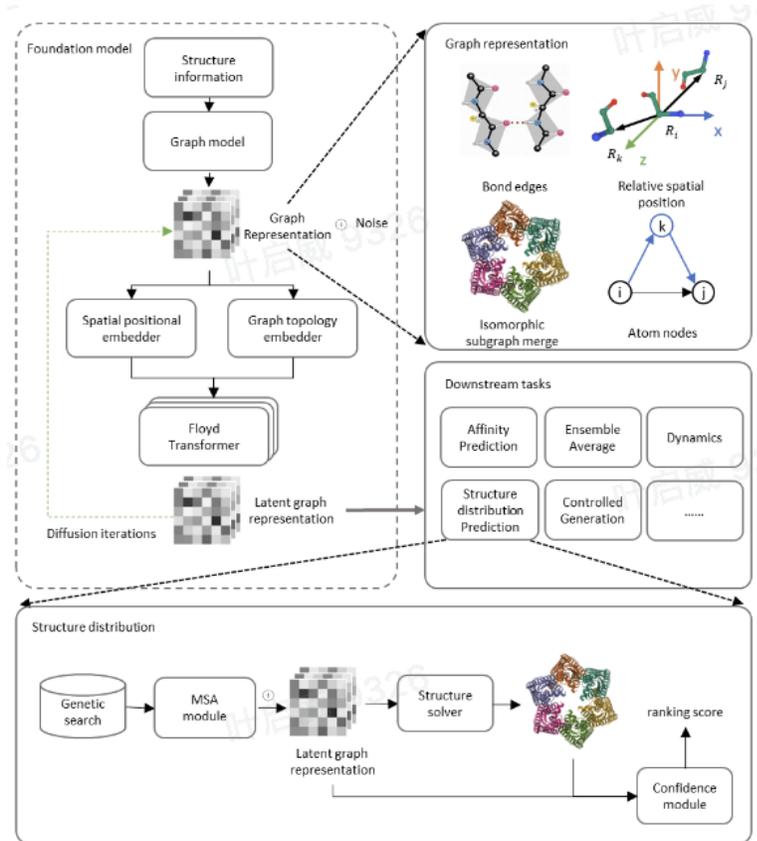


Figure 1. Overview of OpenComplex2 architecture and training flow. OpenComplex2 consists of two training stages: in the pre-training stage, the model focuses on structural relationships between atoms; in the fine-tuning stage, it incorporates additional information for specific tasks.

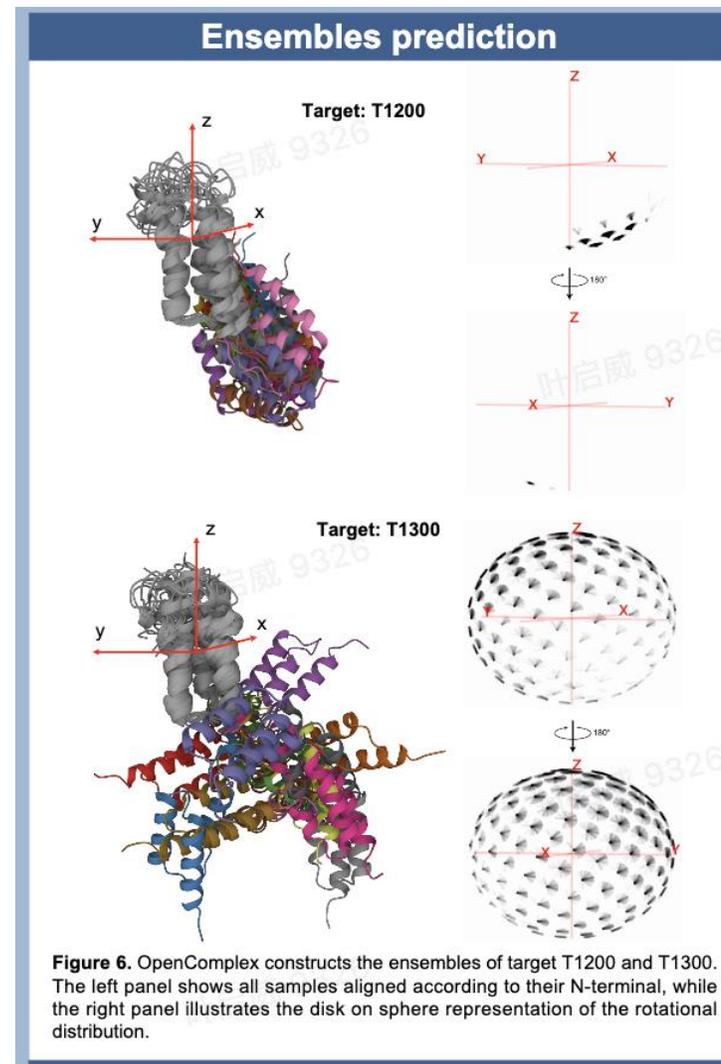


Figure 6. OpenComplex constructs the ensembles of target T1200 and T1300. The left panel shows all samples aligned according to their N-terminal, while the right panel illustrates the disk on sphere representation of the rotational distribution.

Thanks >

 Chivee.ye@gmail.com

 北京市海淀区智源大厦



智源公众号