Integrated structure prediction of protein–protein docking with experimental restraints using ColabDock

Feng et al. 2024

Luke Elder

11/7/24

Outline

- Introduction
- Methods
- Results
- Conclusions

Introduction

- Understanding the structure of protein complexes is crucial for drug discovery and antibody design
- Experimental methods exist but are costly necessitating computational methods
- Traditionally done by docking which uses a scoring function (SF) to evaluate many possible conformations given the individual structure of each protein
  - Limited by accuracy of scoring functions -> can be addressed by incorporating restraints derived from experimental methods

Docking with experimental constraints

- HADDOCK allows users to define active and passive residues in the complex which are converted into ambiguous interaction restraints
- ZDOCK utilizes contacting residues to filter docking conformations
- pyDOCK uses percentage of satisfied restraints as pseudo-energy term in SF
- ClusPro generates a feasible translation set for each restraint and selects translations with high frequency

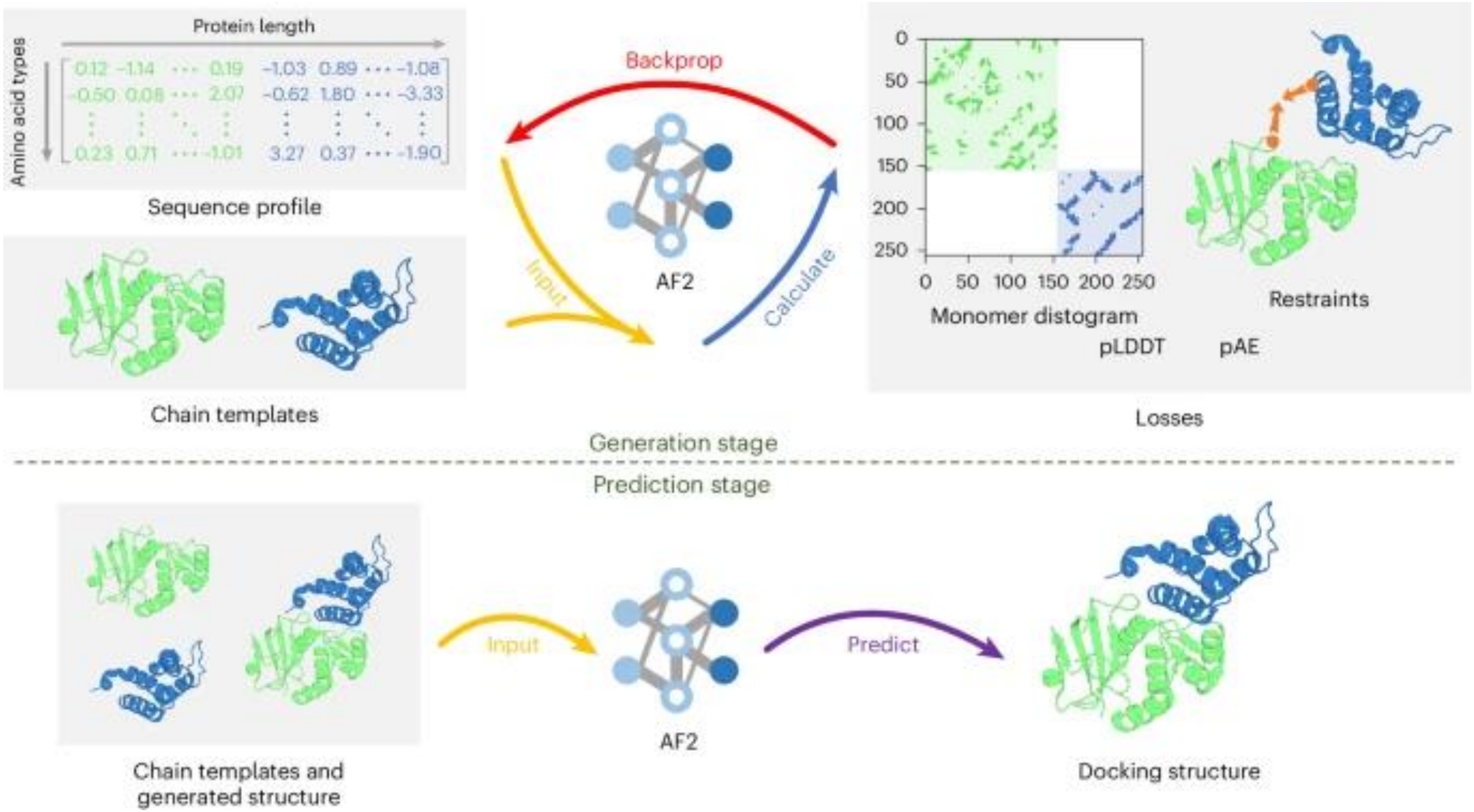
Deep learning approaches

- AlphaFold2 (AF2), AlphaFold-Multimer (AF-Multimer), and RoseTTAFold2 (RF2)
- AF2 has learned approximate biophysical energy landscape and has state-of-the-art quality estimation
- AF2 has been used for protein design but predictions are inconsistent especially for flexible protein-protein interactions

ColabDock

- The motivation of ColabDock is to incorporate experimental restraints into deep learning models to avoid inconsistency between experiment and prediction
- Use gradient backpropagation to effectively integrate prior experimental restraints and the energy landscape of data-driven structure prediction models
  - Search for conformations that satisfy both
- Two stages:
  - Generation stage – generate structure according to constraints while maximizing pLDDT and pAE
  - Prediction stage – structure predicted on basis of generated structure and templates

ColabDock workflow



ColabDock workflow

- Use AF2 with 1 recycle as structure prediction model
  - Trained on protein structures but not complex structures guaranteeing fairness in evaluation
- ColabDock performed multiple times for each complex
  - Final conformations selected by ranking support vector machine (SVM) algorithm
- AF-Multimer or RoseTTAFold2 could also be used
  - They also release ColabDock-Multimer with AF-Multimer

Restraints

- 1v1 restraints – distance of residue pair is below threshold
  - Often derived from cross-linking mass spectrometry (XL-MS)
- MvN restraints - interface level
  - Restraint between two sets of residues on surface that may be in contact
  - Methods:
    - NMR chemical shift perturbation (CSP)
    - Covalent labeling (CL)
    - Deep mutational scanning (DMS)

Synthetic Dataset

- Used for simulated restraints
  - 271 protein complexes curated from protein docking benchmark 5.5
    - Functions including enzyme-inhibitor, enzyme-substrate, antibody-antigen
  - 241 after removing structures with resolution >3 Å
  - Split into benchmark and evaluation sets
- Benchmark set – used to compare with HADDOCK and ClusPro
  - Select complexes with < 1200 residues that AF-multimer performs poorly on
  - 82 complexes - 45 are antibody-antigen complexes
- Evaluation set – used to tune hyperparameters and perform ablation
  - Development set – random sample of 30 out of 157
  - Validation set – 111 with length < 700 out of remaining 127
  - Segment set – 29 with length > 600

Restraint sampling

- 1v1 restraints are sampled from clustered residues pairs in contact
  - Collect all pairs where distance < 8 Å and cluster based on residue index distance
  - Randomly sample from different clusters – 2, 3, or 5 pairs
- MvN restraints – derived from 1v1
  - Expand each residue in restraints by incorporating 4 neighbors
  - Merge residues sampled in same chain
- Loose restraints – proteins containing restraints with large distances
  - Retrieve the inter-chain residue pairs with distances between 8 Å and 20 Å
  - Generate 1v1 restraints using same procedure as above
- Antibody interface restraints – mimic DMS data of antibody-antigen
  - Select residues in contact with other chain – sample 5-10

Experimental Datasets

- Used for experimental restraints
- CSP set – detect contact residues according to chemical shifts
  - 2 proteins
- CL set – labels side chains of solvent accessible AA with reagents which can identify residues at interface
  - 3 proteins – 1 with 3 biological assemblies

Generation stage backpropogation

$$L_{\text{total}} = 1.5 \times L_{\text{dgram}} + 2 \times L_{\text{rest}} + 0.1 \times L_{\text{pLDDT}} + 0.1 \times L_{\text{ipAE}}.$$

$$L_{\text{dgram}} = -\sum_{\text{ichain}} \sum_{i,\,j} \sum_{b=1}^{64} y_{i,\,j,b}^{\text{ichain}} \log p_{i,\,j,b}^{\text{ichain}},$$

$$L_{\text{rest}} = -\sum_{\text{irest}} \log \sum_{b:\text{dist}_b < \text{thres}} p_{\text{irest},b},$$

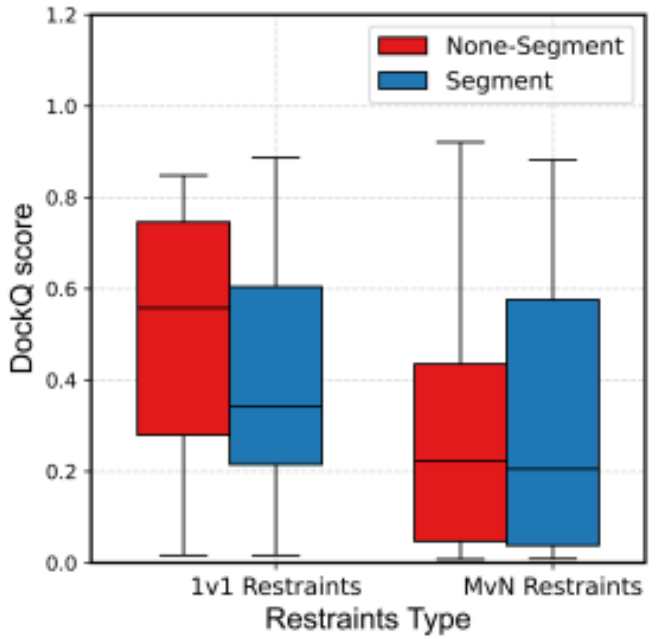
$$L_{\text{pLDDT}} = \sum_{i=1}^{L} \sum_{b=1}^{50} (\text{pLDDT}_{i,b} \times b)/(50 \times L)$$

$$L_{\text{ipAE}} = \sum_{i,\,j=1}^{L} \sum_{b=1}^{64} (\text{pAE}_{i,\,j} \times b) \Big/ \left(64 \times \sum_{i,\,j=1}^{L} \mathbb{J}\right)$$

- Distogram for each monomer – restraint loss for interaction
- Sequence profile is only thing trained
- Learning rate = 0.1

Segment based optimization

- Backpropagation uses large amounts of GPU memory
- Set of residues is first cropped out of the sequence at the beginning of each step
  - 50% probability the cropped residues contain restraints (all 4 loss terms)
  - 50% randomly chosen with no restraints (no L_rest)
  - Only profile of cropped residues are updated
- Crop 200 residues
- 100 optimization steps



Ranking algorithm

- Performance largely impacted by stochastic initialization and optimization
  - Each protein is run multiple times (rounds)
  - Structures from all rounds sorted by ranking algorithm
- Ranking algorithm built on basis of RankingSVM (RSVM)
  - 5 features: ipTM, contact number, pLDDT, number of satisfied restraints, and average error
- One RSVM selects top 5 structures for each round
- Second RSVM ranks all selected structures
- Both trained on development set

Evaluation metrics

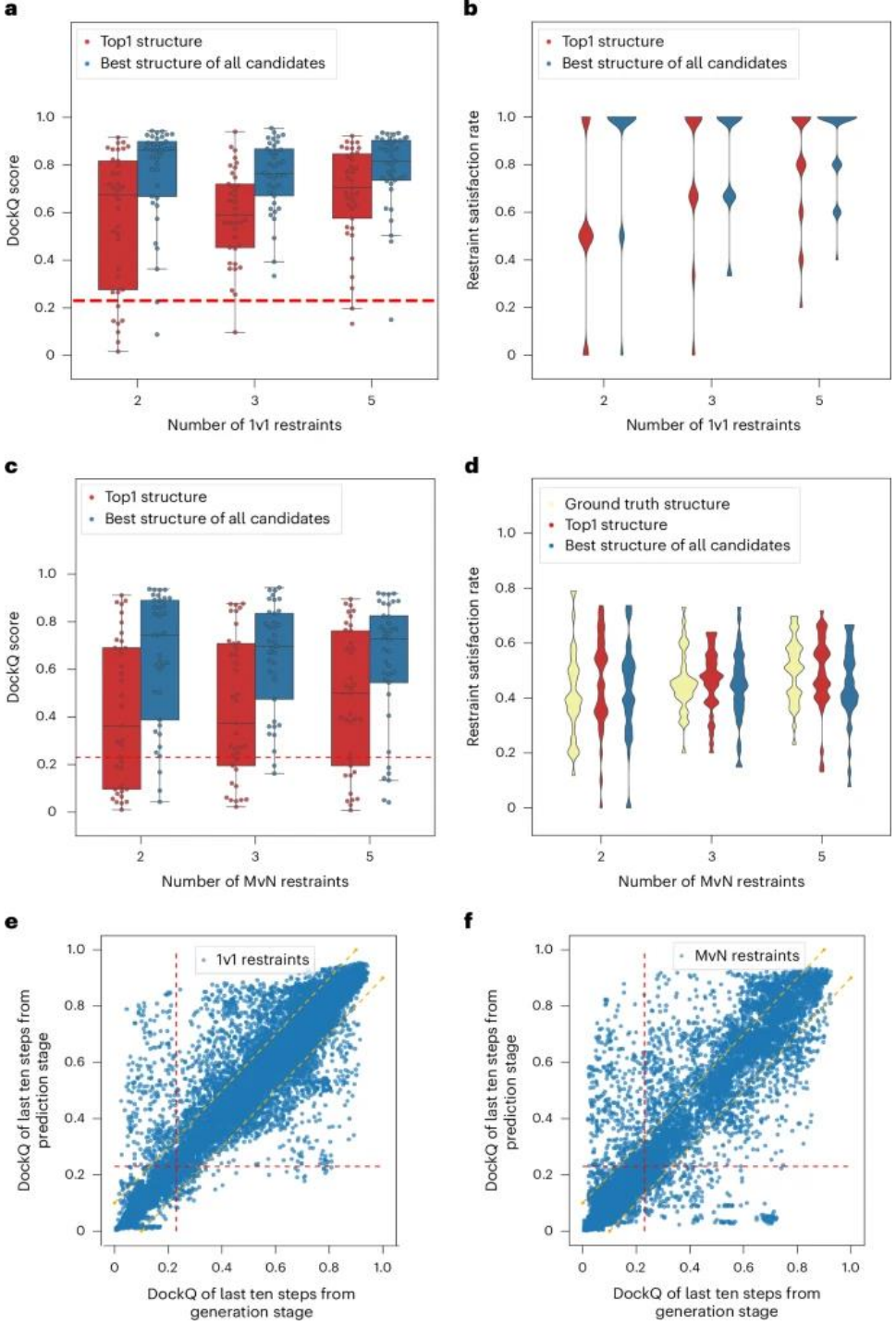
- Structure prediction
  - DockQ – focus on quality of interface – 0 to 1, DockQ > 0.23 is correct

$$DockQ(F_{nat}, LRMS, iRMS, d_1, d_2) = (F_{nat} + RMS_{scaled}(LRMS, d_1) + RMS_{scaled}(iRMS, d_2))/3$$

  - Cα-r.m.s.d. – global structure
- Restraint satisfaction rate

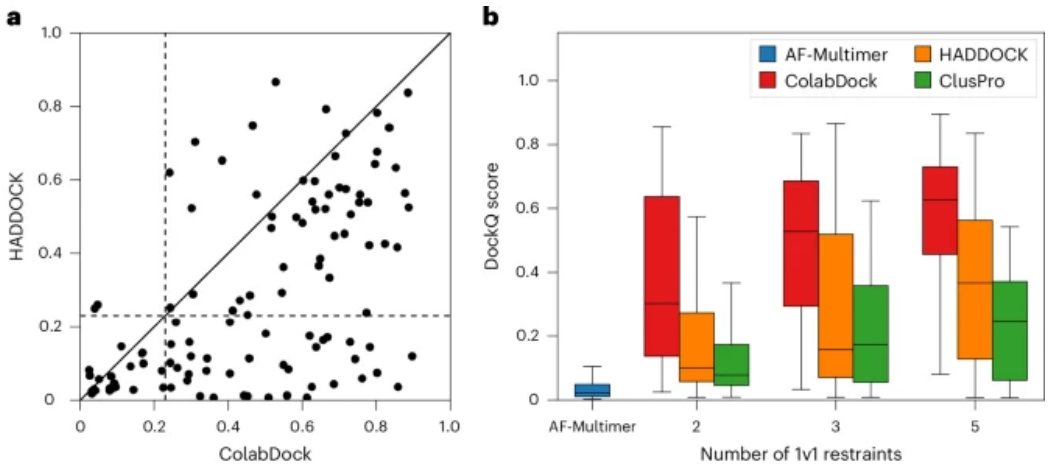
Validation set performance

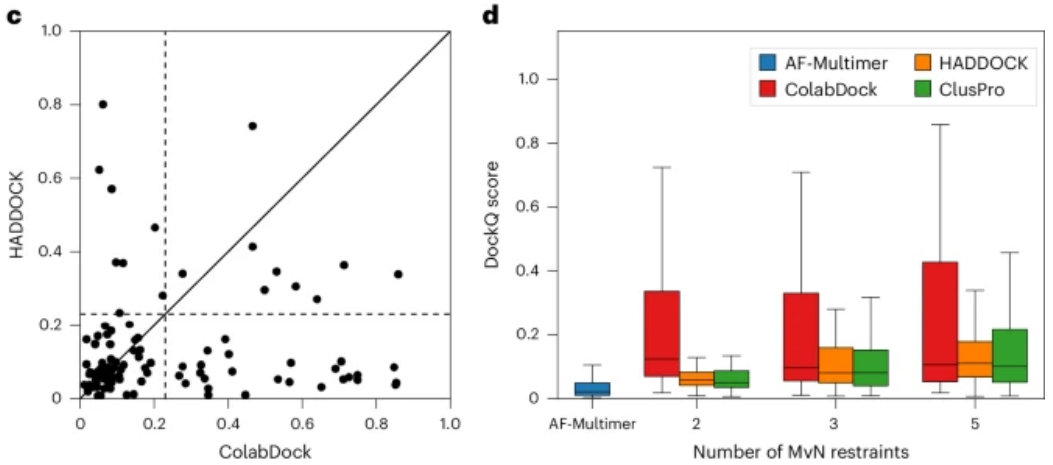
- Set of 111 complexes
- 37 samples for each level of restraints
- Prediction stage performs better on ~69% of complexes

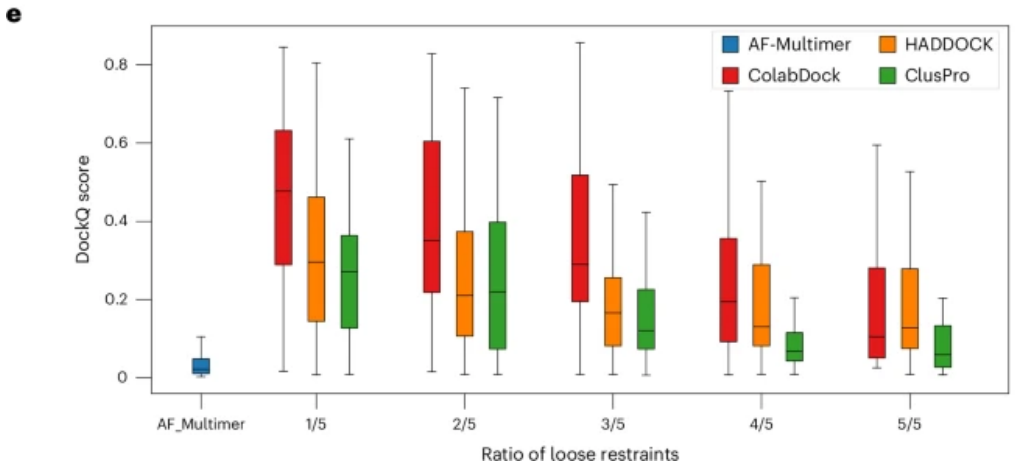


Comparison with restrained docking methods

- 37 complex benchmark set
- (a) top1 structures with 1v1
- (c) top1 structures with MvN
- ColabDock-Multimer outperforms AF-Multimer but is worse than ColabDock
  - Simple explanation

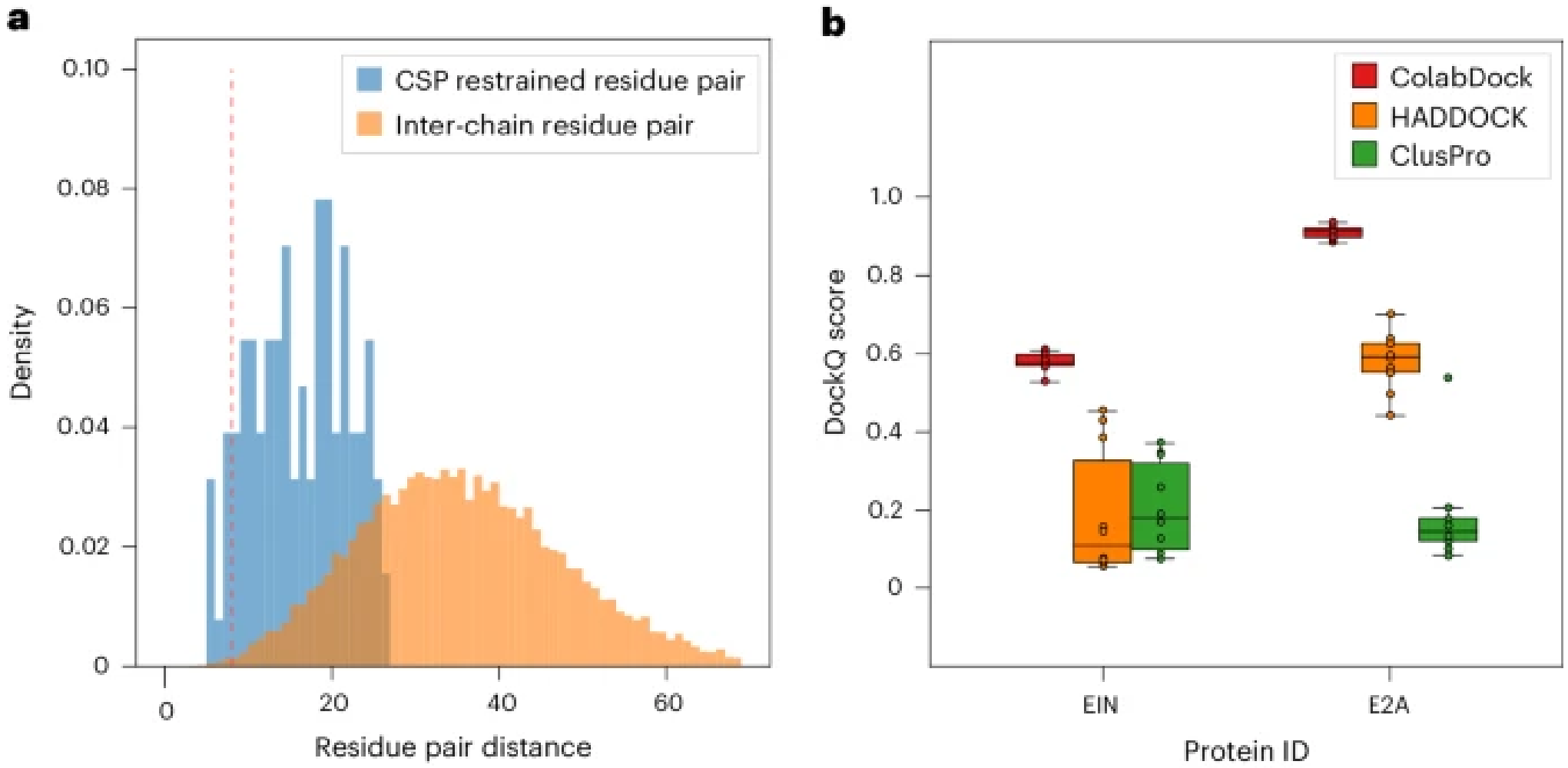






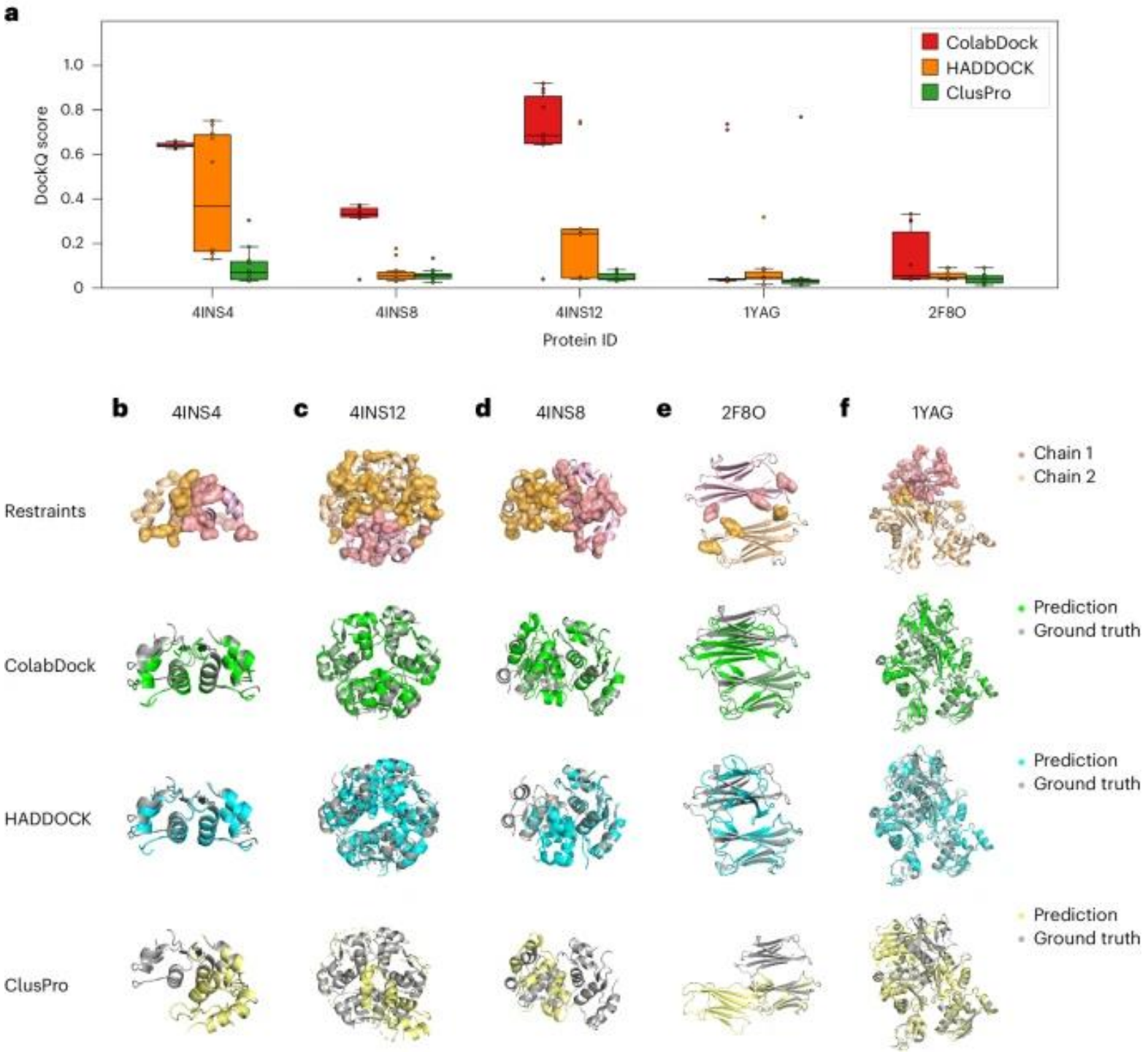
CSP restraints

- CSP uses NMR to provide range of residues located at interface



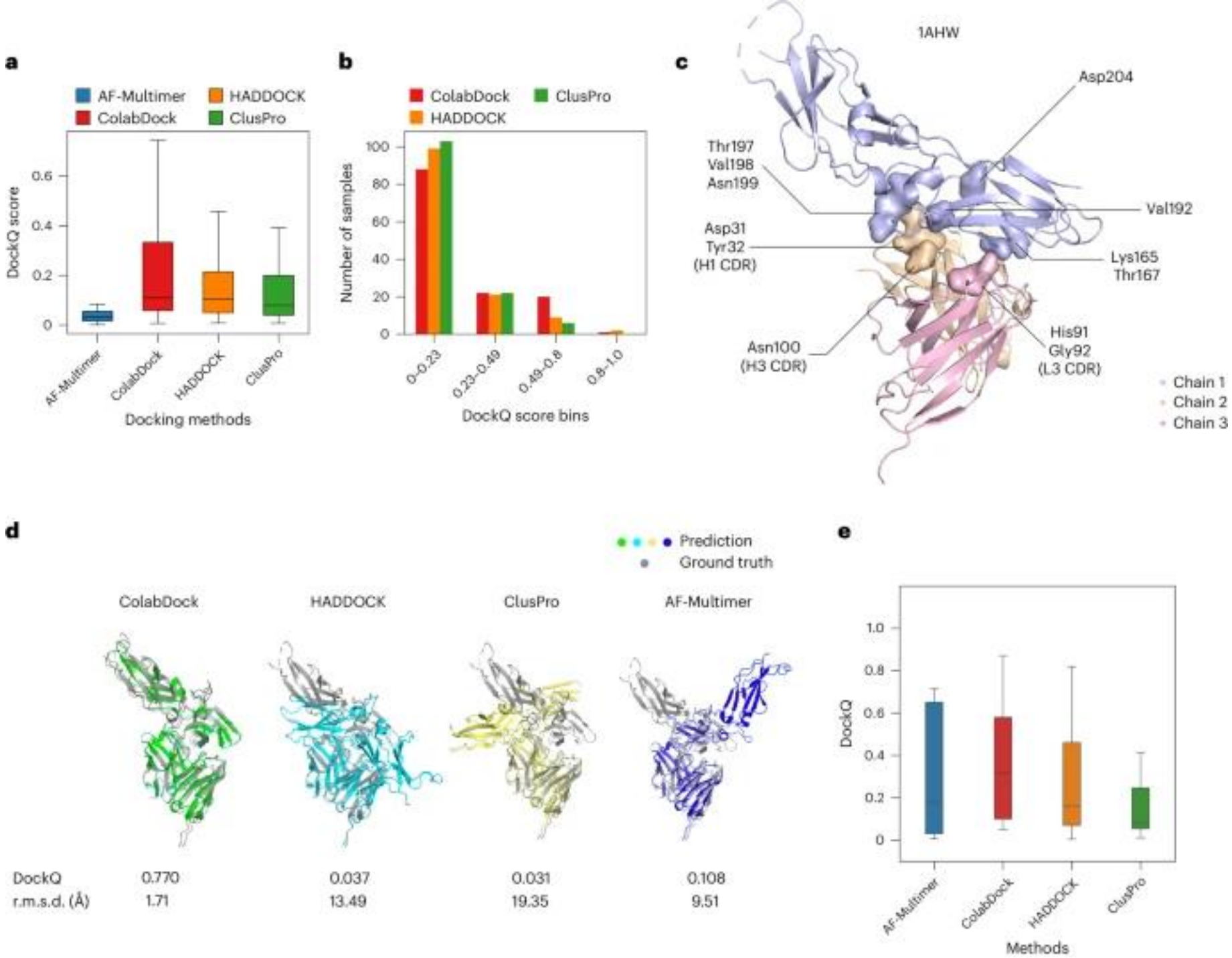
CL restraints

- Labels side chain of residues with reagents
  - residues with substantial modification more likely to be on interface
  - Weaker than CSP restraints

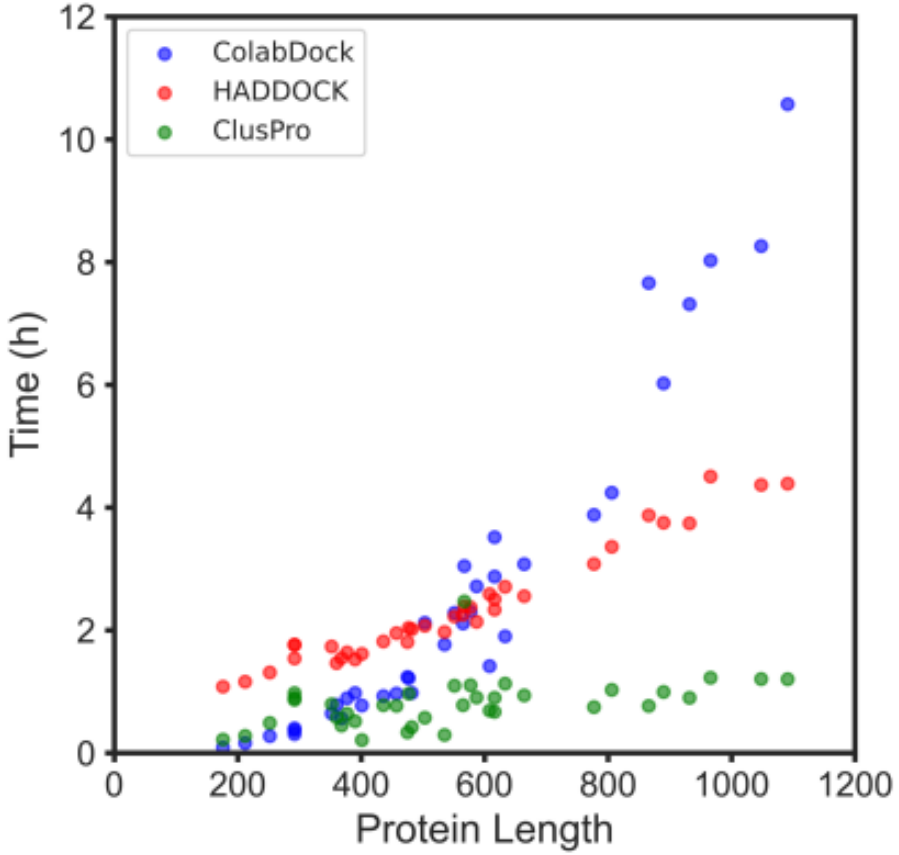


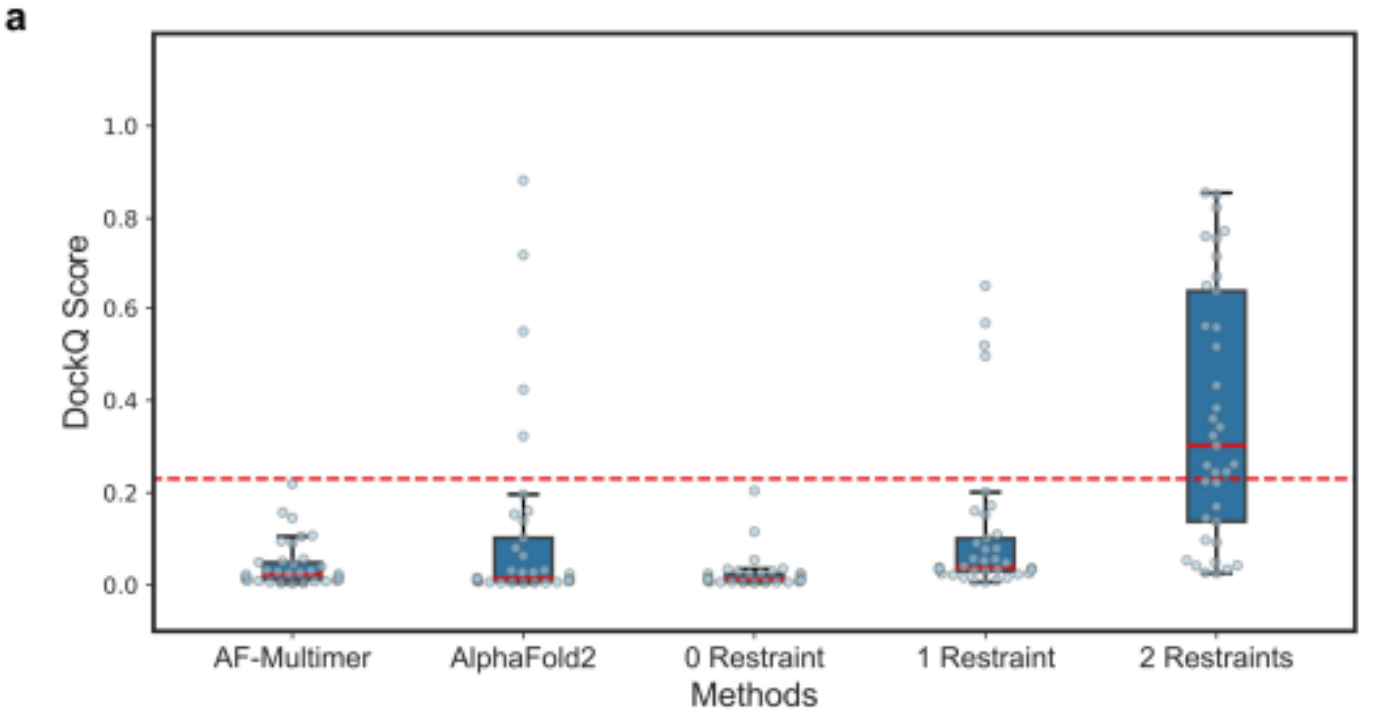
Antibody-antigen complexes

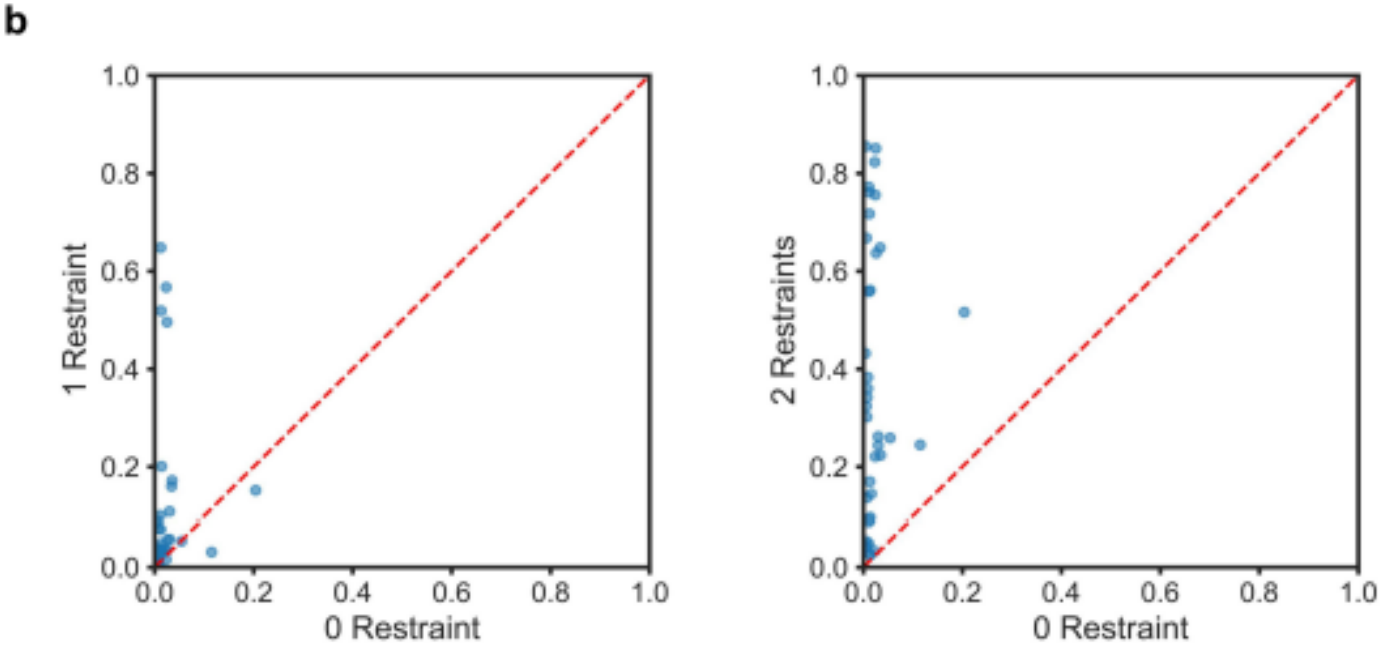
- Simulate DMS restraints for 45 complexes in antibody-antigen benchmark set
- (e) unbiased set of 8 antibody-antigen complexes



Other results







Conclusion

- ColabDock is able to effectively incorporate experimental restraints with deep learning methods to improve protein docking
- ColabDock-Multimer shows that transferability of framework
  - Performs significantly better than AF-Multimer (unfair comparison)
- The more restraints it has, the better it performs
  - Performs much better than other methods on synthetic and experimental restraints

Future directions and limitations

- In the future ColabDock can be extended to other docking tasks
  - Protein-ligand, protein-RNA/DNA
- Limitations:
  - Only accept restraints with distance below 22 Å (AF2 distogram limit)
  - Without segment-based optimization can only handle complexes less than 1,200 residues on NVIDIA A100 GPU
  - Method is very time consuming
- Published after AF3 but no comparisons at all

Questions?