# Unraveling Protein Structural Motifs across the Protein Universe with

# FOLDDISCO

Hyunbin Kim[1,2], Rachel Seongeun Kim[1,2], Milot Mirdita[2] and Martin Steinegger[1,2,*]

Personal website

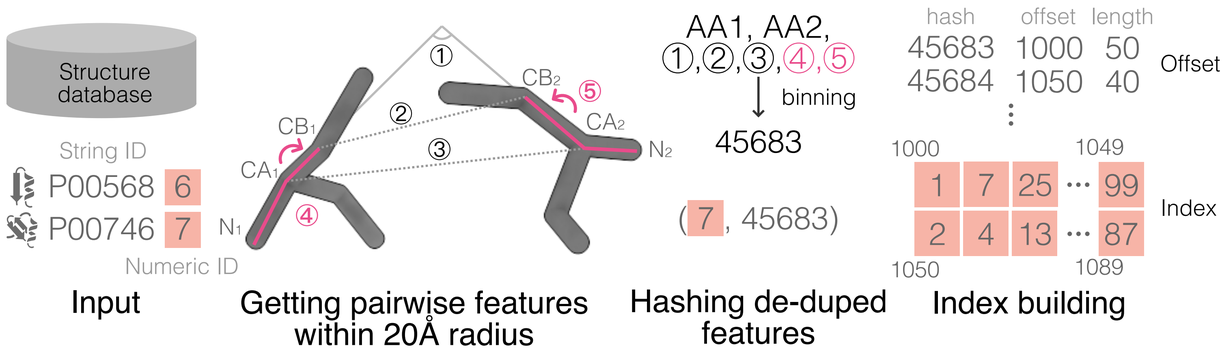
[1]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea
[2]School of Biological Sciences, Seoul National University, Seoul, Republic of Korea

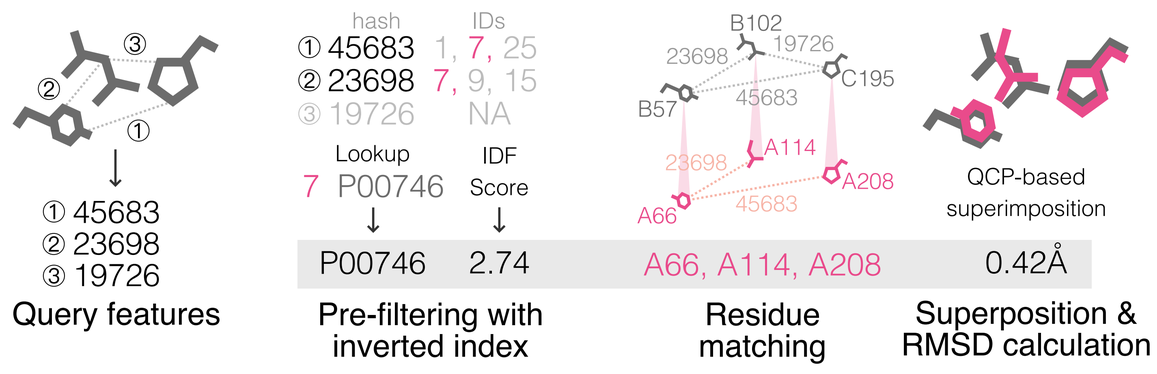
## CURRENT MOTIF SEARCH METHODS STRUGGLE TO SCALE

Protein structural motifs are recurring patterns involved in conserved functions such as small-molecule binding, enzyme catalysis, and interaction partner recruitment. These motifs often occur in weakly conserved sequences or as discontinuous patches, challenging detection by sequence aligners. Graph-based approaches address this problem but suffer limited scalability, making them inapplicable to the AlphaFold2 structural data explosion.

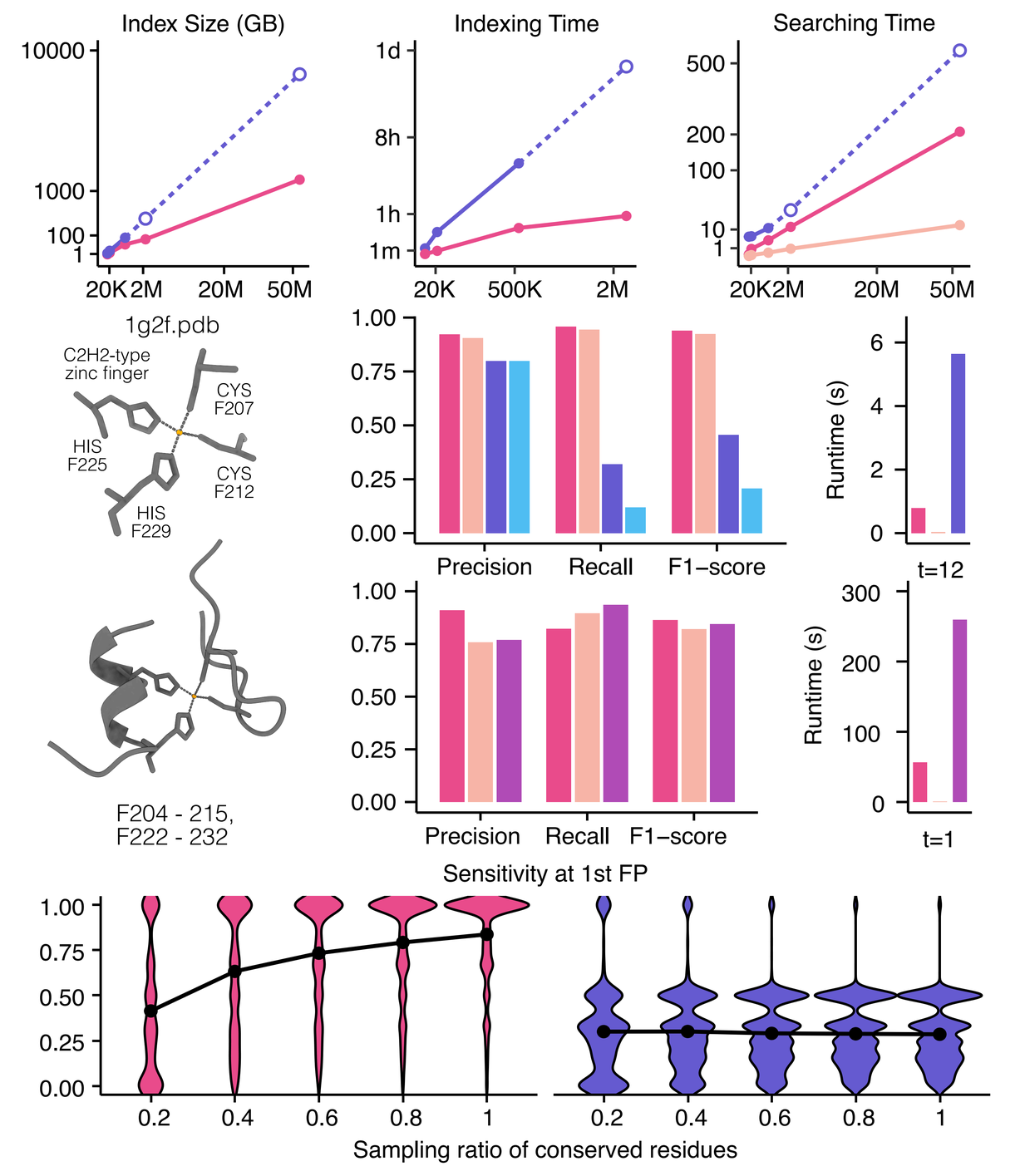
## HOW FOLDDISCO WORKS

- optimizing the index without location information
- introducing a new pairwise feature to capture side chain orientation
- scoring hits by weighting rarity of features.

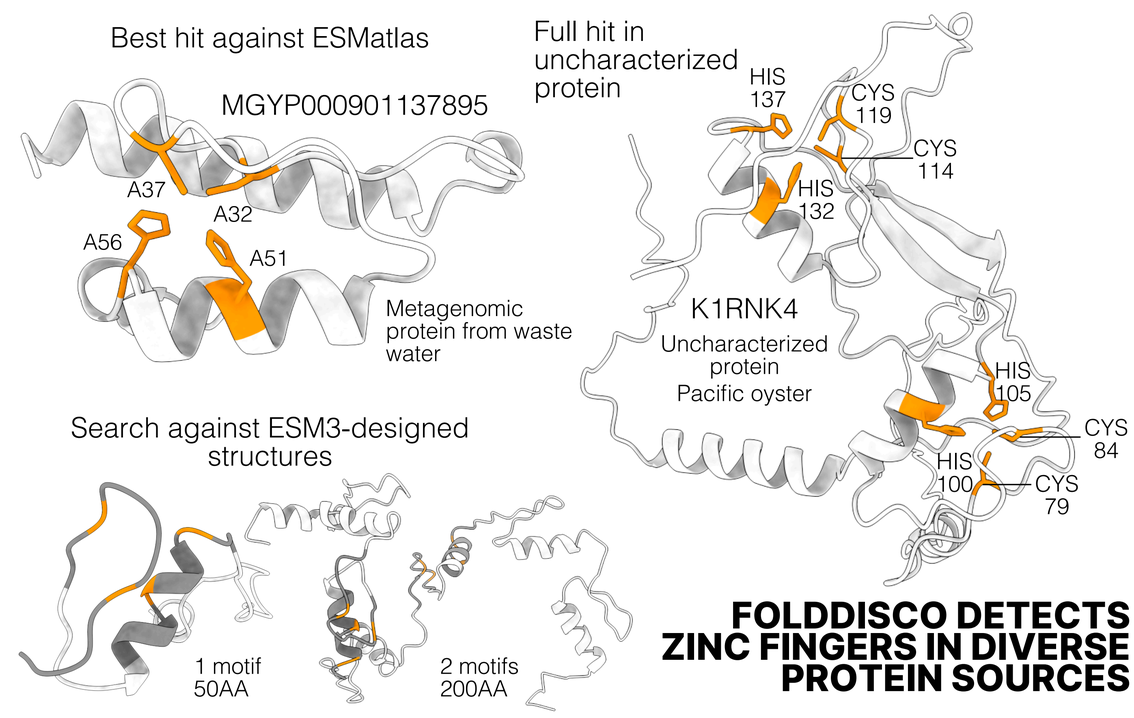



## FOLDDISCO ENABLES ACCURATE MOTIF SEARCH AT SCALE

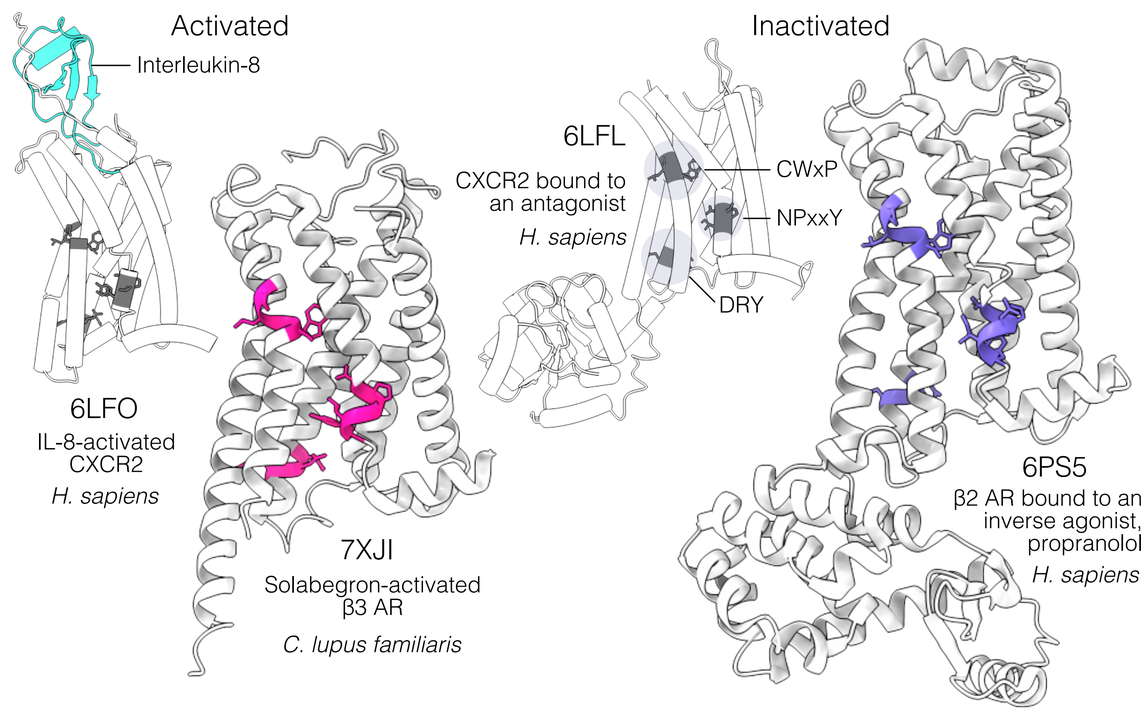



Folddisco achieved higher F1-scores compared to pyScoMotif (discrete motif) or MASTER (discontinuous segment) in querying zinc finger motifs.
We also demonstrate Folddisco's capability for searching SCOPe domains with conserved residues identified with FoldMason.

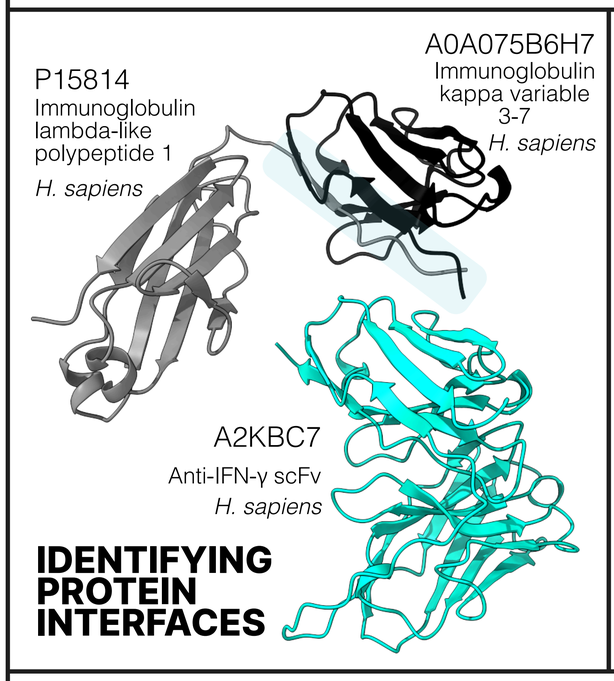



## FOLDDISCO DETECTS ZINC FINGERS IN DIVERSE PROTEIN SOURCES

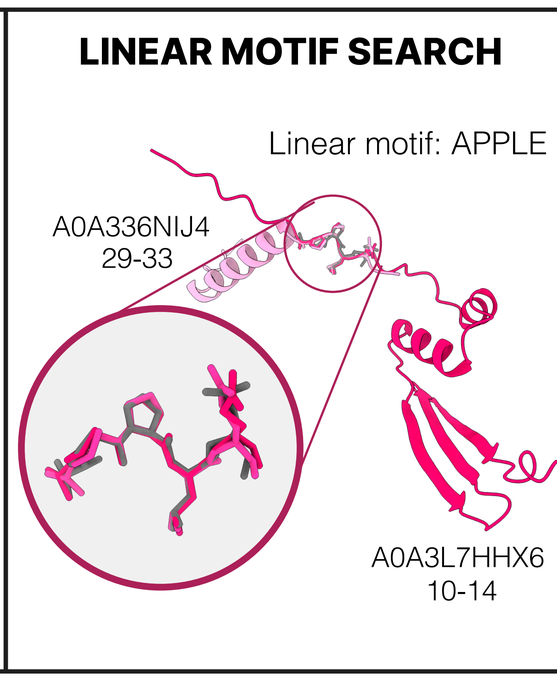
## MOTIF-BASED DETECTION OF GPCR CONFORMATION BY FOLDDISCO



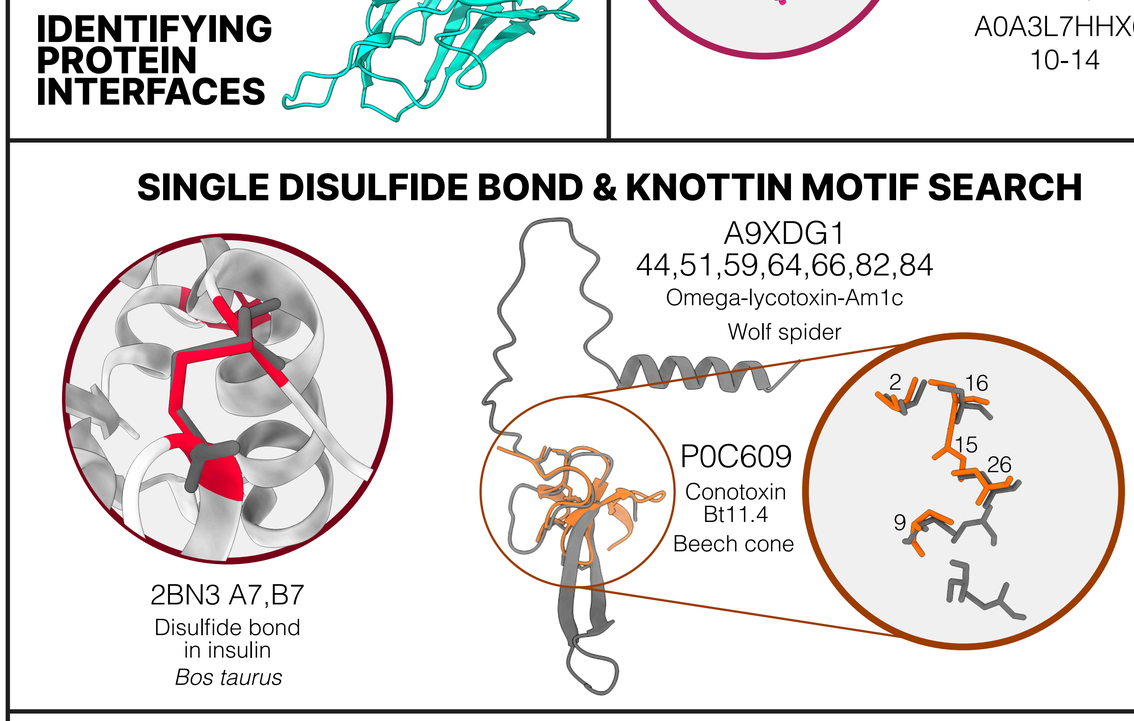

## IDENTIFYING PROTEIN INTERFACES




## LINEAR MOTIF SEARCH




## SINGLE DISULFIDE BOND & KNOTTIN MOTIF SEARCH




Reference

- Bittrich, Sebastian et al. PLoS computational biology, 2020.
- Cia, Gabriel, et al. Bioinformatics Advances, 2023.