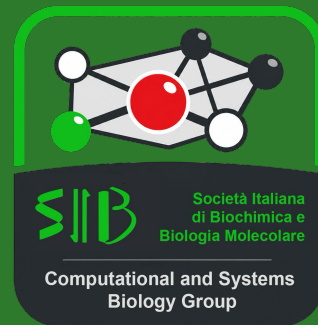


S3 · 14:15–14:30

Alpha&ESMhFolds: An Updated Database for the Comparison and Functional Annotation of Predicted Structural Models for the Human Proteome

Matteo Manfredi

University of Bologna



- Method for predicting the **structure** of a protein
- Winner of **CASP14** (2020)
- Huge **revolution** in the field of structural biology

nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [articles](#) > article

Article | [Open access](#) | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

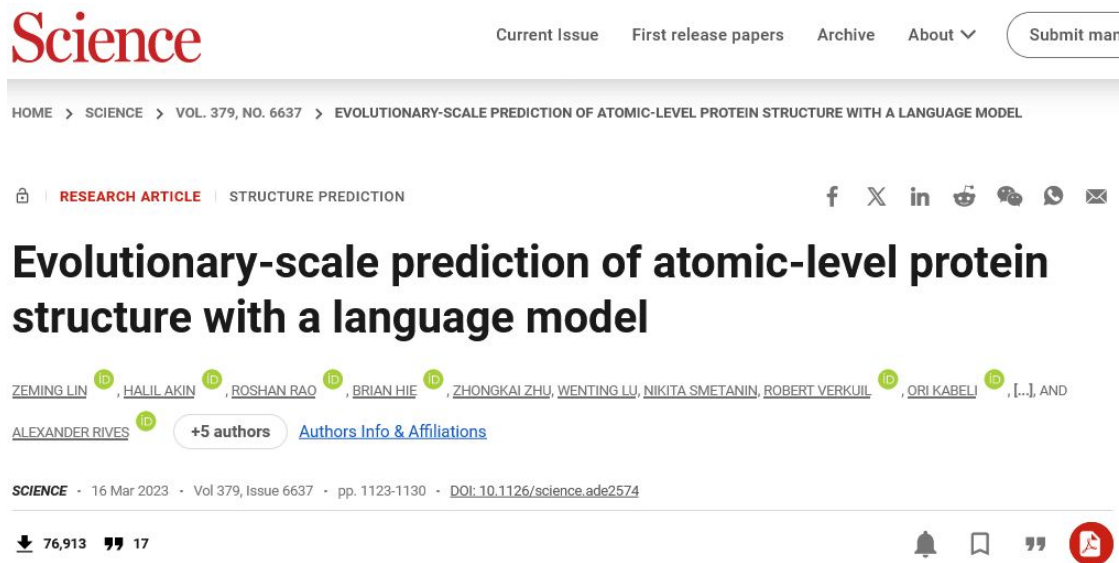
[John Jumper](#) , [Richard Evans](#), [Alexander Pritzel](#), [Tim Green](#), [Michael Figurnov](#), [Olaf Ronneberger](#), [Kathryn Tunyasuvunakool](#), [Russ Bates](#), [Augustin Žídek](#), [Anna Potapenko](#), [Alex Bridgland](#), [Clemens Meyer](#), [Simon A. A. Kohl](#), [Andrew J. Ballard](#), [Andrew Cowie](#), [Bernardino Romera-Paredes](#), [Stanislav Nikolov](#), [Rishub Jain](#), [Jonas Adler](#), [Trevor Back](#), [Stig Petersen](#), [David Reiman](#), [Ellen Clancy](#), [Michal Zielinski](#), ... [Demis Hassabis](#) 

[+ Show authors](#)

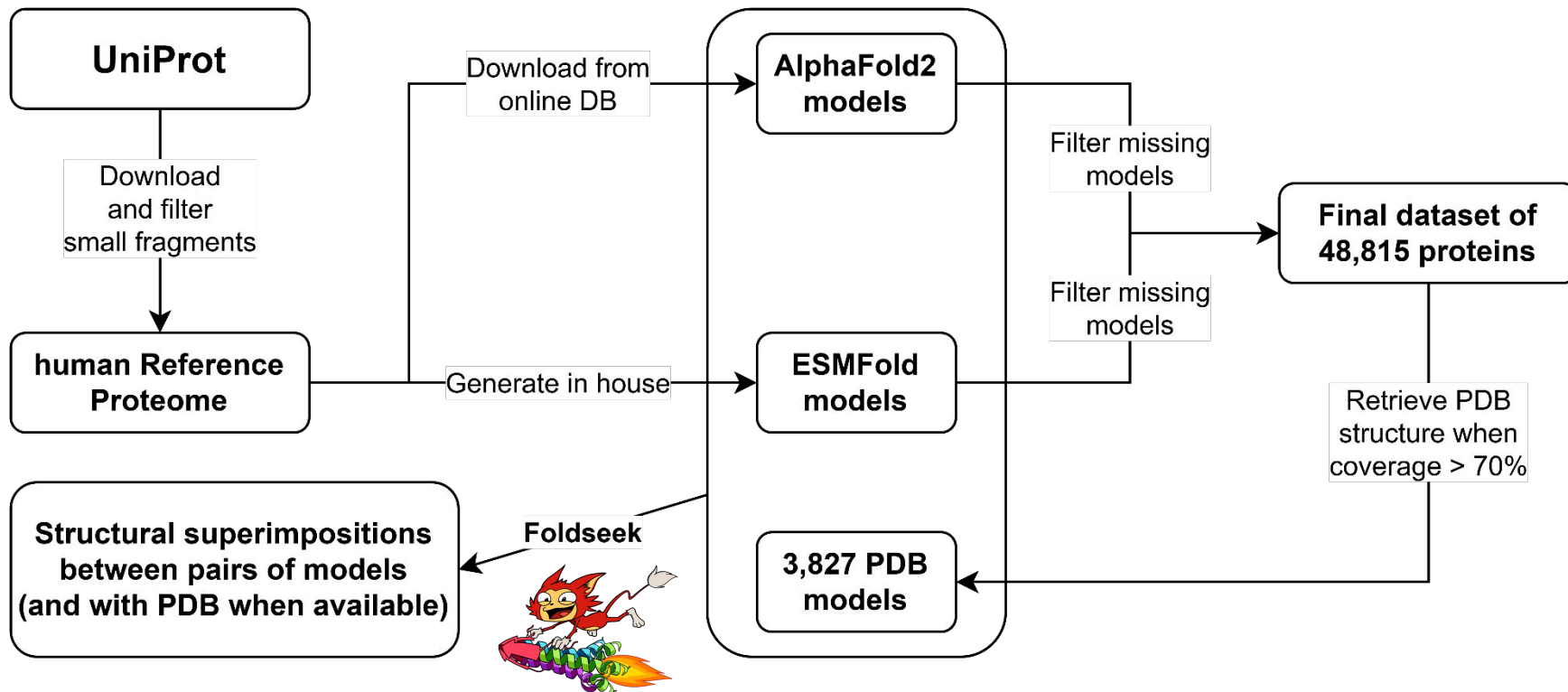
[Nature](#) **596**, 583–589 (2021) | [Cite this article](#)

1.60m Accesses | **13k** Citations | **3591** Altmetric | [Metrics](#)

- Adoption of only protein **embeddings** in input
 - Based on **Protein Language Models**
- Two advantages:
 - **Faster** protein encoding
 - Not reliant on **deep** Multiple Sequence Alignments



- Database storing **48,815** pairs of predicted structural models
- Focus on the **Human Proteome**
- Motivations:
 - Large-scale **comparison** of the two methods
 - Online resource for **choosing the best model**

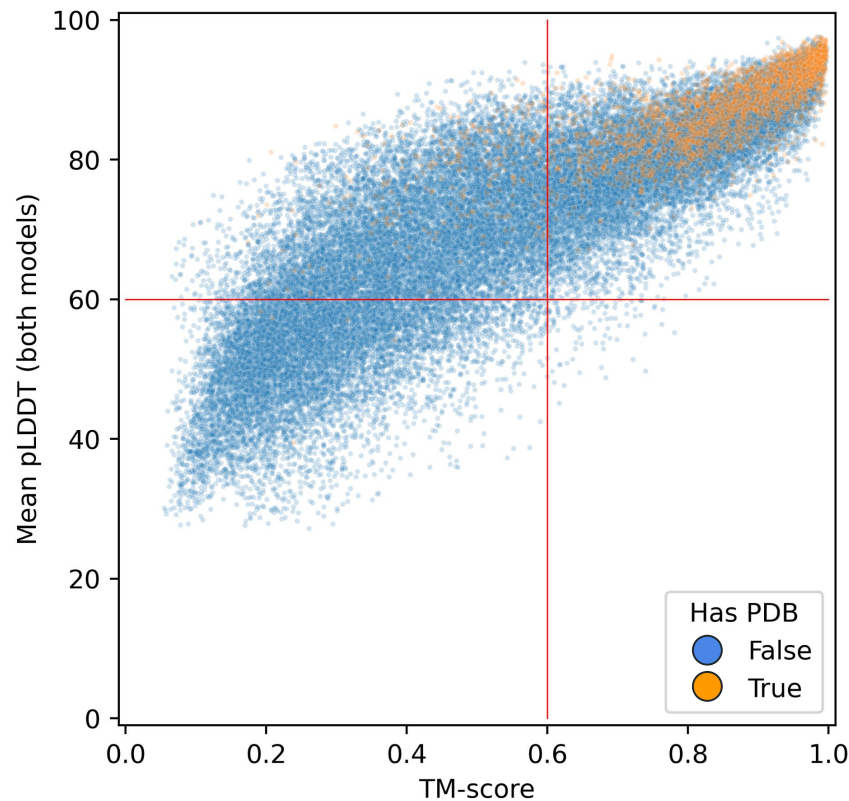


- Huge update of UniProt at the end of 2025
- AlphaFold DB replaced all models, upgrading them to their 6th version
- New version of Alpha&ESMhFolds
 - **6190** new proteins included
 - **394** proteins with a sequence update
 - **19,270** with a different AlphaFold model
 - **22,961** proteins unchanged
 - **927** more proteins with an experimental PDB

Model quality and similarity

CSB Meeting 2026

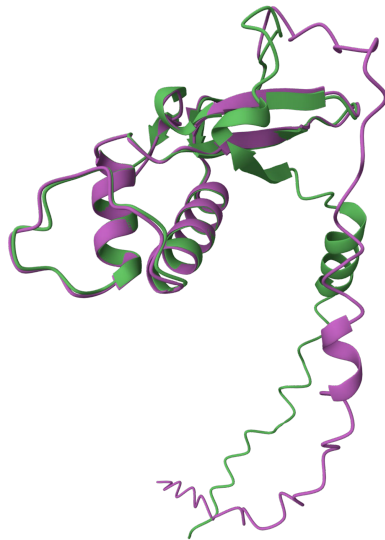
- **Mean pLDDT**
 - Metric of the quality of a model
 - Self-assessed by both methods
 - **60** → Standard threshold
- **TM-score**
 - Metric of similarity of a pair
 - Computed by Foldseek
 - **0.6** → Standard threshold
- **High correlation (0.83 PCC)**



- Self-assessed pLDDT
 - Models trained in different ways
 - Good indication of reliability, but **hard to compare directly**
 - **Different spread** of values
- Use of state-of-the-art Quality Assessment tools
 - **DeepAccNet**
 - **QMEANDisCo**
 - **QATEN**

TM-score ≥ 0.6	DeepAccNet	QMEANDisCo	QATEN	Consensus
AlphaFold	88%	80%	55%	81%
ESMFold	12%	20%	45%	19%

TM-score < 0.6	DeepAccNet	QMEANDisCo	QATEN	Consensus
AlphaFold	73%	49%	34%	51%
ESMFold	27%	51%	66%	49%



ESMFold vs AlphaFold2

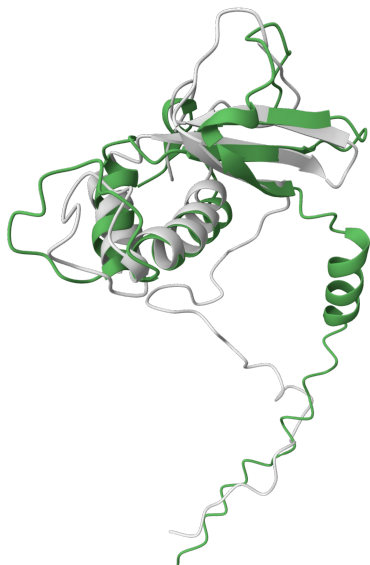
70 pLDDT 78

TM-score: 0.65

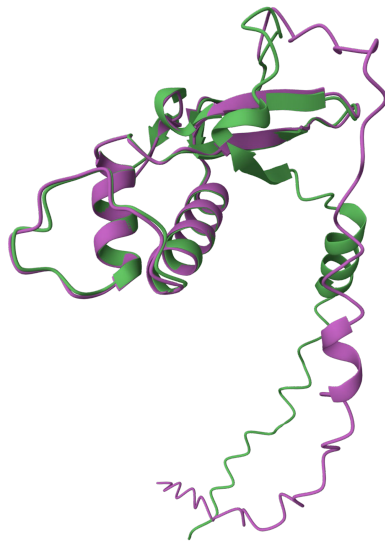
DNA-directed RNA polymerase

Example

CSB Meeting 2026



ESMFold vs PDB
TM-score = **0.62**

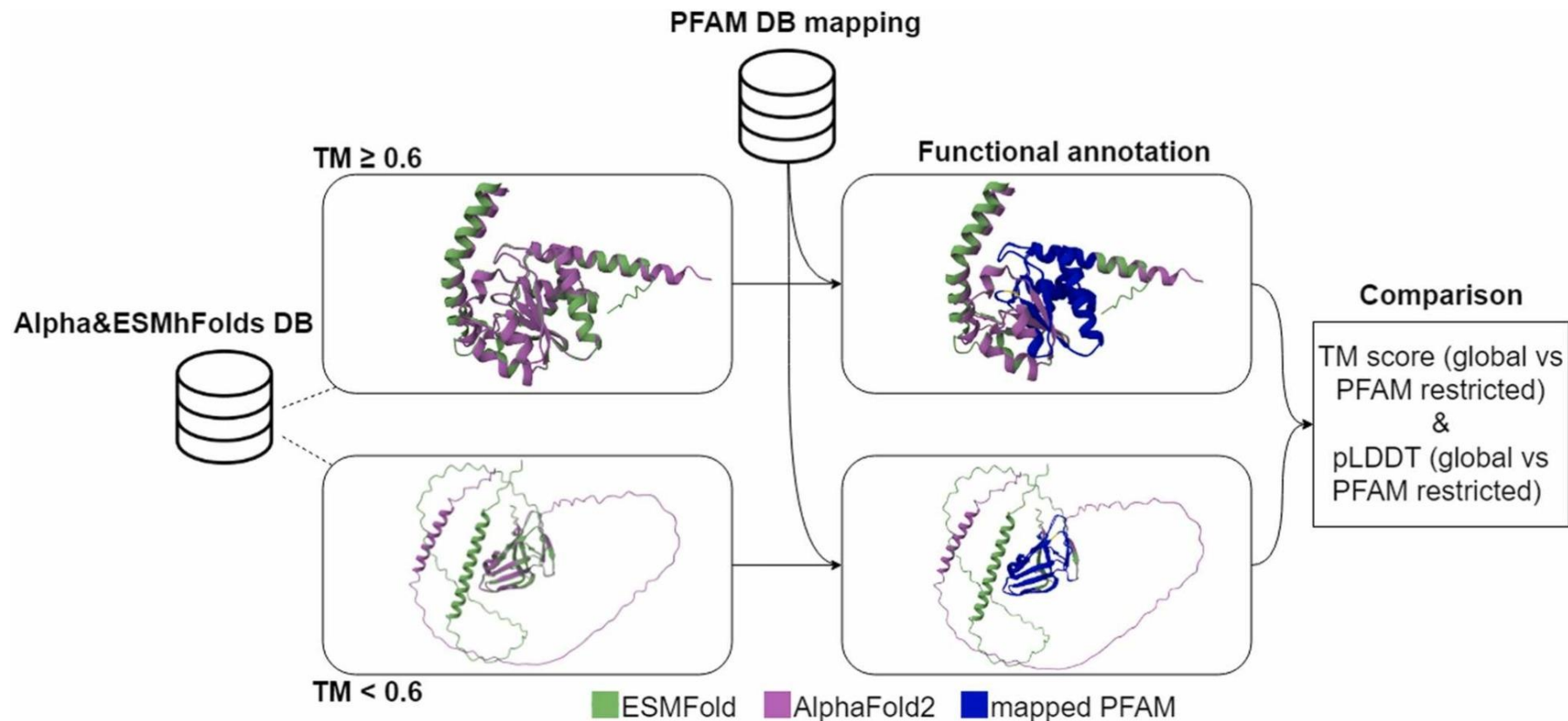


ESMFold vs AlphaFold2
70 pLDDT 78
3 External QA 0



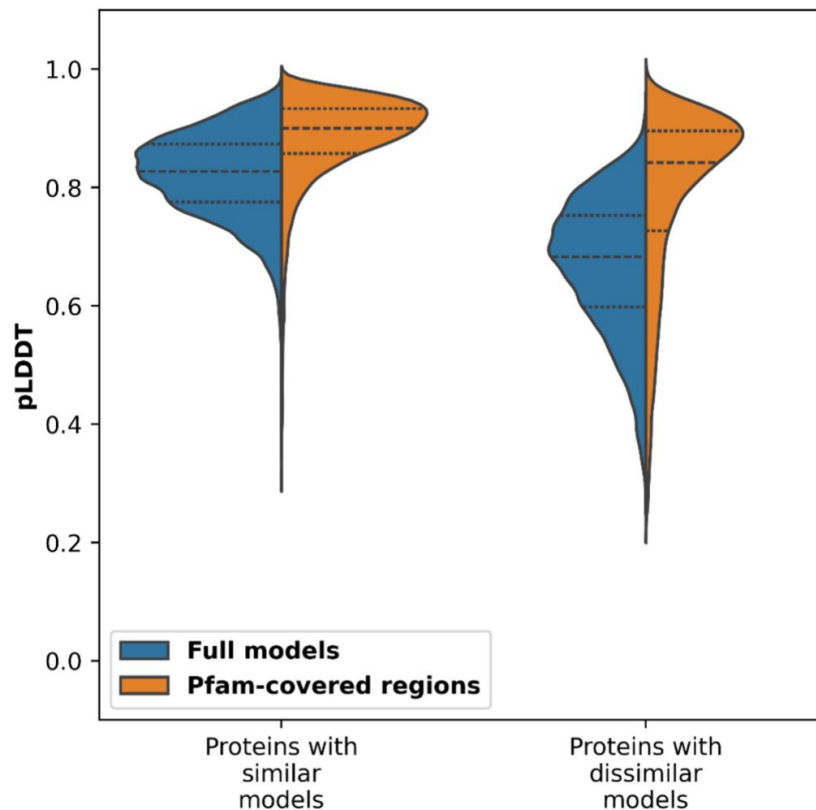
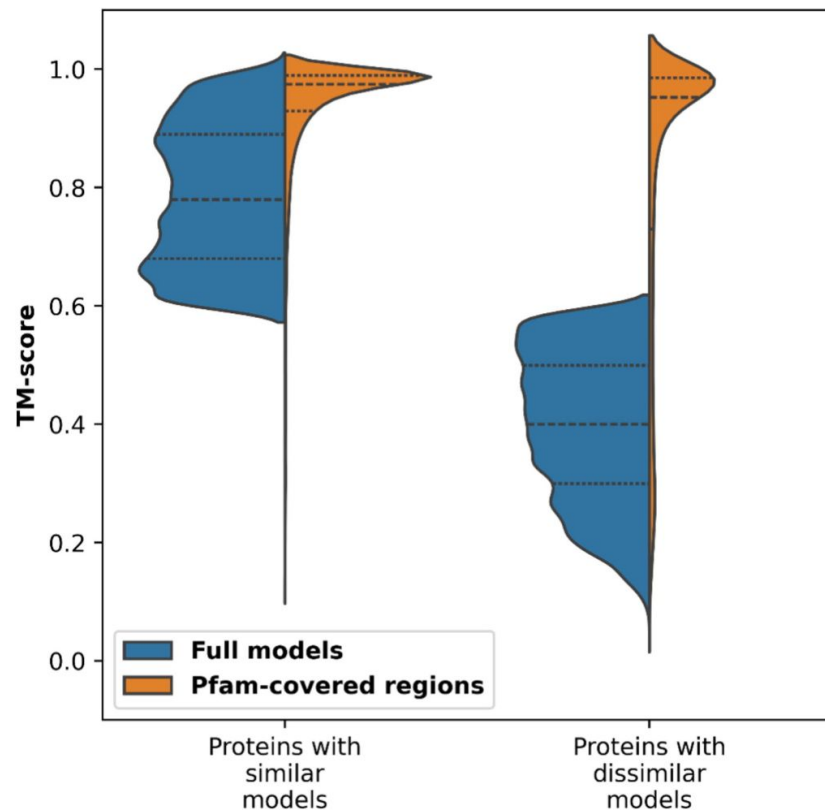
AlphaFold2 vs PDB
TM-score = **0.48**

DNA-directed RNA polymerase



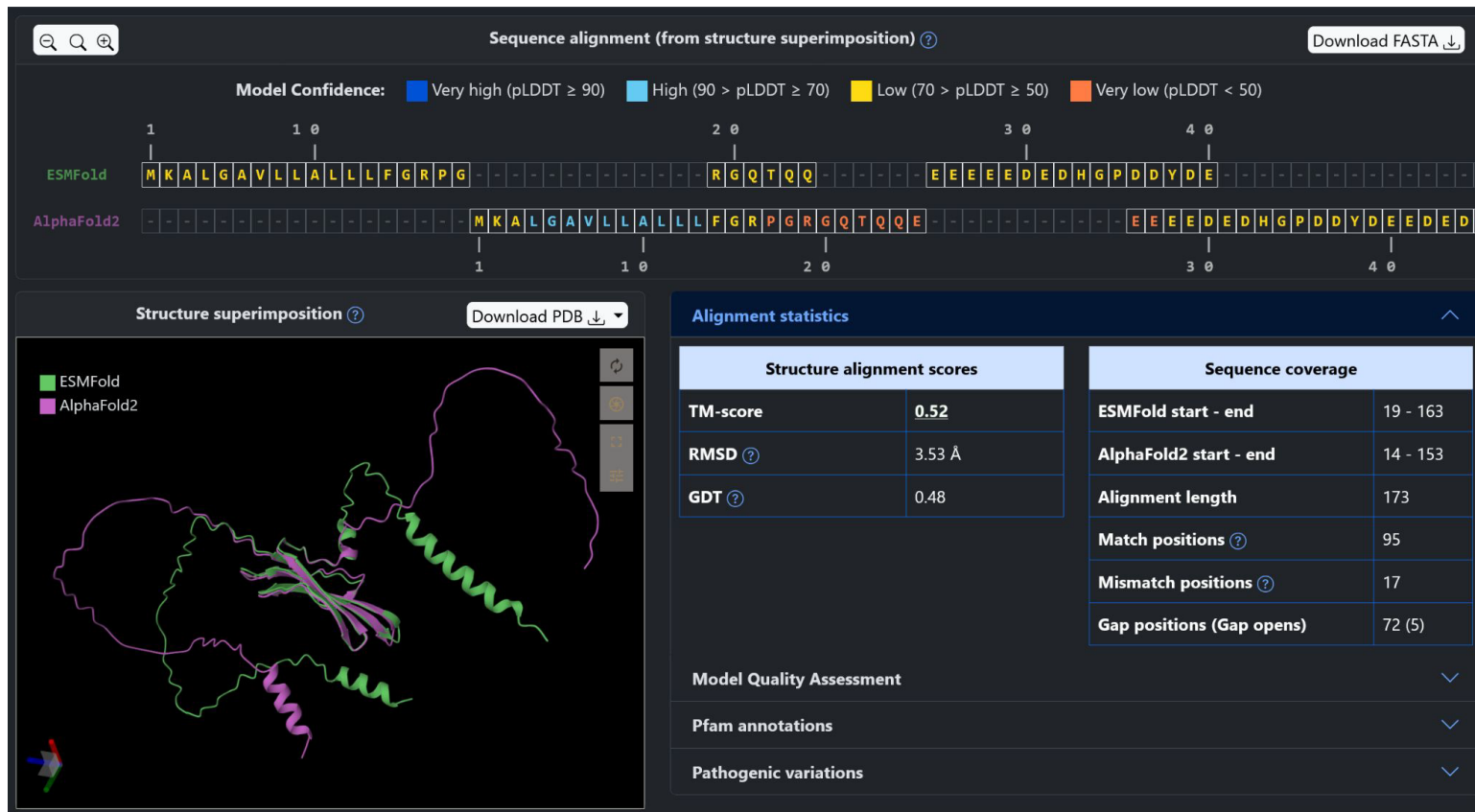
Functional Annotation

CSB Meeting 2026



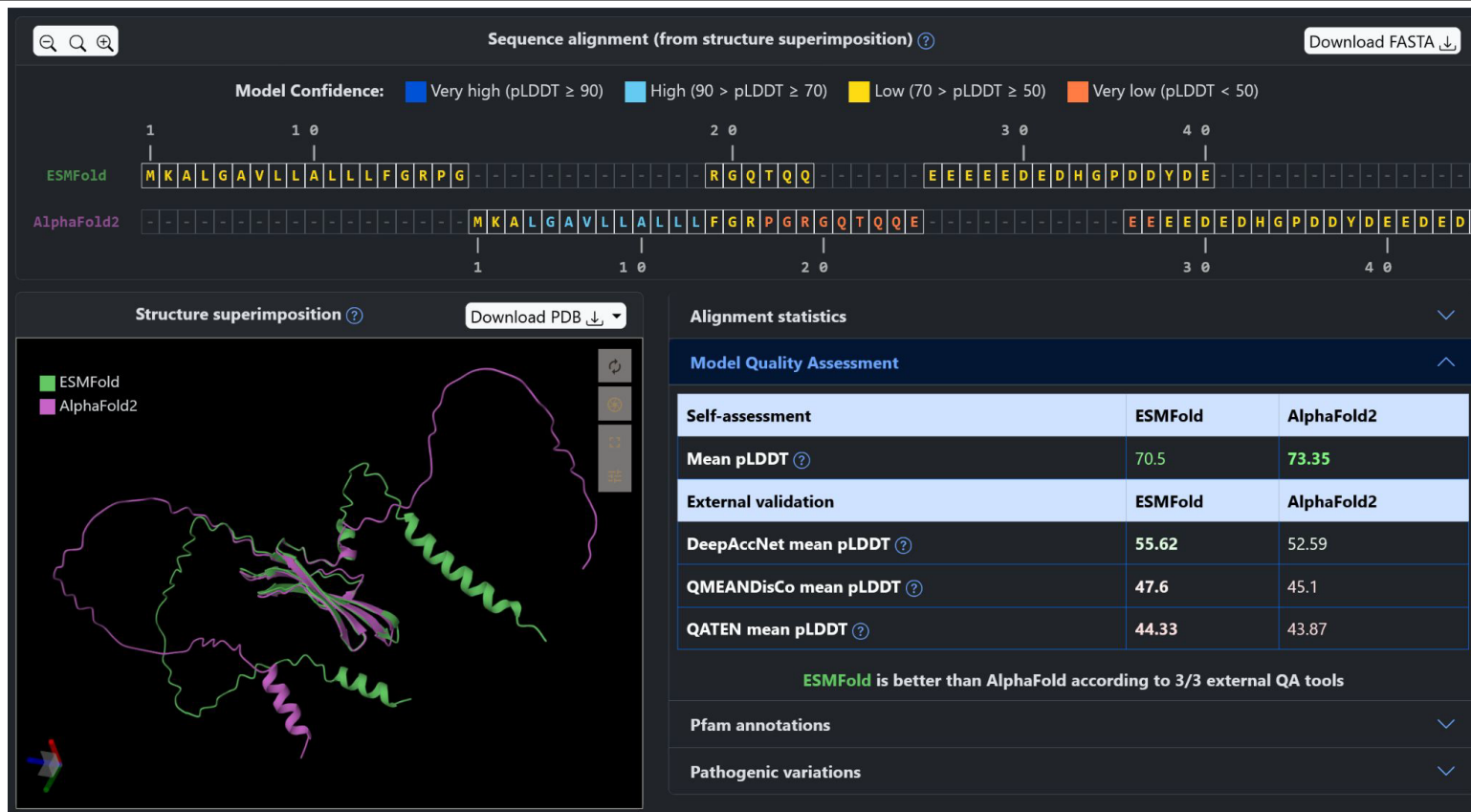
Web server visualization

CSB Meeting 2026



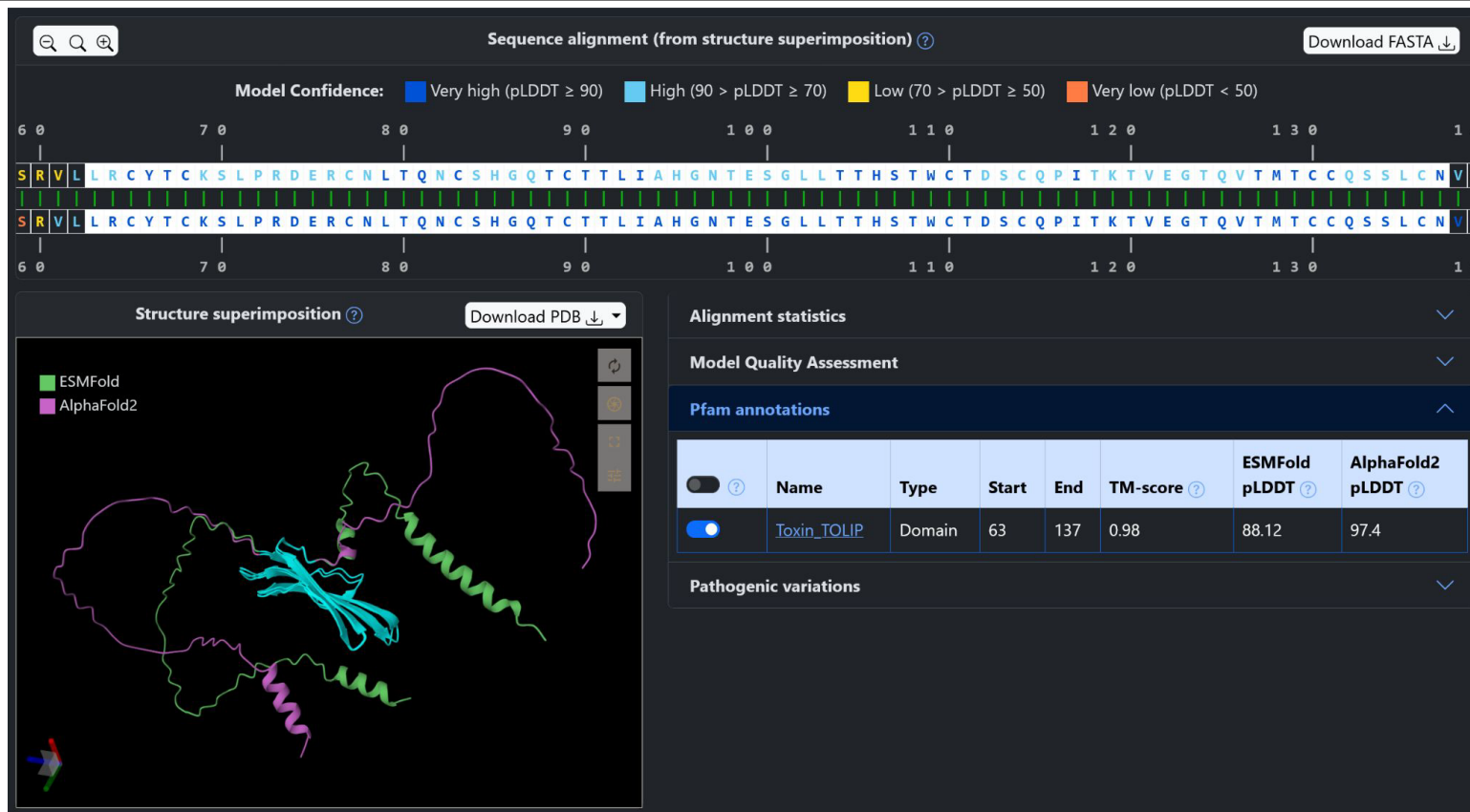
Web server visualization

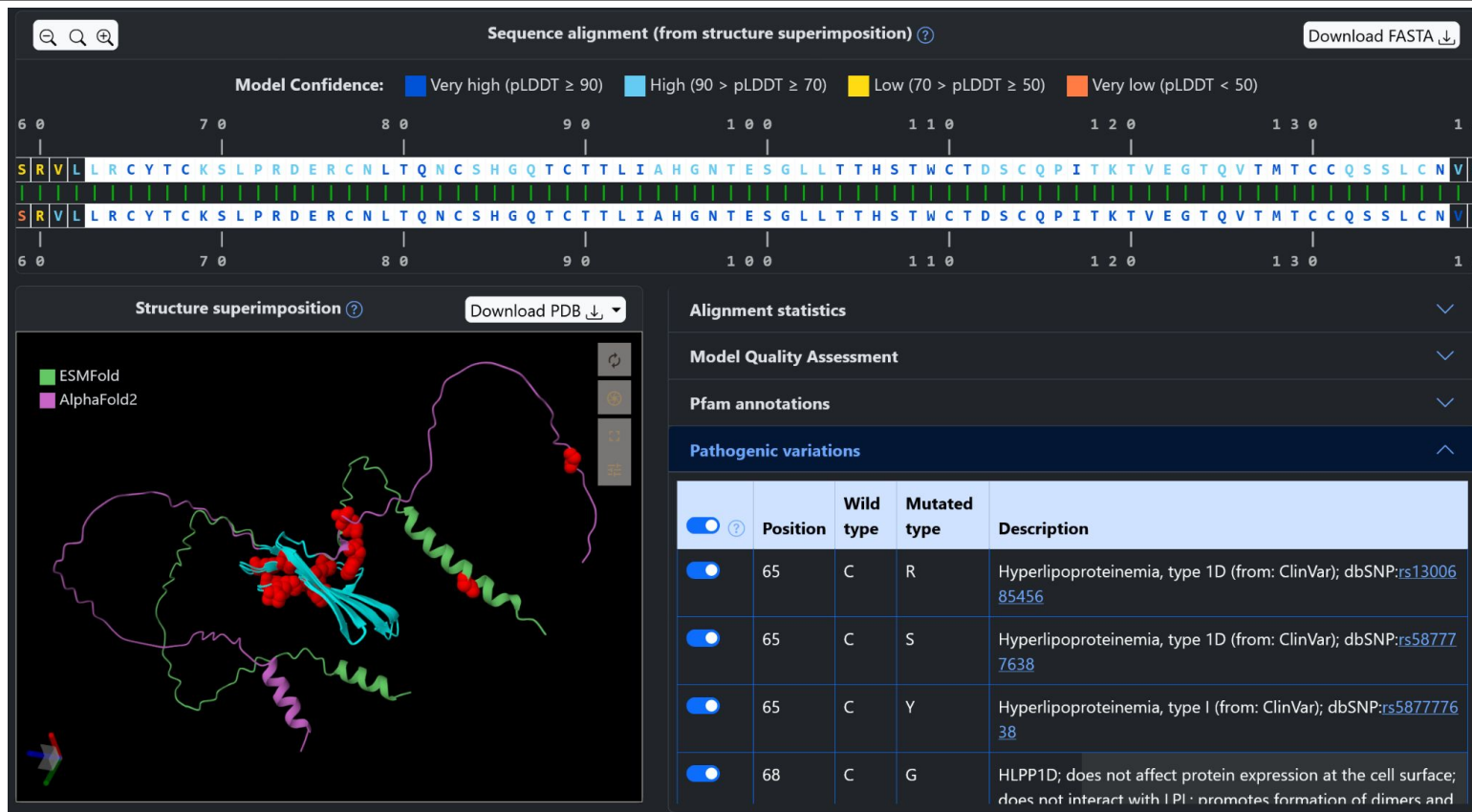
CSB Meeting 2026



Web server visualization

CSB Meeting 2026





Acknowledgments

CSB Meeting 2026

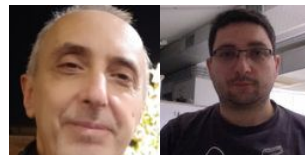
Biocomputing Group

- Rita Casadio
- Pier Luigi Martelli
- Castrense Savojardo
- Manfredi Matteo



INFN-CNAF

- Davide Salomoni
- Alessandro Costantini

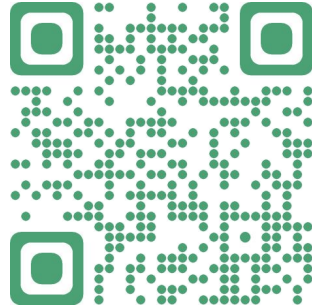


Fundings

- HEAL ITALIA
- ELIXIR x
NextGenerationIT



Thank you for your attention



← Link to the web server