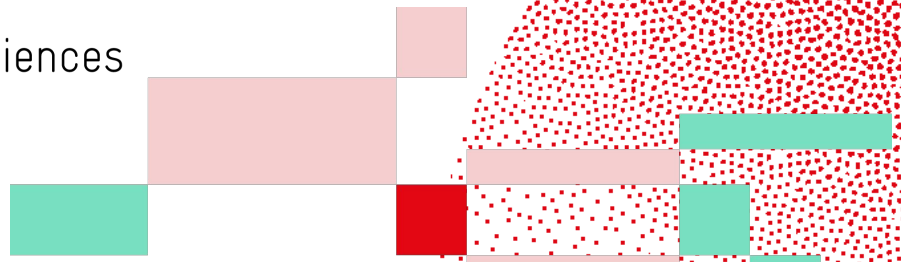
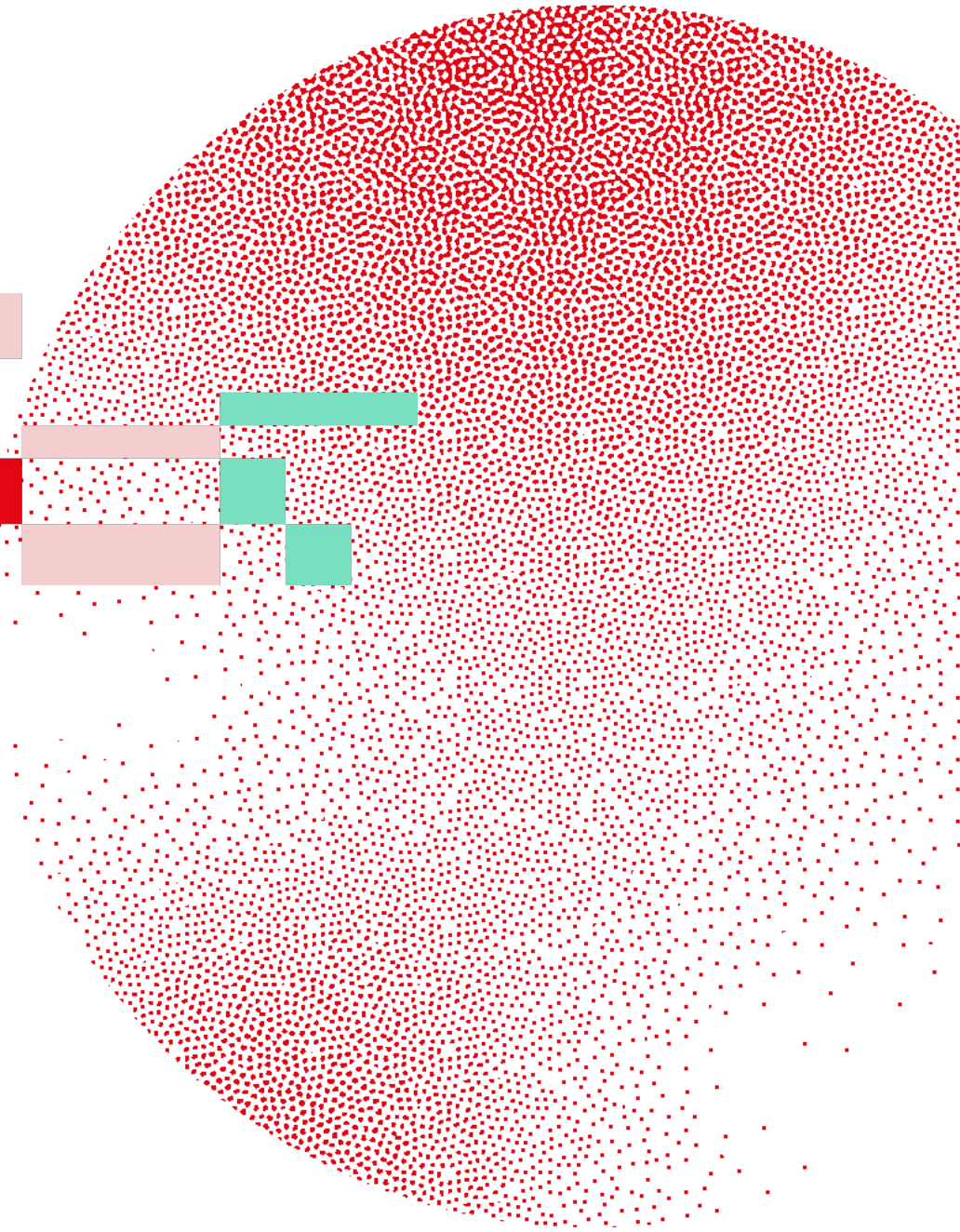
Swiss Institute of Bioinformatics

BIOZENTRUM

Universität Basel
The Center for
Molecular Life Sciences



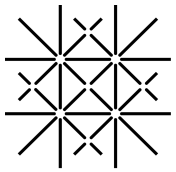


CAMEO

Automated benchmarking of protein complex predictions

Peter Škrinjar

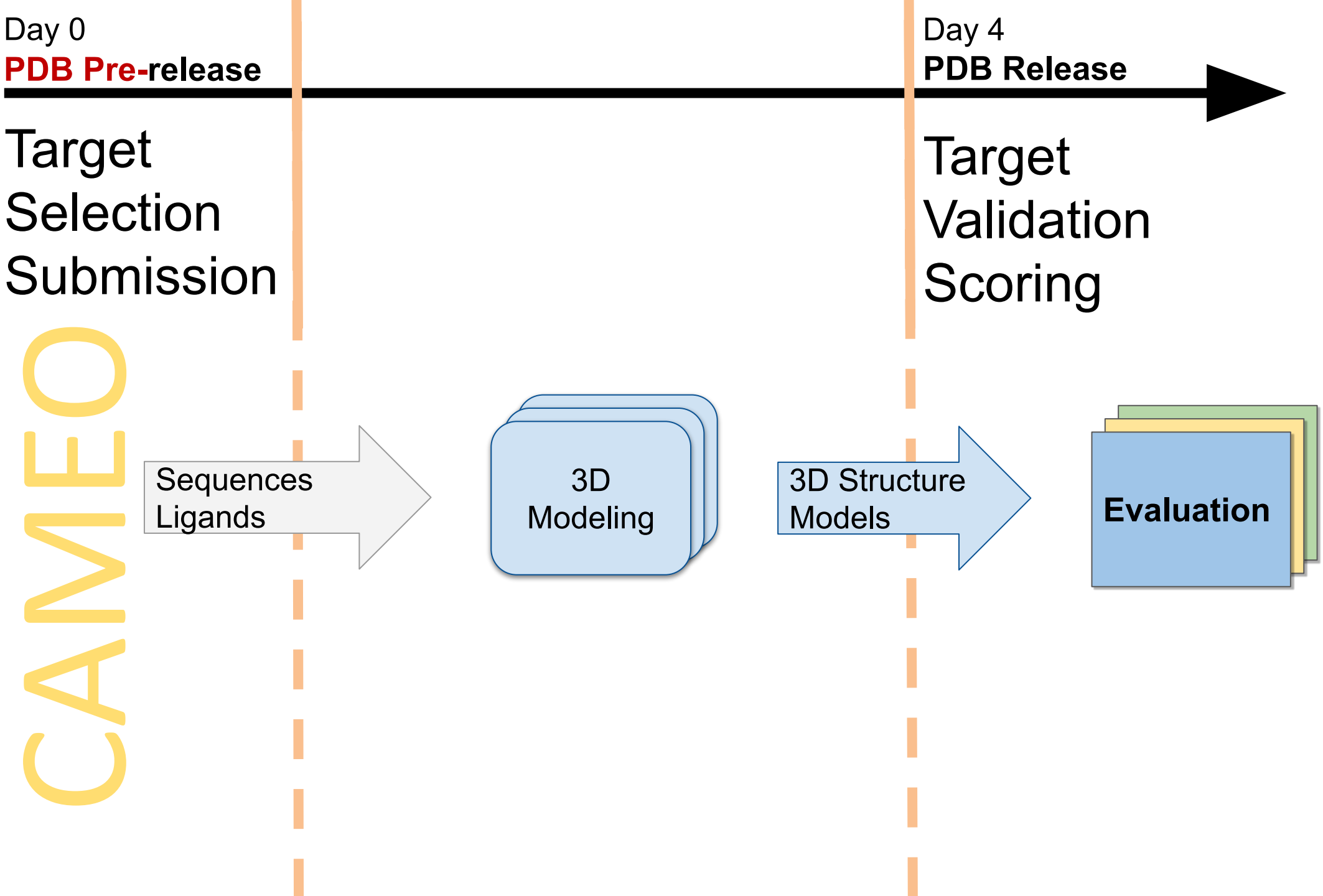
CASP16
December 1st, 2024



Universität Basel

Outline

- Introduction
- Updates in CAMEO
- Challenges in automated benchmarking of protein complexes:
  - PLINDER
- Outlook


Day 0
PDB Pre-release
Day 4
PDB Release
Target
Selection
Submission
Target
Validation
Scoring
CAMEO
Sequences
Ligands
3D
Modeling
3D Structure
Models
Evaluation

CAMEO is complementing CASP

CASP

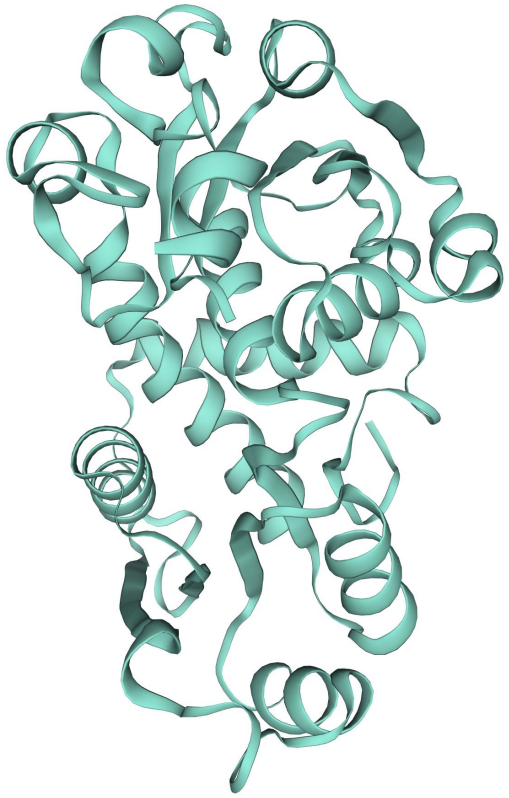
- Blind predictions
- Every 2 years
- Human curation
- Human & Server predictors
- Hand-picked targets
  - Obtained from crystallographers
  - Hard folding targets
- Experts assessment

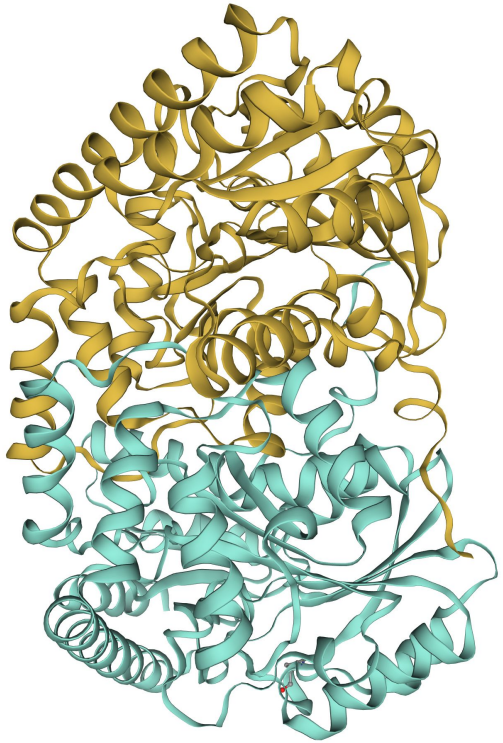
CAMEO

- Blind predictions
- Weekly
- Fully automated
- Only servers
- Targets from PDB pre-release
  - Selection of interesting, diverse targets
  - Typically easier targets
- Automated evaluation

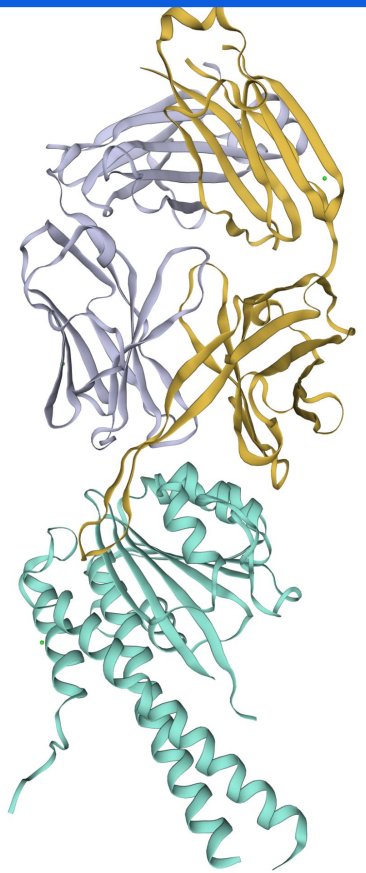
CAMEO - Beyond Single Protein Chains

Single chain and homomeric complexes

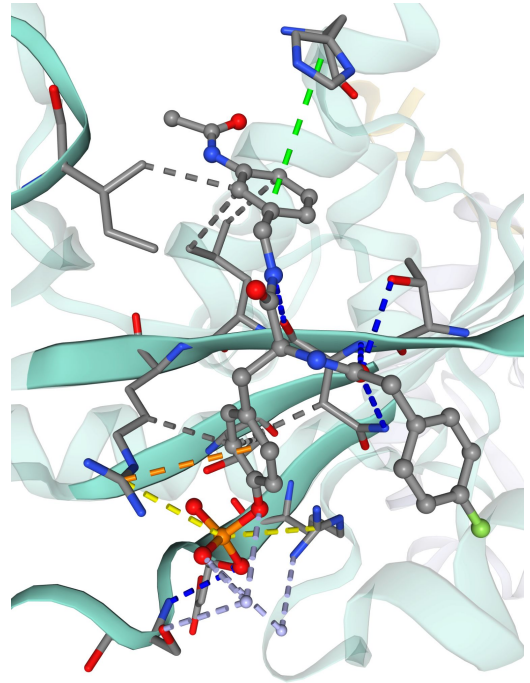




Heteromeric complexes



Ligands



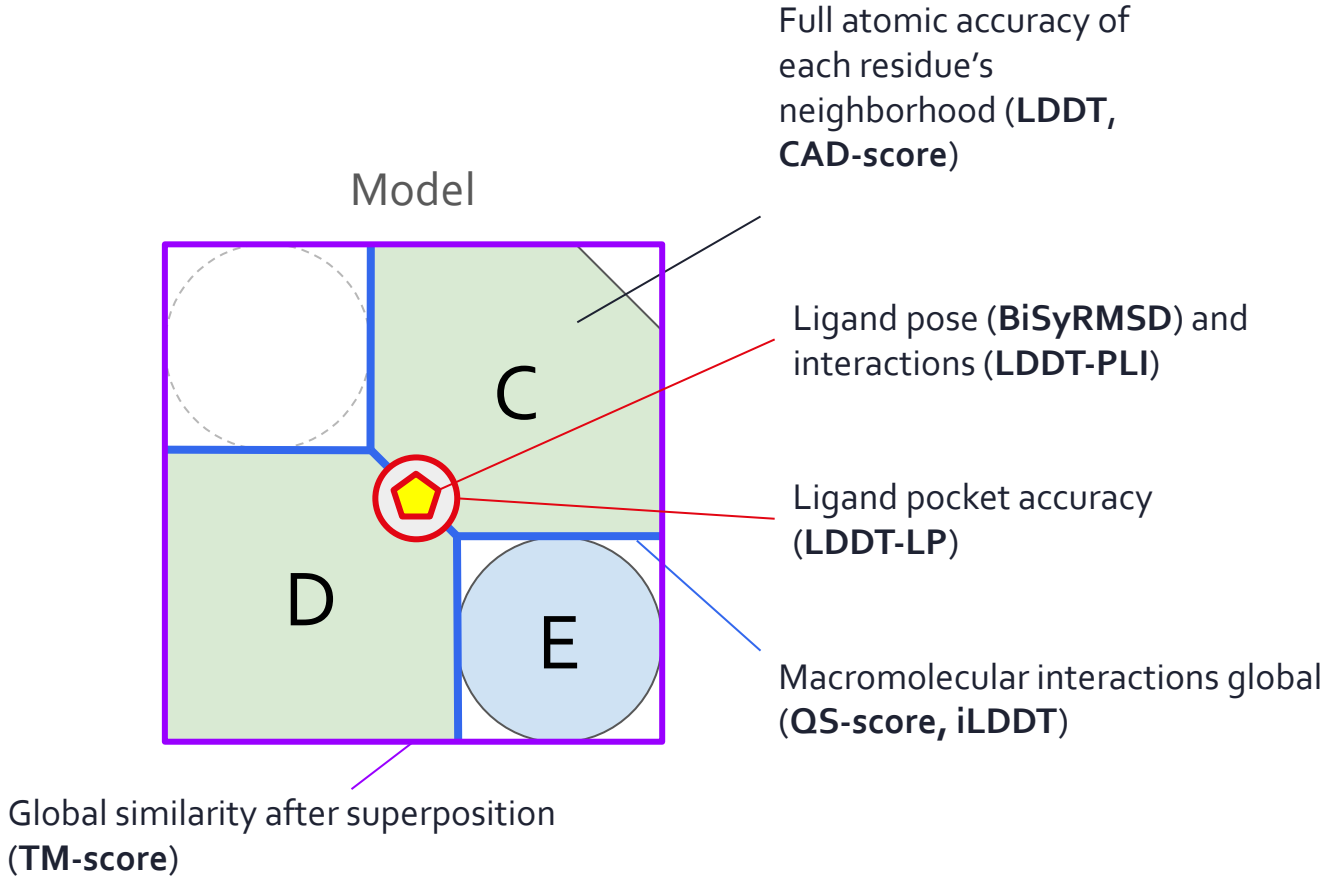
CAMEO - Beyond Single Protein Chains

https://beta.cameo3d.org/

help-cameo3d@unibas.ch

CAMEO Update

- Included ligand targets
- Included ligand scores for ligands and binding pockets:
  - Symmetry-corrected RMSD
  - LDDT-PLI
  - LDDT-LP
- Baseline predictors:
  - AlphaFold 3 (Server)
  - SWISS-MODEL + Schrödinger Glide
  - SWISS-MODEL + AutoDock Vina

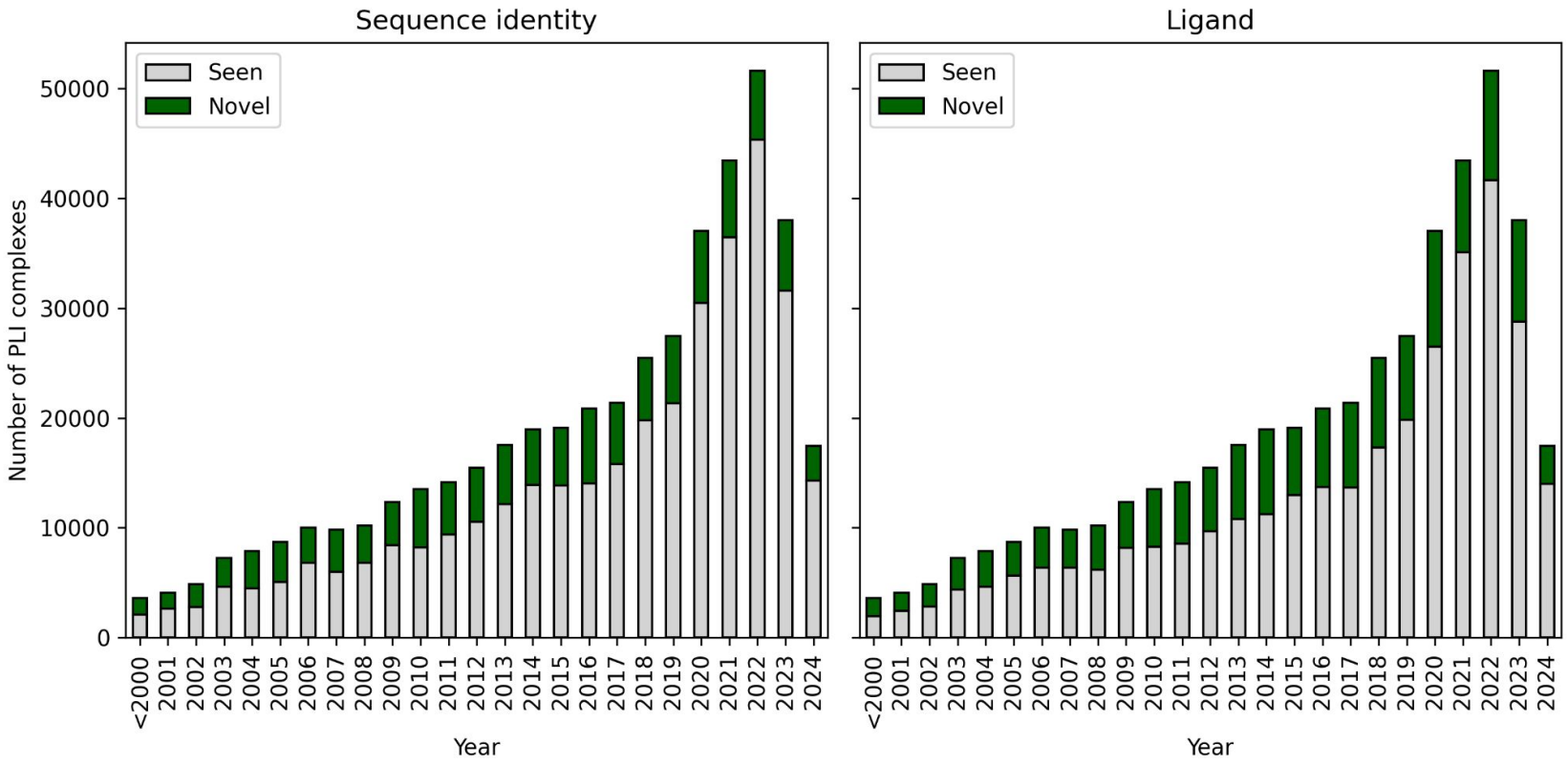


Challenges in automated benchmarking of protein complexes

- Defining difficulty/novelty for complexes:
  - “New” PDBs != Novel complexes
  - “Novelty” definition depends on the prediction task
  - DL-methods tend to memorize, not generalize
- Quality of the ground truth structure:
  - Information not known in the preprocessing stage
- Restricted use of latest methods:
  - AlphaFold3 Server
  - Chai-1 Server

“New” PDBs != Novel complexes

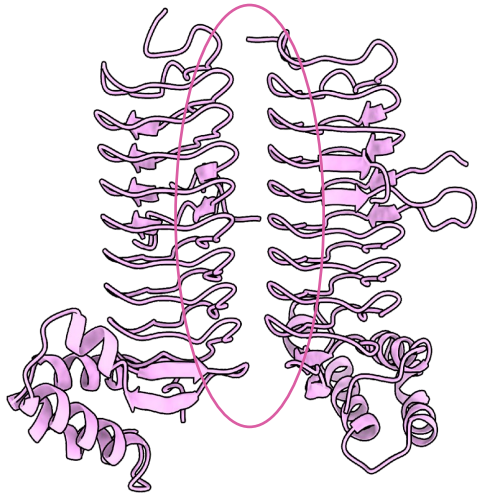
- ~80% PLC in 2023 have >30% seq. identity to previous
- DL-methods trained on time split overestimate performance
- Sequence and ligand identity not enough to decide difficulty



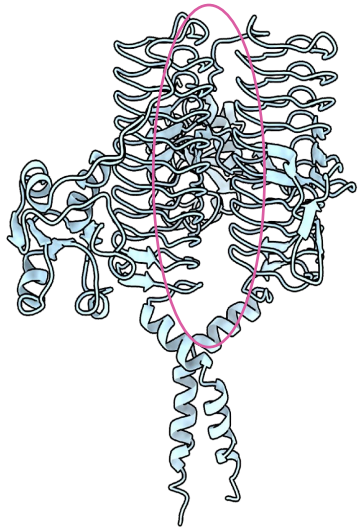
Leemann, Michèle, et al. *Proteins: Structure, Function, and Bioinformatics* 91.12 (2023): 1912-1924.

“Novelty” definition depends on the task

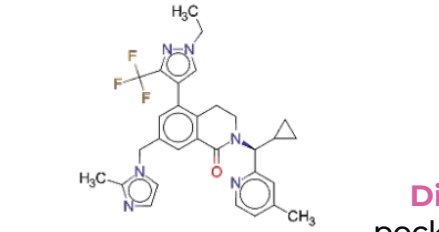
- Generalisation difficulty = Interaction dissimilarity

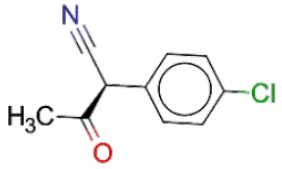


7OJQ
ProteinFlow
Train Split

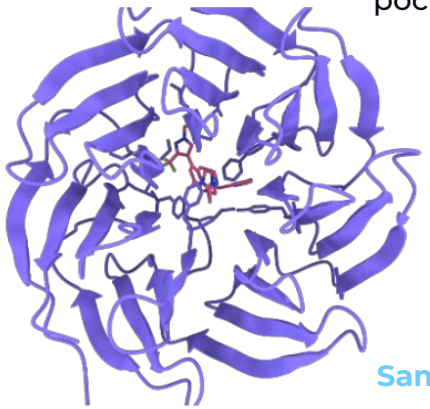


2IUA
ProteinFlow
Test Split



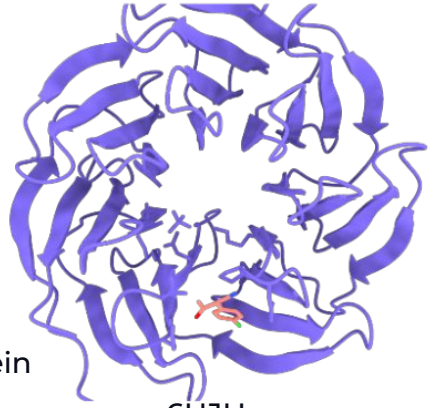


Different
pocket, ligand



7UAS

Same protein

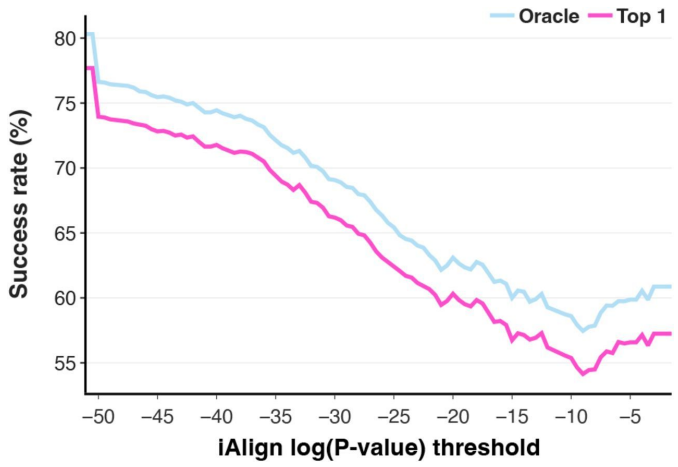


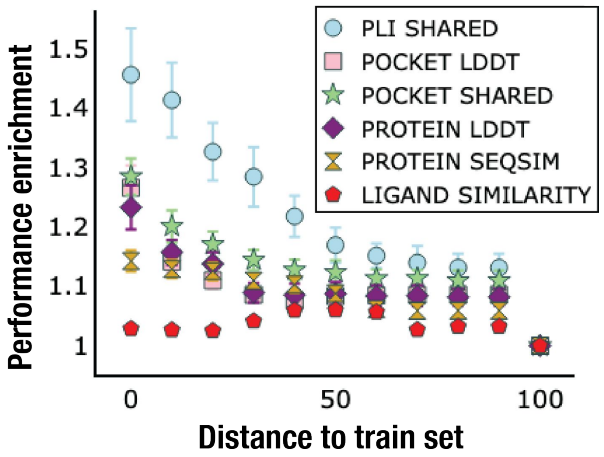
6UJH

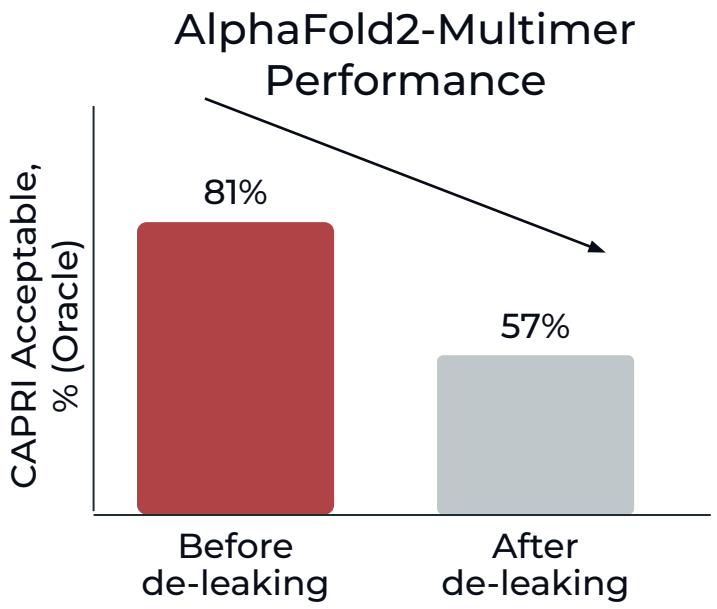
Durairaj, Janani, et al. "PLINDER: The protein-ligand interactions dataset and evaluation resource." bioRxiv (2024): 2024-07
Kovtun, Daniel, et al. "PINDER: The protein interaction dataset and evaluation resource." bioRxiv (2024): 2024-07.

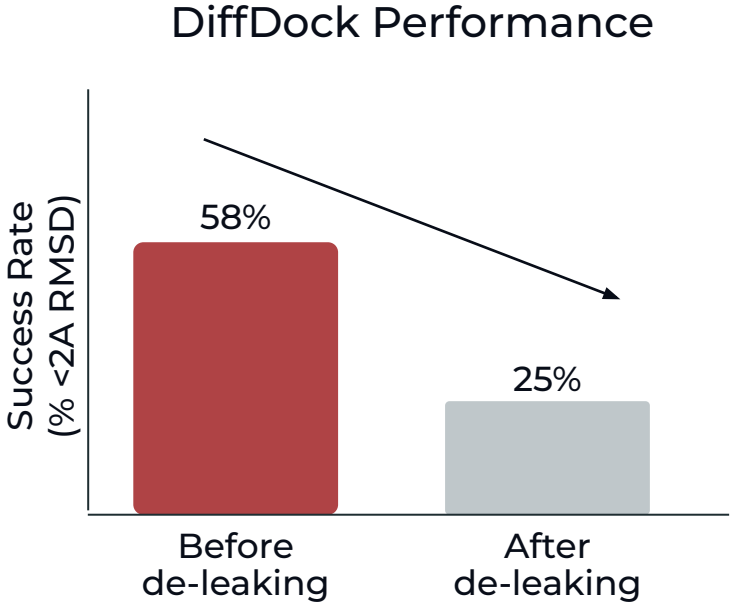
Lack of generalisation ~= memorisation

- Reported performance evaluations of models don’t account for interaction similarity
- Performance drops significantly when evaluation set is de-leaked for interactions





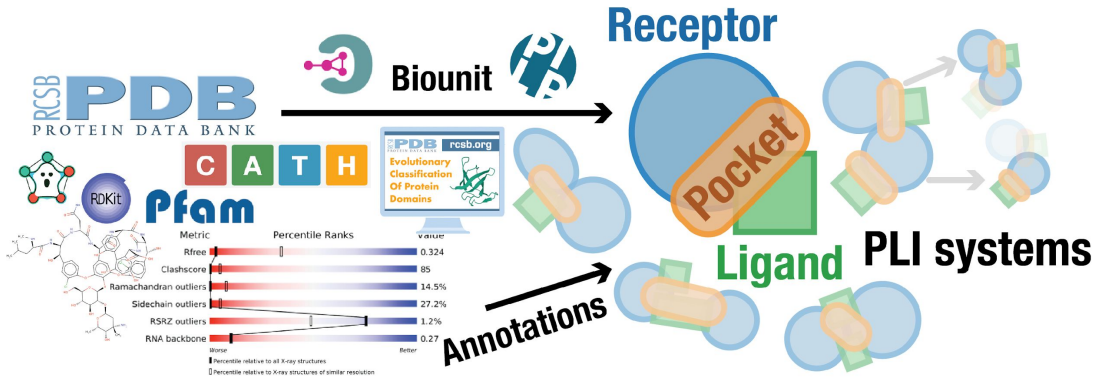




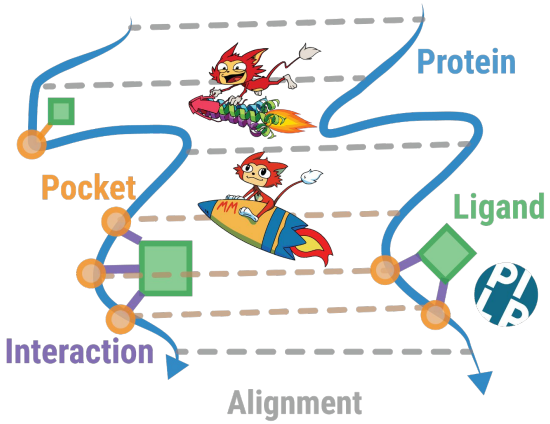
Durairaj, Janani, et al. "PLINDER: The protein-ligand interactions dataset and evaluation resource." bioRxiv (2024): 2024-07
Kovtun, Daniel, et al. "PINDER: The protein interaction dataset and evaluation resource." bioRxiv (2024): 2024-07.

PLINDER - how it solves these issues

Collect PLI complexes and annotate quality

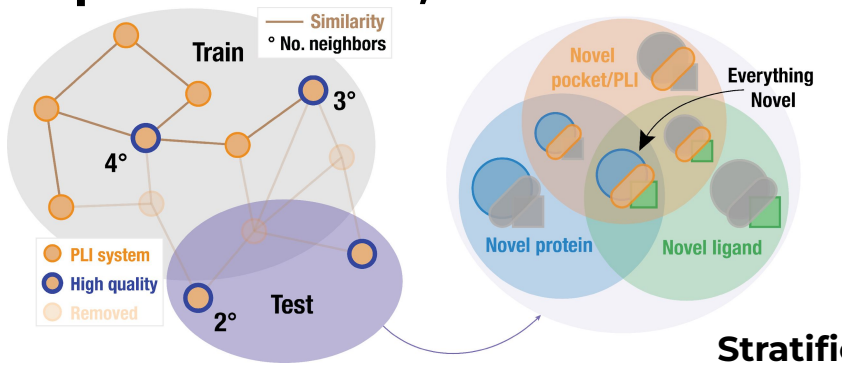


Cluster on PLI similarity



Sequence & ***structure*** similarity at **protein**, **pocket**, **ligand**, and **interaction** levels

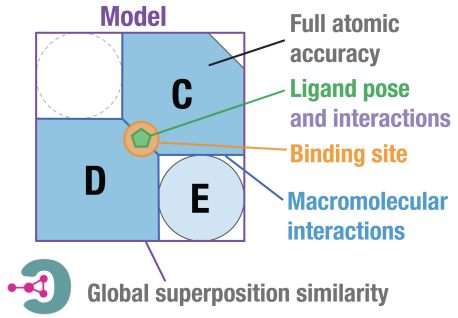
Split into train, val and test sets



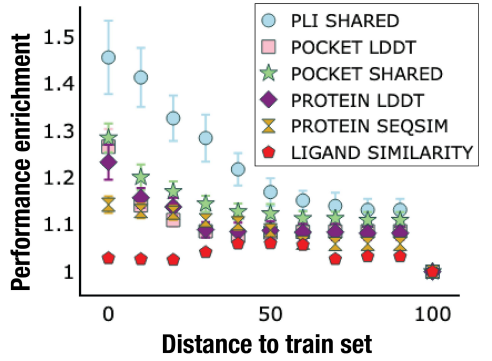
Maximizing test set **quality and diversity** while *minimizing* information **leakage**

Stratification for different tasks

Evaluate accuracy and generalisability



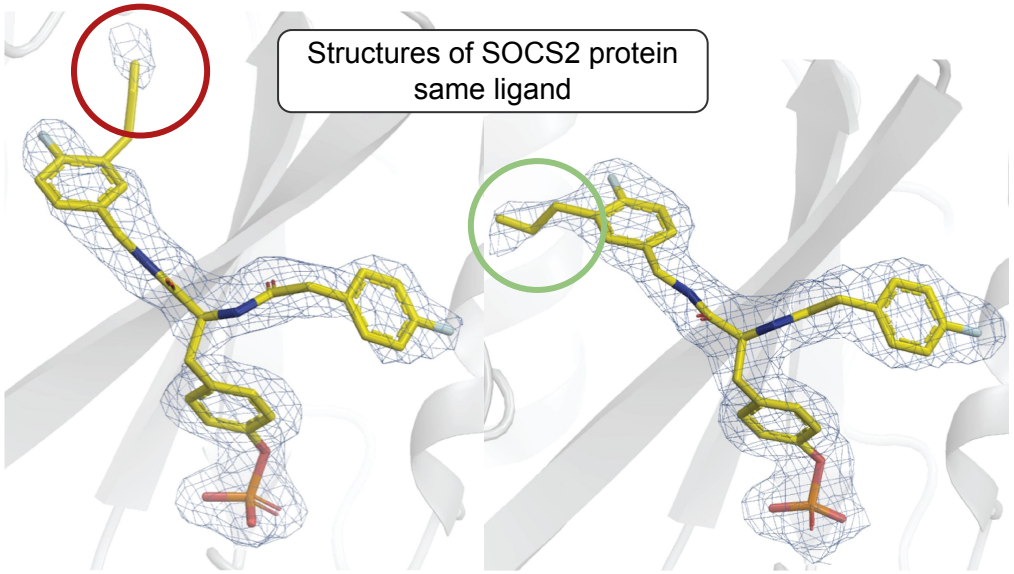
Comprehensive **evaluation** harness



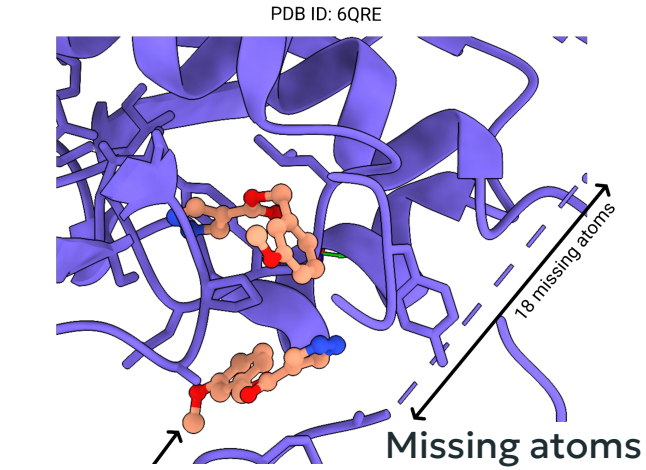
Resources to **inspect biases** in your models

Quality of the ground truth structures

- When using data from PDB validation reports only ~50% pockets pass validation criteria
- cryo-EM criteria not (yet) well defined



- BUT this data not available at pre-release
- Solutions:
  - do filtering in post-release
  - include quality into scoring



Usage restrictions of new tools = difficult to assess performance

- Chai-1

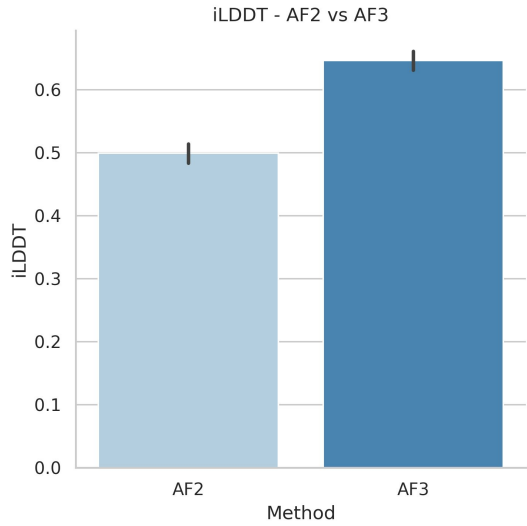
> **(d)** Output Use Restrictions. You further agree that you will not: (i) use, or permit others to use, the Services or any Output or Derived Materials in connection with any benchmarking or other comparison of the Services with any other services or technology, or (ii) upload, distribute, or otherwise provide any Output or Derived Materials to any other artificial intelligence or machine learning model, tool, neural network, or other technology for any purpose, including but not limited to training.

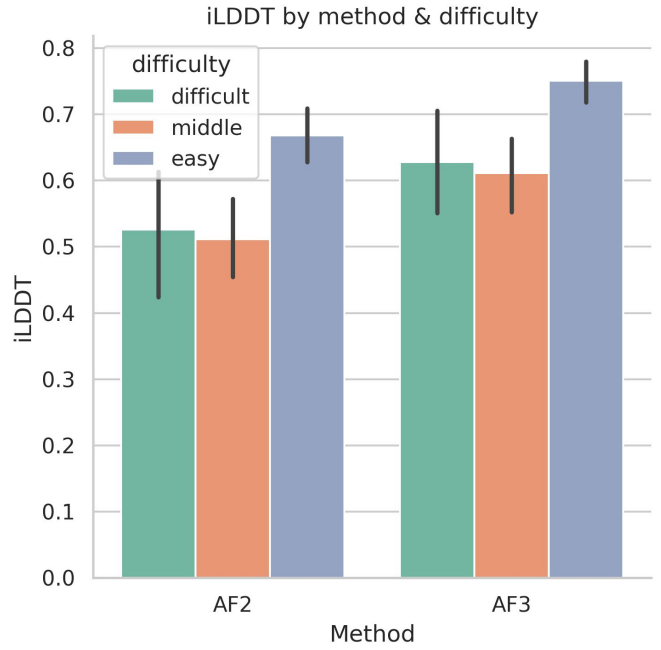
- AF3 Server:
  - Cannot be done in automated fashion
  - Restricted use of ligands
  - Both points resolved with the release of code

- Biologically common ligands: ATP, ADP, AMP, GTP, GDP, FAD, NADP, NADPH, NDP, heme, heme C, myristic acid, oleic acid, palmitic acid, citric acid, chlorophylls A and B, bacteriochlorophylls A and B
- Biologically common ions: Ca^{2+}, Co^{2+}, Cu^{2+}, Fe^{3+}, K^{+}, Mg^{2+}, Mn^{2+}, Na^{+}, Zn^{2+}, Cl^{-}

Usage restrictions of new tools = difficult to assess performance

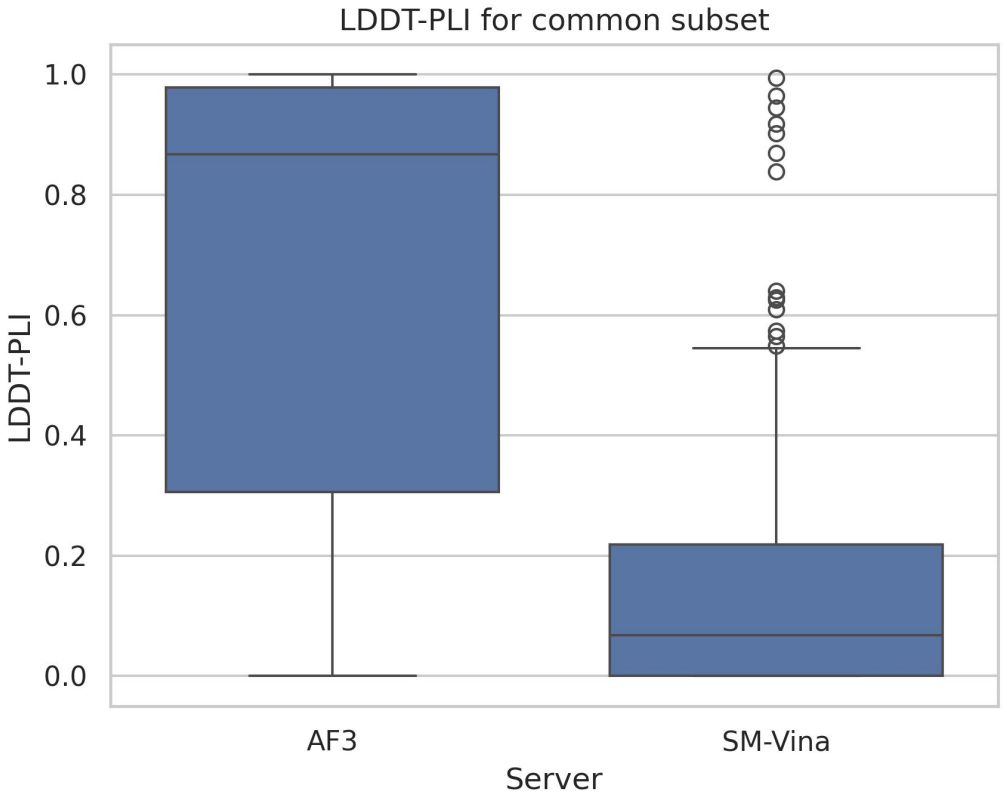
- Data for heteromers modeled with AF2 and AF3 from 2024-05-18 on
- N = 222 proteins (heteromers with 1 chain per protein)
- Difficulty based on foldseek TM-score of a complex:
  - 0-0.6: hard (N=39)
  - 0.6-0.8: medium (N=92)
  - 0.8-1.0: easy (N=91)

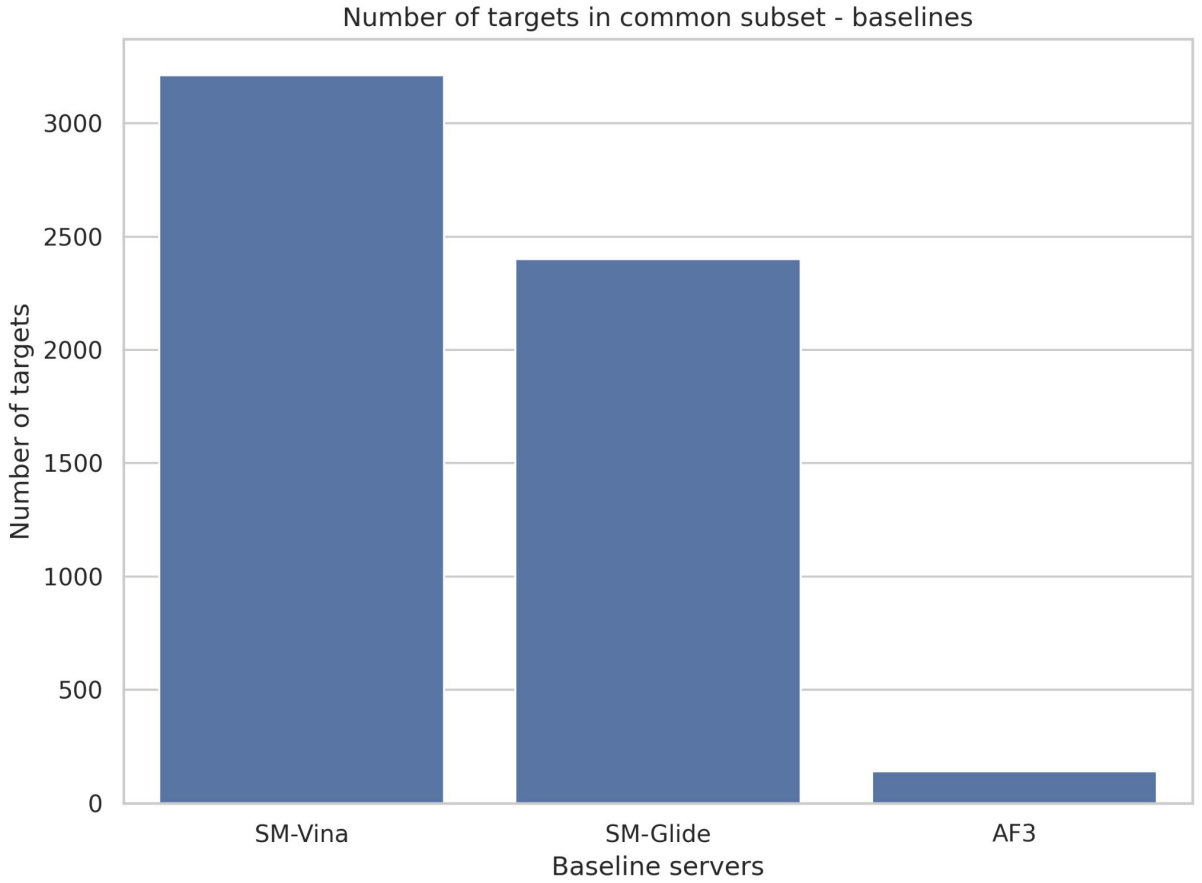




Usage restrictions of new tools = difficult to assess performance

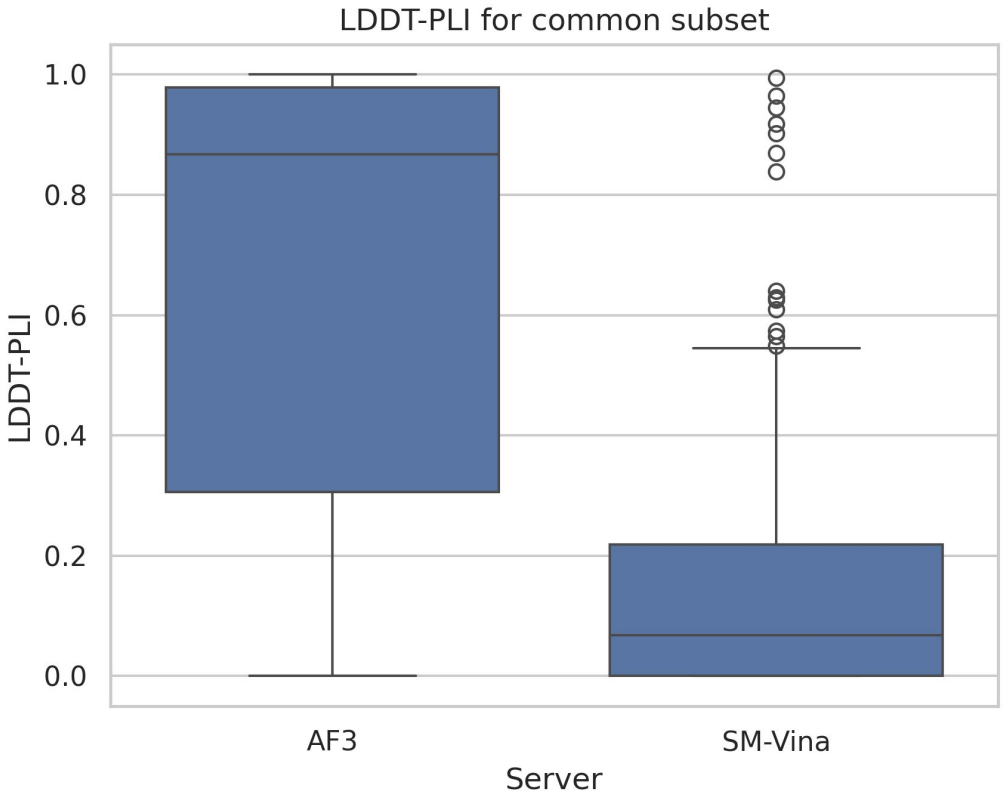
- Analysis of predictions with small ligands is inconclusive

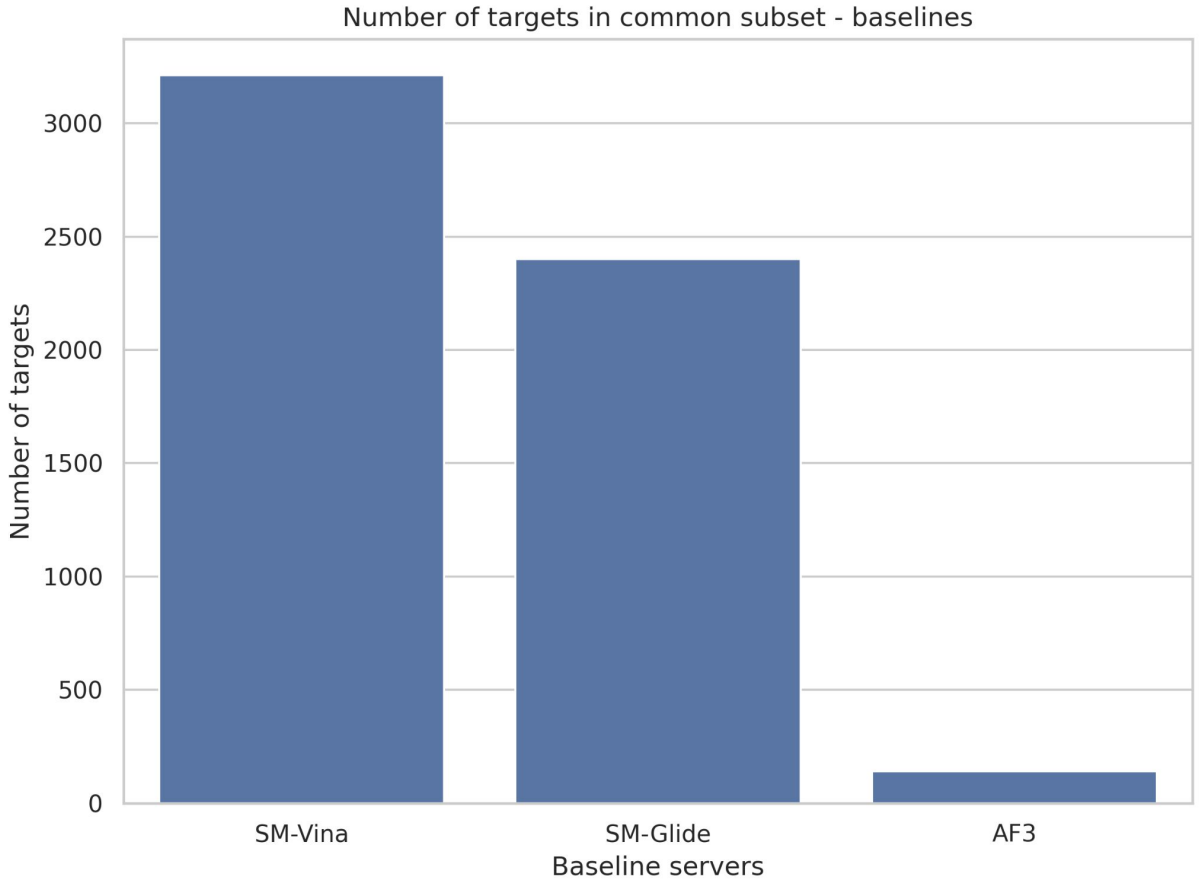




Usage restrictions of new tools = difficult to assess performance

- Analysis of predictions with small ligands is inconclusive





For more AF3 results - come to Ligand category presentation on Tuesday!

CAMEO Outlook

- Target classification:
  - Incorporation of PLINDER definitions of novelty/difficulty and quality criteria
- Web interface for results
- We’re going to migrate 3D participants to the new Structures & Complexes category (early 2025)
  - Participants will receive single sequence targets
- Including ground truth quality into scoring
- Include model confidence into evaluation
- Keep up with the modeling challenges

Register your servers

- Servers are useful for:
  - Access to everyone
  - Reproducibility
  - Open Science
- Benefits for server developers:
  - Continuous benchmarking of server performance
  - Many targets to collect statistics
  - Development servers anonymized to benchmark new features

https://beta.cameo3d.org/  help-cameo3d@unibas.ch

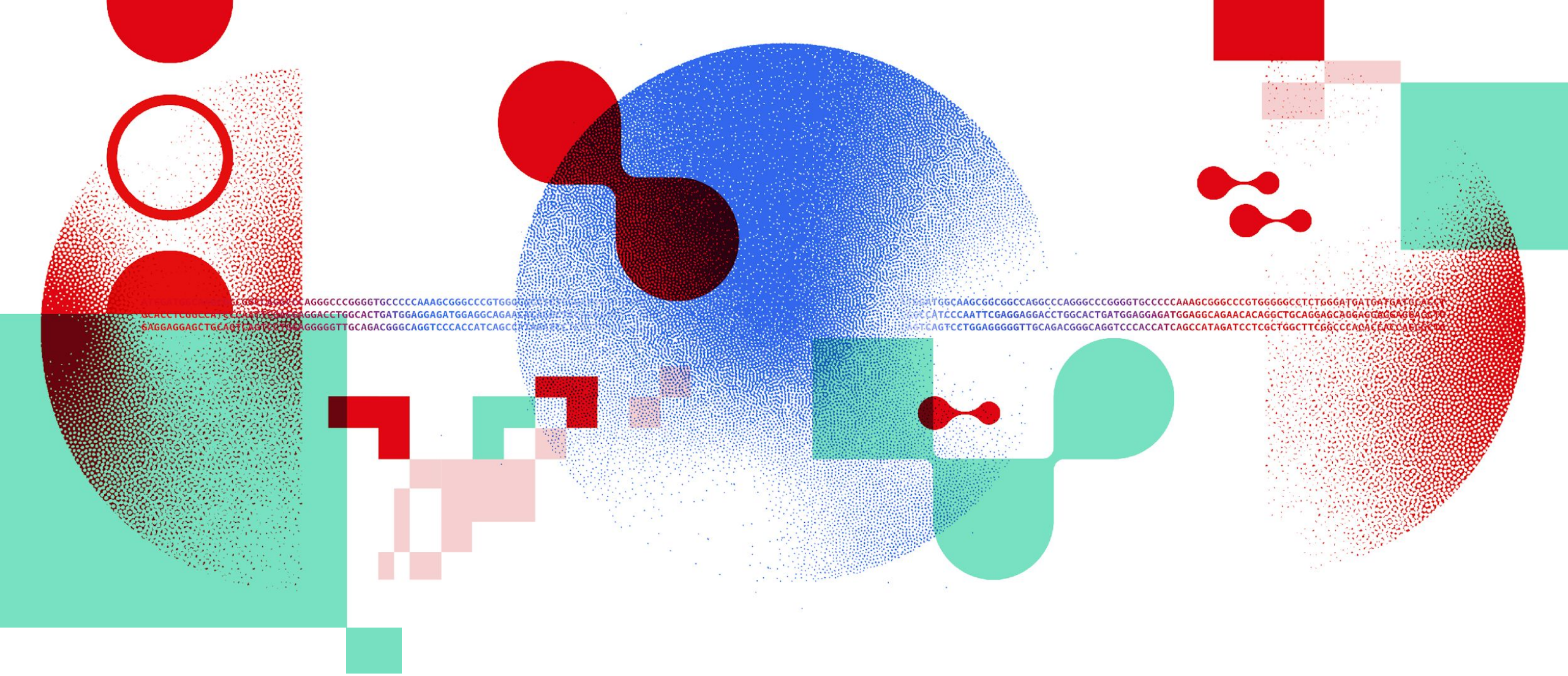
Xavier Robin

CAMEO contributors: Xavier Robin, Juergen Haas, Rafal Gumienny & many others

PDB

CAMEO participants

A. Sali, L. McGuffin, T. Schwede, J. Soeding, D. Baker, A. Fiser, M. Sternberg, Y. Zhang, C. Floudas, S. Tosatto, J. Xu, Y. Zhou, O. Brock, B. Wallner, A. Eloffson, D. Labudde, C. Venclovas, J. Cheng, O. Taştan Bishop, Y. An-Suei, T. Sosnick, C. Kaesar, P. Winn, C. Seok., S. Wang, G. Zhang, F. Wang, HeliXon, Y. Q. Gao, L. Zhang, S. Z. Li, Y. Lan



Thank you

BIOZENTRUM
Universität Basel
The Center for
Molecular Life Sciences

SIB

Target selection

- Cluster for redundancy
  - 99% sequence identity for polymer sequences
  - Complexes sharing the same set of clustered sequences
  - Ligands targets: additional clustering of complexes with the same ligands
- Select interesting prediction targets
  - BLAST templates for individual chains
  - Templates overlap across chains
    - Novel complex: no template covering all proteins
    - Medium/high difficulty based on BLAST results
  - Assign labels to interesting targets
  - “Easy” - template covering all entities with > 85% identity, coverage
  - “**Medium**” or “**hard**” otherwise
  - “**Ligands**”: easy target with novel ligands