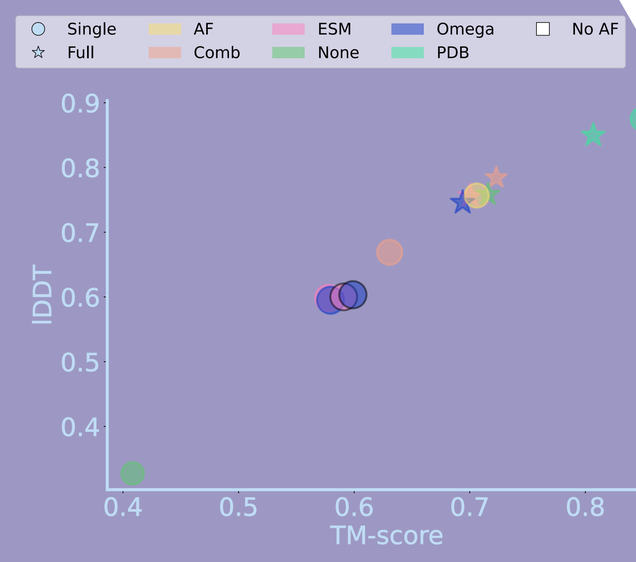
Aggregating and optimising AlphaFold2 structure predictions with templates

Jannik Adrian Gut, Thomas Lemmin
Institute of Biochemistry and Molecular Medicine, University of Bern
Graduate School for Cellular and Biomedical Sciences
jannik.gut@unibe.ch, thomas.lemmin@unibe.ch

Monomer refinement

Goal: Refine structural templates using AF2
Input: Single sequence and template from diverse sources



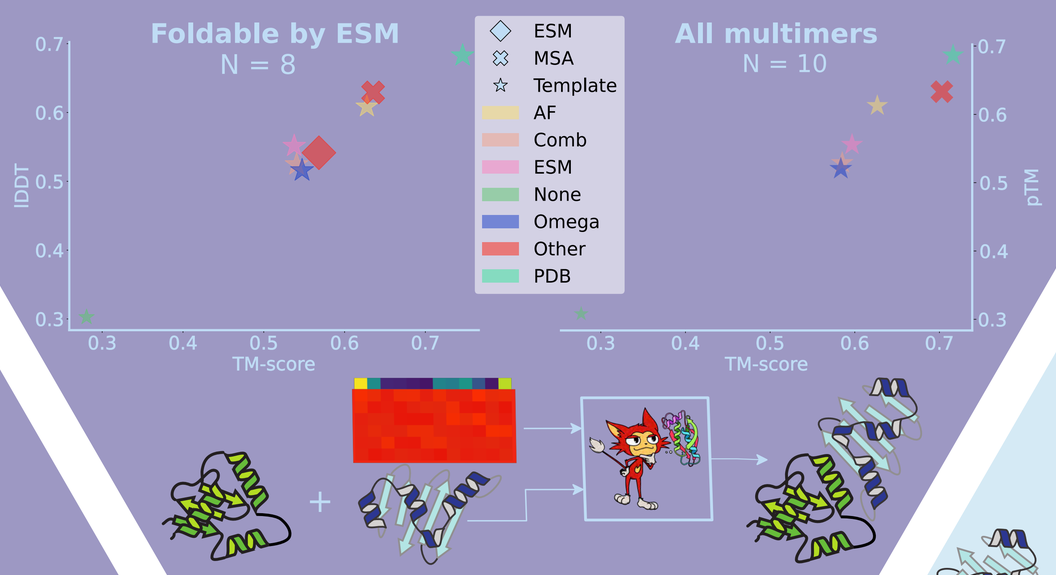
CASP14 monomers

No improvement of poor predictions or enhancement of accurace ones
No degradation in the quality of accurate predictions

CASP14 multimer refinement

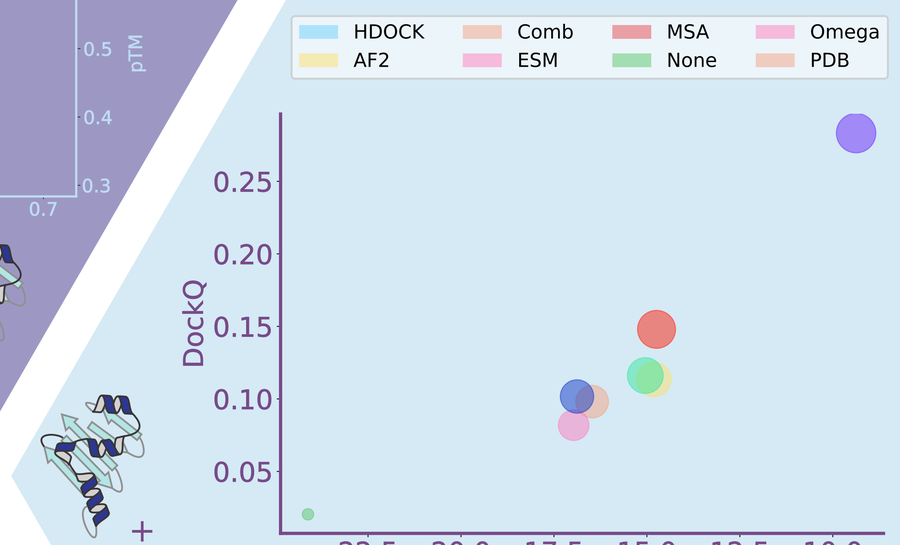
Goal: Predict protein complexes with AF2
Input: Single sequences and monomer templates from diverse sources

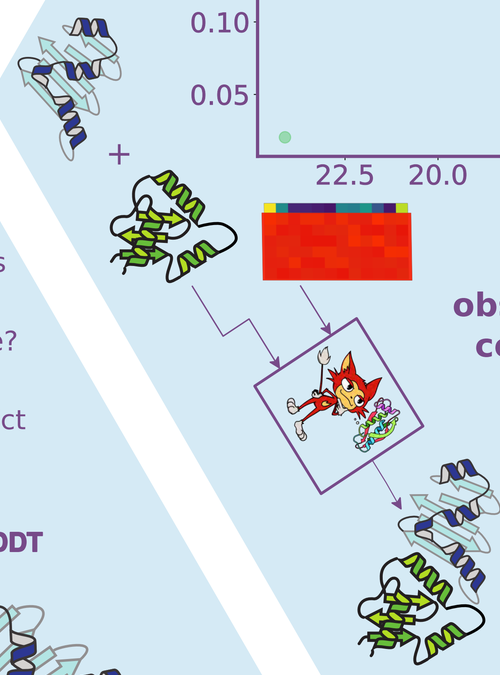
Proficient recovery of protein multimers when provided with structural templates derived from oligomeric states; however, performance deteriorates when templates deviate from the oligomeric conformation



Protein-protein interaction

Goal: Rigid body protein blind docking
Input: Single sequences and monomer templates from DIPS[4]
DIPS 100 dataset*





Limited success observed in recovering complexes, especially when compared to a multimer setting, likely attributed to the dynamic and transient nature of the interactions

AlphaFold 2 (AF2)

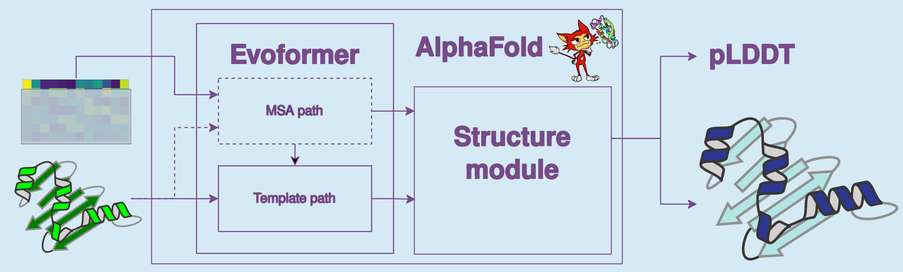
Current state of the art for protein folding
Can be used to rank quality of prediction [1]
Relies on Multiple Sequence Alignment (MSA) for predictions

Research question

To what extent does AlphaFold 2 understand protein structure?

Experimental approach

Evaluate accuracy and limitations by analyzing ability to predict protein structures using single sequence and template(s)

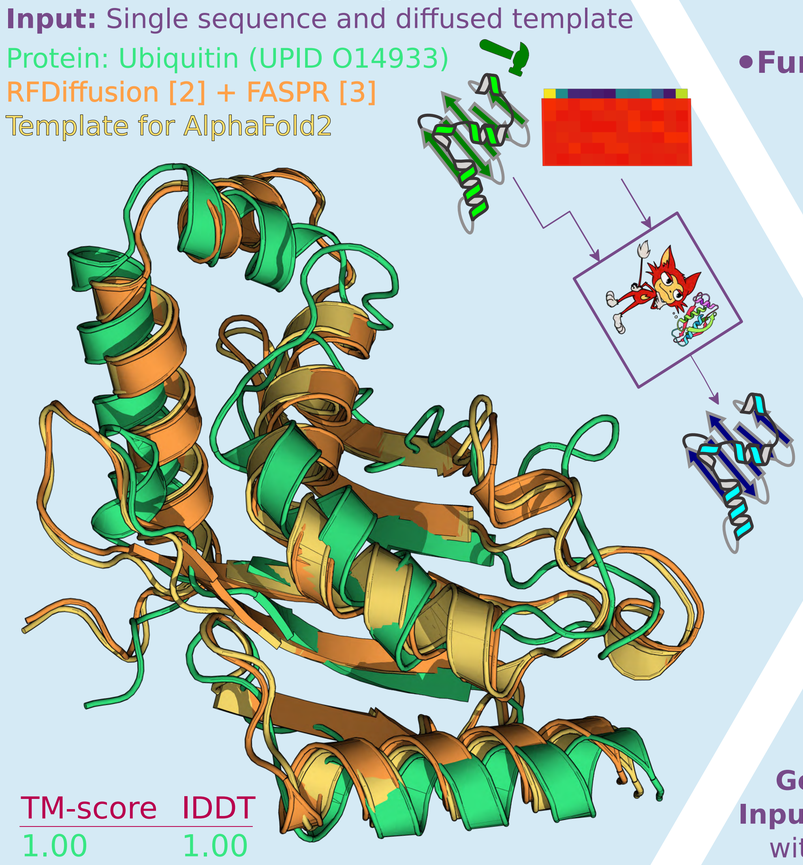


•AlphaFold 2 has difficulties in managing structural variations, indicating limitations in capturing dynamic molecular transformations
•Successful packing of side chains suggests a better understanding of local structural features
•Further supported by good performance for protein docking when provided with perfect templates
•AlphaFold 2 has learnt a valid machine learning energy function, yet it has not fully mastered a robust and comprehensive physical energy function
•Without MSA information, the local neighborhood to unroll a prediction to a better structure is prohibitively narrow

Recovery

Goal: Recover original structure post-alteration through diffusion
Input: Single sequence and diffused template

Protein: Ubiquitin (UPID O14933)
RFDiffusion [2] + FASPR [3]
Template for AlphaFold2

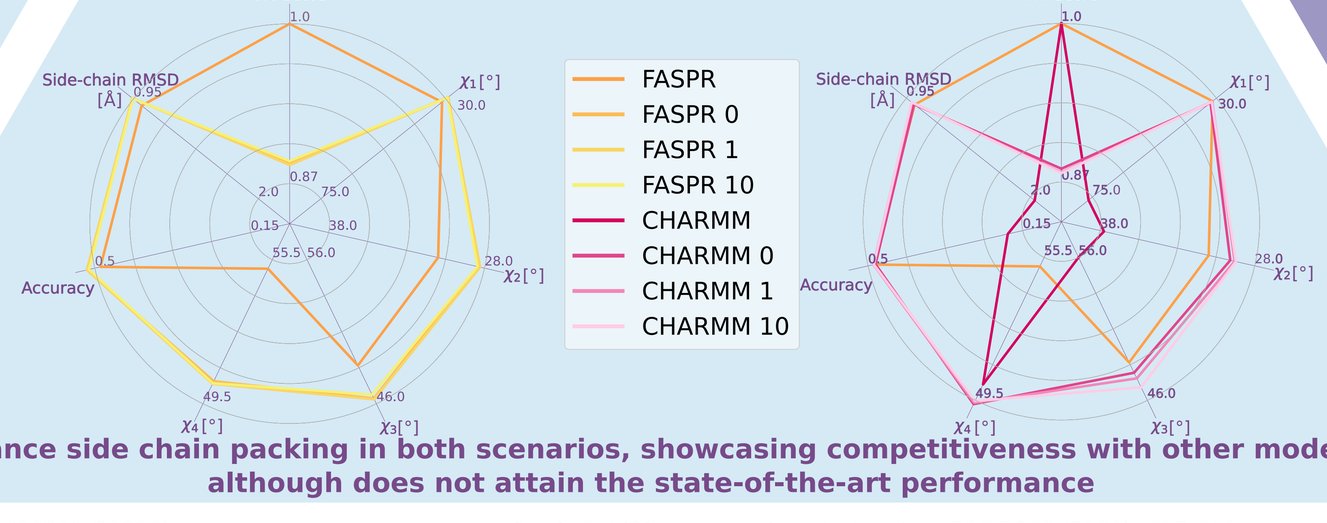


TM-score	IDDT
1.00	1.00
0.69	0.49
0.71	0.54

Incapable of restoring the structure after alteration through RFdiffusion

Side-chain packing

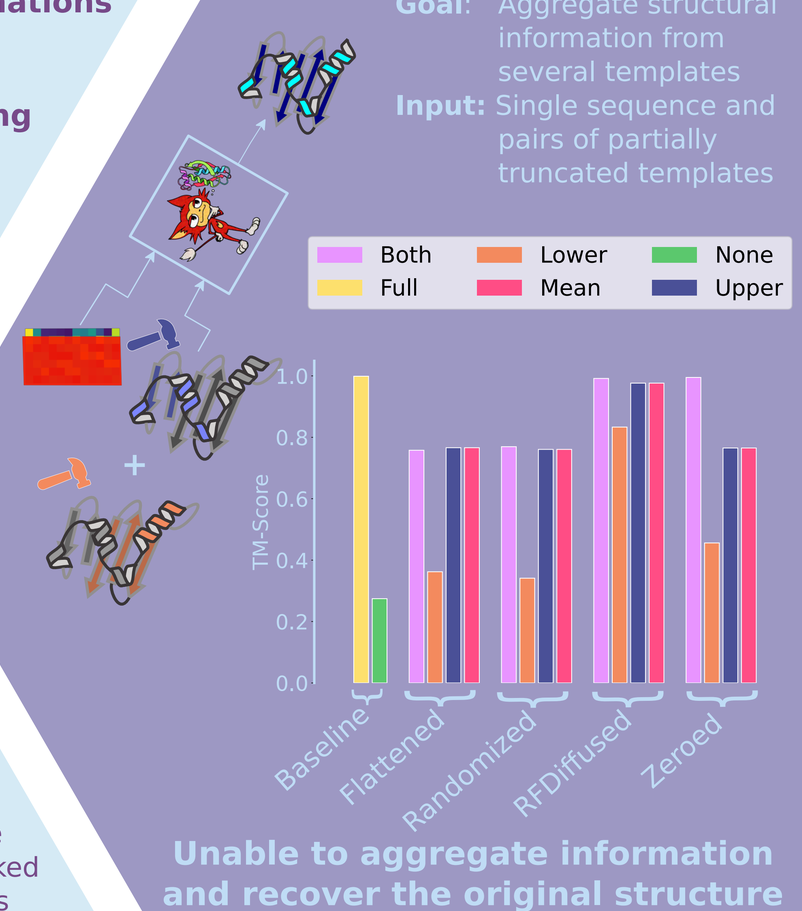
Goal: Side chain packing assessment of AF2 using a provided backbone
Input: Single sequence and a template from CASP14 with side chains packed with FASPR [3] or reconstructed from CHARMM36 [5] internal coordinates



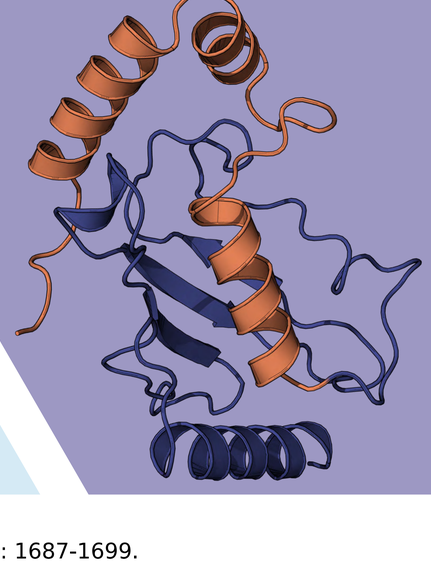
Enhance side chain packing in both scenarios, showcasing competitiveness with other models, although does not attain the state-of-the-art performance

Template aggregation

Goal: Aggregate structural information from several templates
Input: Single sequence and pairs of partially truncated templates



Unable to aggregate information and recover the original structure



[1] Roney, James P., et al. Physical Review Letters 129.23 (2022): 238101.
[2] Watson, Joseph L., et al. Nature 620.7976 (2023): 1089-1100.
[3] Huang, Xiaoqiang, et al. Bioinformatics 36.12 (2020): 3758-3765.
[4] Townshend, Raphael, et al. Advances in Neural Information Processing Systems 32 (2019).
[5] Best, Robert B., et al. Journal of chemical theory and computation 8.9 (2012): 3257-3273.
[AlphaFold2] Jumper, John, et al. Nature 596.7873 (2021): 583-589.
[CASP] Pereira, Joana, et al. Proteins: Structure, Function, and Bioinformatics 89.12 (2021): 1687-1699.
[ColabFold] Mirdita, Milot, et al. Nature methods 19.6 (2022): 679-682.
[ESM] Lin, Zeming, et al. BioRxiv 2022 (2022): 500902.
[Omega] Wu, Ruidong, et al. BioRxiv (2022): 2022-07.

Thanks to Symela and Noah for design help :)