

BeyondFold: Advancing structural biology with computational approaches

The field of structural biology is being revolutionized with the development of accurate protein structure prediction methods like AlphaFold2 and RoseTTAFold. Our aim is to integrate the strengths of SciLifeLab structure biology research groups and infrastructure units with such methods to establish a service for computational analysis of molecular structures and interactions.

A unique ecosystem in Structural Biology in Sweden

As the demand for expertise in structural biology increases in Sweden, we are connecting a number of key players in the field to build a strong knowledge base and provide better services to the community.

Data

In this initial stage of the TDP, we are working with Marta Carroni at the cryo-EM facility in SciLifeLab. The main goal here is to use forced sampling in AlphaFold to get a broader picture of a protein's conformational landscape. This will allow to test and fit multiple conformations in cryo-EM densities. A second stage of the project involves NMR modelling at the Swedish NMR Center.

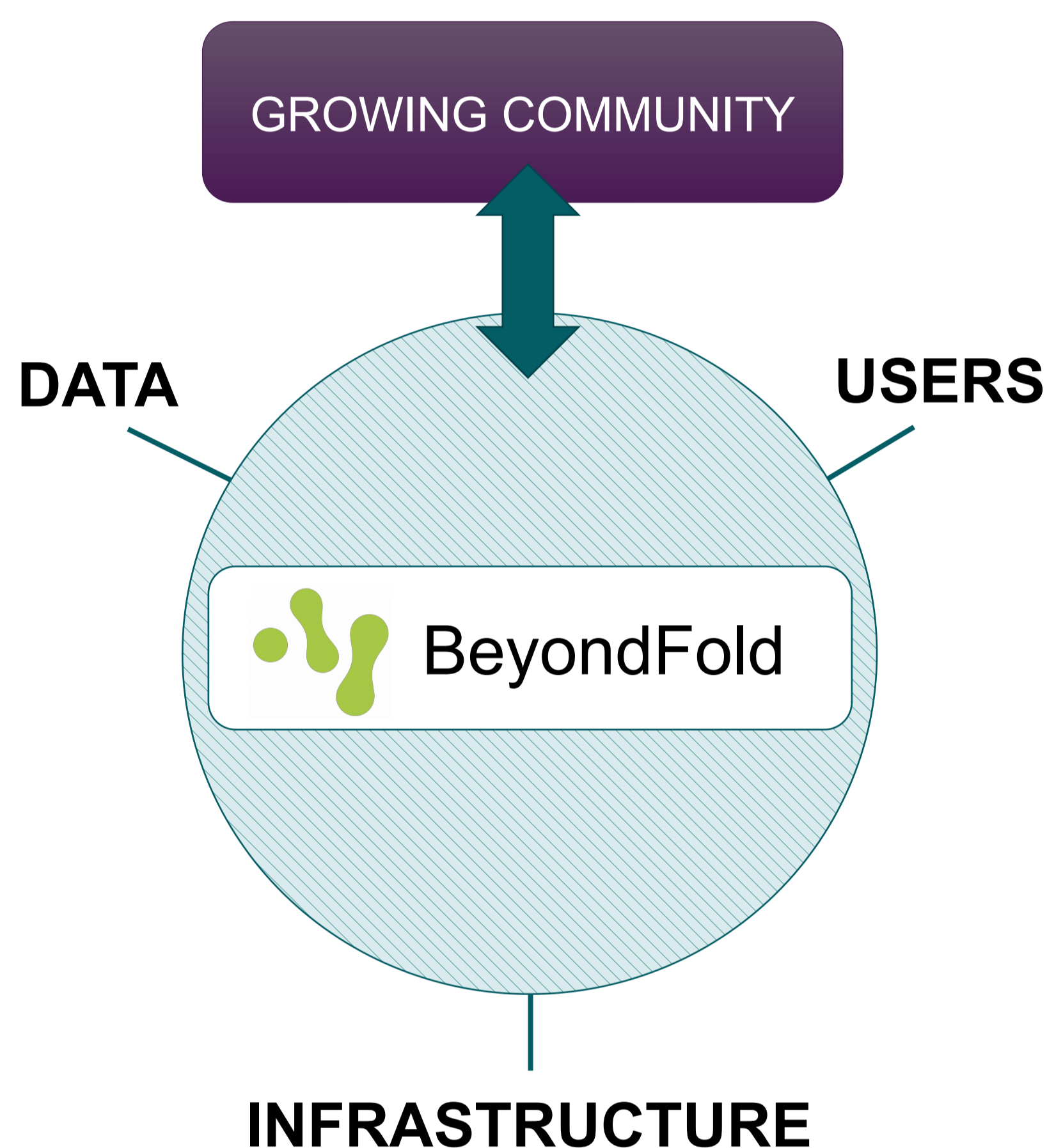
Users

A pilot project with Amunts lab uses AlphaFold's EvoFormer to extract and experimentally validate distance constraints between protein chains, which will be used in Rosetta to guide docking experiments. This approach will be useful whenever AlphaFold fails to model larger protein complexes.

Infrastructure

As part of the Integrative Structural Biology platform and InfraLife, we can track the changing needs of the community and deliver training on computational methods at the InfraLife ISB workshop.

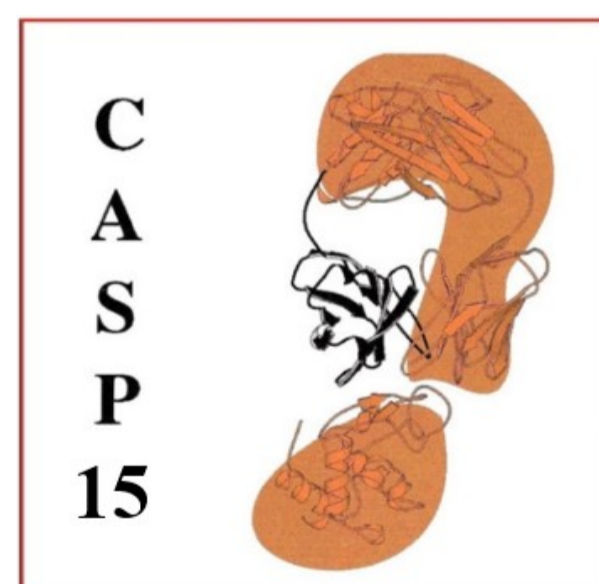
We also work with the Data Centre to facilitate user access to computing infrastructure and tools.



This TDP leverages expertise from data providers, infrastructure and user base in order to provide a better service to the growing structural bioinformatics community in Sweden

CASP15

We are involved in the organization of the latest Critical Assessment of Techniques for Protein Structure Prediction by running an AlphaFold2 server. All models and data were shared with the assessment group as well as all participants, and will establish a baseline against which all other methods are benchmarked.



Find all data at:

duffman.it.liu.se/casp15/

Tools Development

We are collaborating with computational groups (Elofsson and Wallner lab) to provide new tools to benefit the whole community:

AF_server

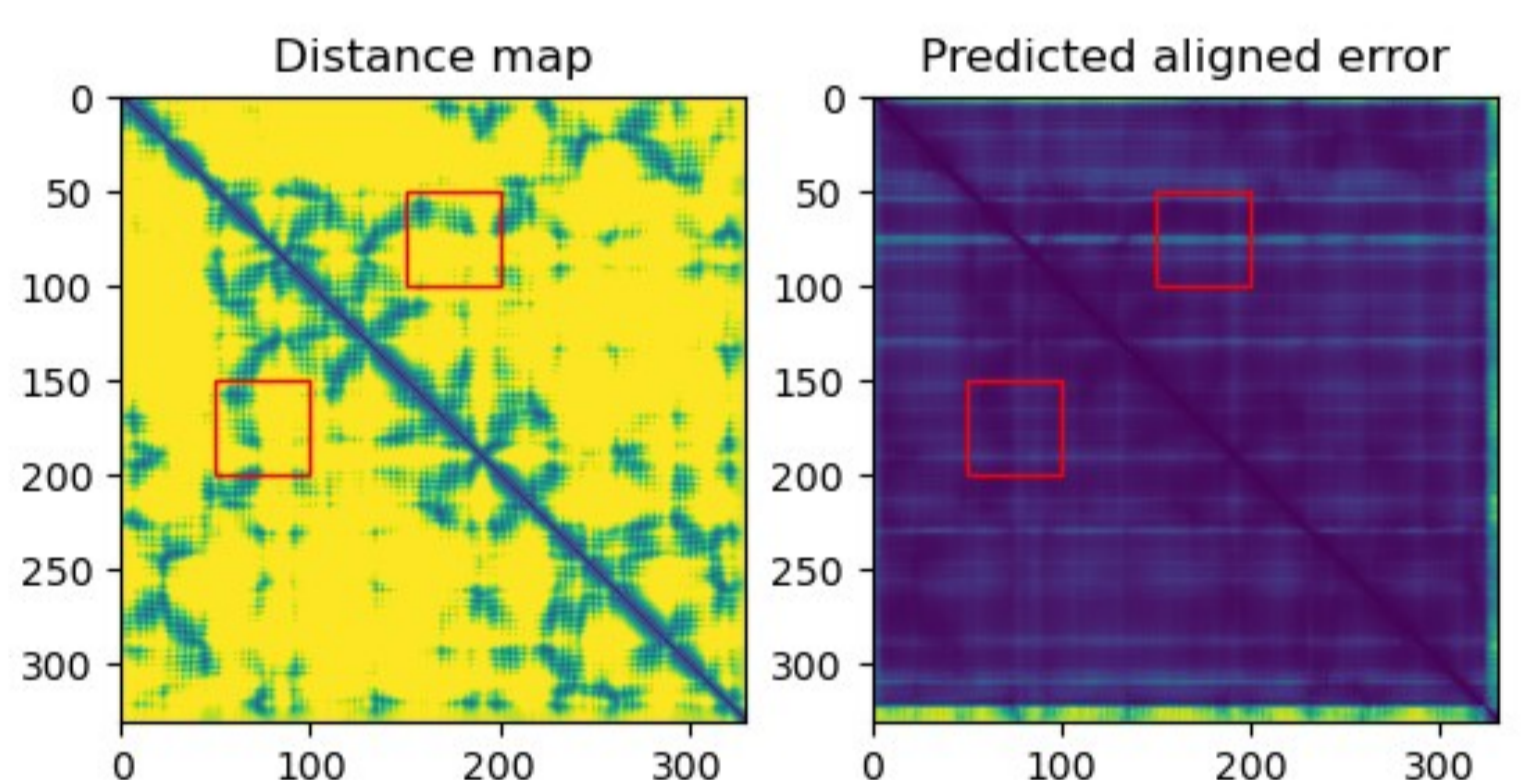
[available on demand](#)

A Snakemake pipeline that allows to query an AlphaFold2 service and receive the results via e-mail. The service can run on HPC systems such as BerzeLiUs, UPPMAX.

Dgram2dmap

github.com/clami66/dgram2dmap

This tool extracts residue-residue distance predictions from an AlphaFold EvoFormer output for visual inspection and convert them to a constraint file for further experiments on other tools such as Rosetta.



Dgram2dmap plots the distance constraints that form a protein model and extracts constraints from desired ranges of amino acids

IDDT_align

github.com/clami66/IDDT_align

IDDT is the most commonly used measure of the quality of a protein model. This tool allows to compare models in a roto-translational invariant manner, even when these don't have identical sequences.