nf-core/proteinfold

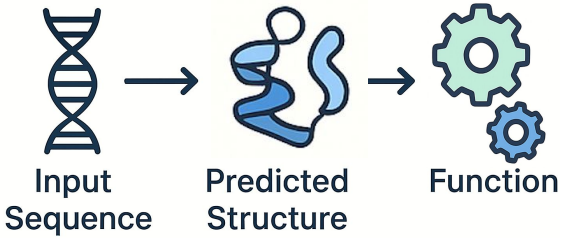
A standardised pipeline-hub for protein structure prediction methods

Jose Espinosa-Carrasco (Notredame's lab)

31st of January, FOSDEM 2026 - Brussel

FOSDEM'26

Why is it important to predict structures?



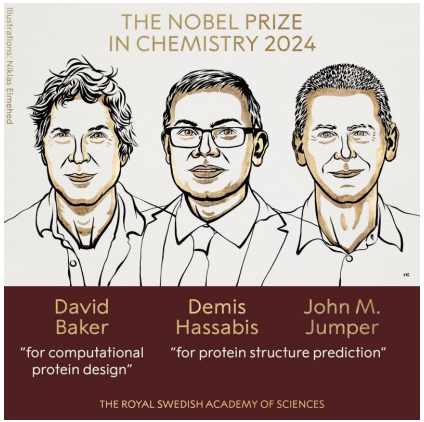
Article | Open access | Published: 15 July 2021

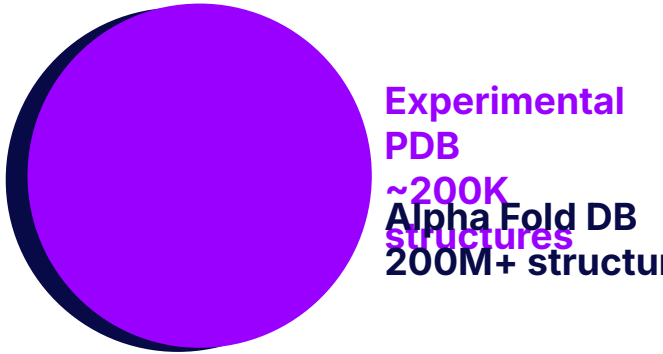
Highly accurate protein structure prediction with AlphaFold

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

1.28m Accesses | **9258** Citations | **3528** Altmetric | Metrics





AI methods revolutionised structural biology

An explosion of AI tools for protein prediction

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

Evolutionary-scale prediction of atomic-level protein structure with a language model

ZEMING LIN, HALIL AKIN, ROSHAN RAO, BRIAN HIE, ZHONGKAI ZHU, WENTING LU, NIKITA SMETANIN, ROBERT VERKUIL, ORI KABELI, [...], AND ALEXANDER RIVES +5 authors Authors Info & Affiliations

Technical Report of HelixFold3 for Biomolecular Structure Prediction

Lihang Liu, Shanzhuo Zhang, Yang Xue, Xianbin Ye, Kunrui Zhu, Yuxin Li, Yang Liu, Jie Gao, Wenlai Zhao, Hongkun Yu, Zhihua Wu, Xiaonan Zhang, Xiaomin Fang

Accurate prediction of protein structures and interactions using a three-track neural network

MINKYUNG BAEK, FRANK DIMAIO, IVAN ANISHCHENKO, JUSTAS DAUPARAS, SERGEY OVCHINNIKOV, GYU RIE LEE, JUE WANG, QIAN CONG, LISA N. KINCH, [...], AND DAVID BAKER +22 authors Authors Info & Affiliations

SCIENCE · 19 Aug 2021 · Vol 373, Issue 6557 · pp. 871-876 · DOI: 10.1126/science.abj8754

Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA

Minkyung Baek, Ryan McHugh, Ivan Anishchenko, Hanlun Jiang, David Baker & Frank DiMaio

Nature Methods **21**, 117–121 (2024) | Cite this article

Accurate structure prediction of biomolecular interactions with AlphaFold 3

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, ... John M. Jumper + Show authors

Nature **630**, 493–500 (2024) | Cite this article

ColabFold: making protein folding accessible to all

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov & Martin Steinegger

Nature Methods **19**, 679–682 (2022) | Cite this article

BOLTZ-1 Democratizing Biomolecular Interaction Modeling

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, Regina Barzilay

doi: https://doi.org/10.1101/2024.11.19.624167

Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, Regina Barzilay

doi: https://doi.org/10.1101/2025.06.14.659707

Computing structures is all about **trade-offs**:

speed ⚔️ accuracy ⚔️ cost

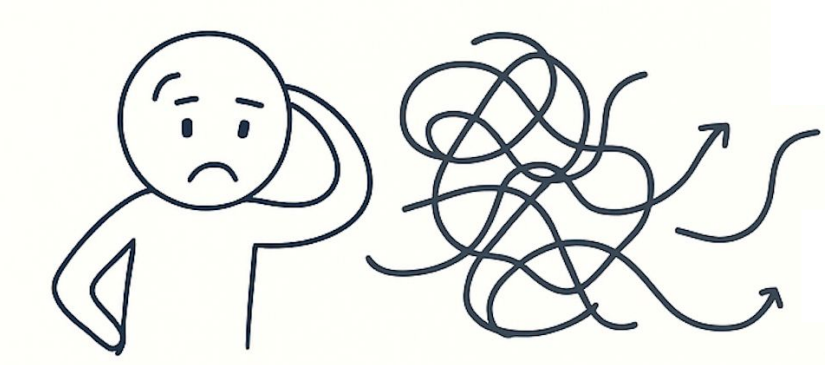
Why running structure prediction tools is hard?

Multiple dependencies on several software libraries and databases!!!

Alphafold2 | Colabfold | RosettafoldAA

Boltz 2 | Esmfold | Alphafold3

Rosettafold2na | Helixfold | Boltz 1



cudatoolkit | biopython

tensorflow | Jax | scipy

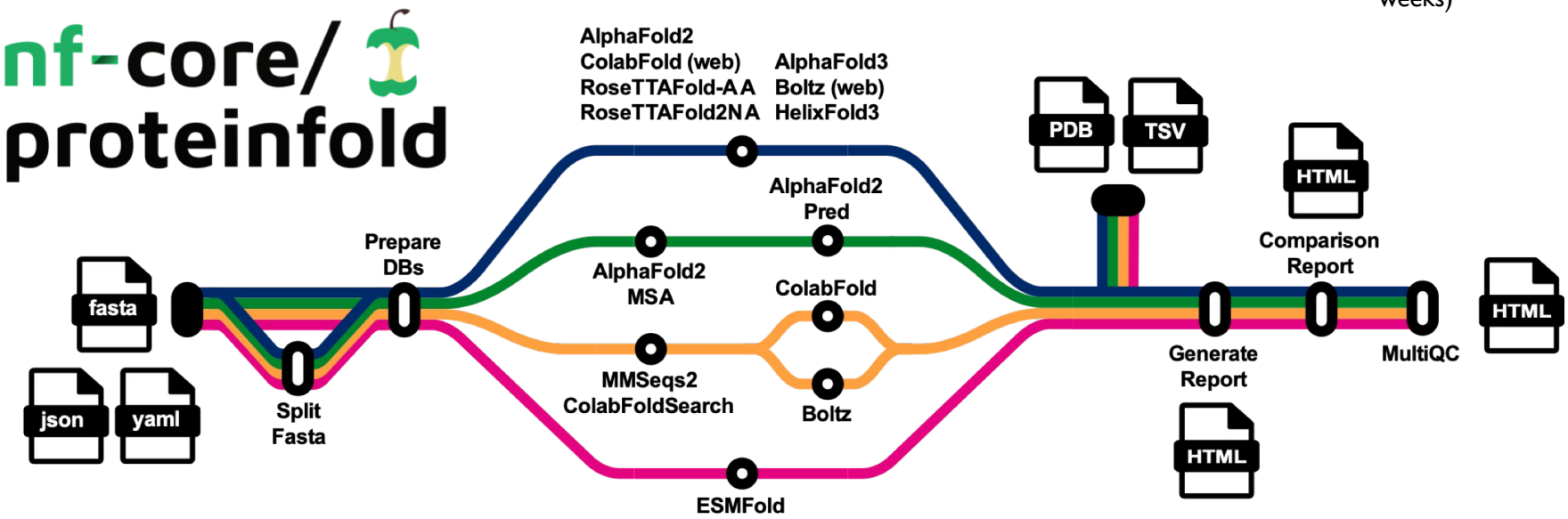
hmmer | wandb | pytorch

uniprot30 | pdb70

colabfolddb | pdbseqres | uniclust30

mgnify | bfd | pdb_mmcif

*Current development versión (release planned for following weeks)



License: CC BY

nf-core/proteinfold

Powered by nextflow

Reproducible • Scalable • Portable

DBs and parameters downloads

Containarized Software and dependencies

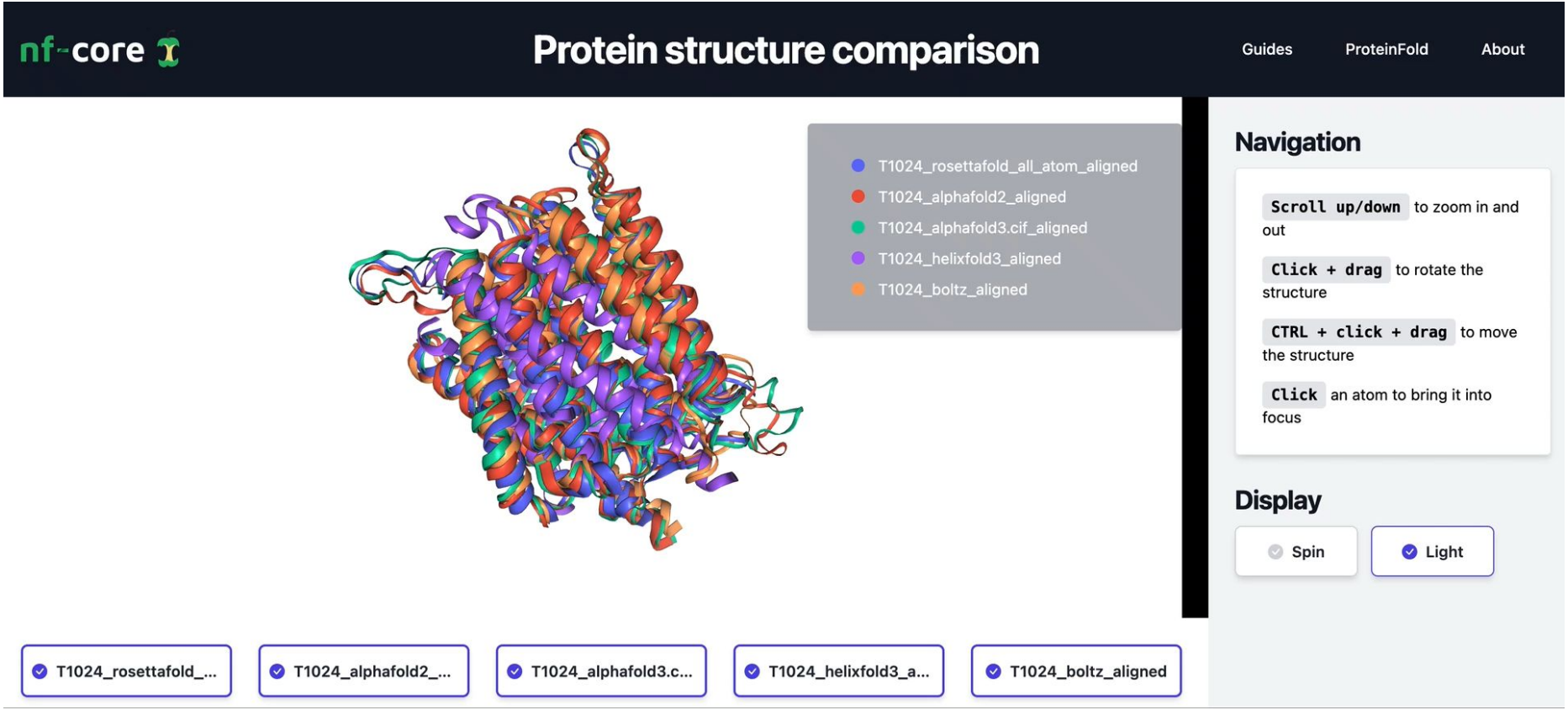
Easy configuration (e.g. gpu/cpu)

Enables methods benchmarking

Provides reporting capabilities

Part of nf-core

REPORTING EXAMPLE



A real application: predict a molecular inventory of an organism

Keiran Rowell

UNSW SYDNEY

Do eukaryotic cells evolved via symbiosis between sulfate-reducing bacteria and hydrogen-producing archaea?

UNSW Ramaciotti Centre for Genomics

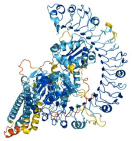
'equence genome

4544 proteins (*Nerearchaeum marumarumayae*)

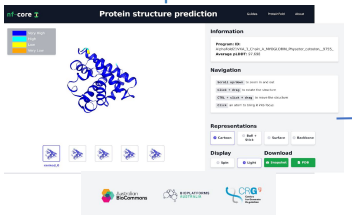
A mix of AlphaFold2, ESMFold and AlphaFold3

nf-core/ proteinfold

Struct. Pred.

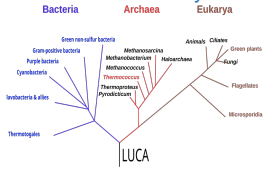


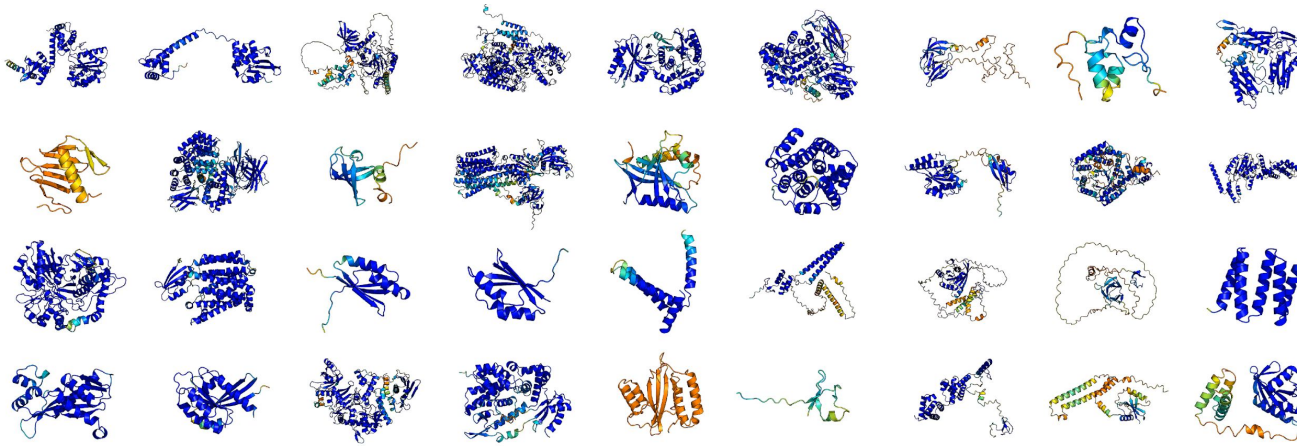
Reports



Struct. Search (Foldseek)

Struct. Bio. Analysis





An Asgard archaeon from a modern analog of ancient microbial mats

Stephanie-Jane Nobs, Matthew D. Johnson, Timothy J. Williams, Julia Meltzer, Xabier Vázquez-Campos, Fraser I. MacLeod, Keiran Rowell, Miranda Pitt, Bindusmita Paul, Doulin C. Shepherd, Katharine A. Michie, Iain G. Duggin, Debnath Ghosal, Brendan P. Burns

doi: https://doi.org/10.1101/2025.07.22.663070

Community built pipelines, community level benefits

Open Source development delivers benefits for everyone

♻️ **Reusing instead of reinventing**

✅ **Reproducible across sites and systems**

✨ **Lower maintenance through shared standards**

🔍🤖 **More reliable through review and automated testing**

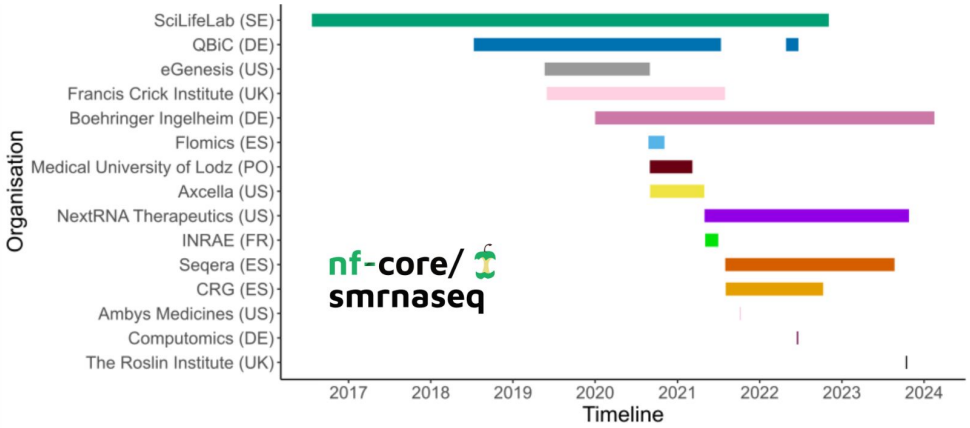
👁️ **More transparent and auditable science**

⬇️ **Lower long-term maintenance costs through shared improvements**

Collective maintenance makes pipelines last

Paper here!

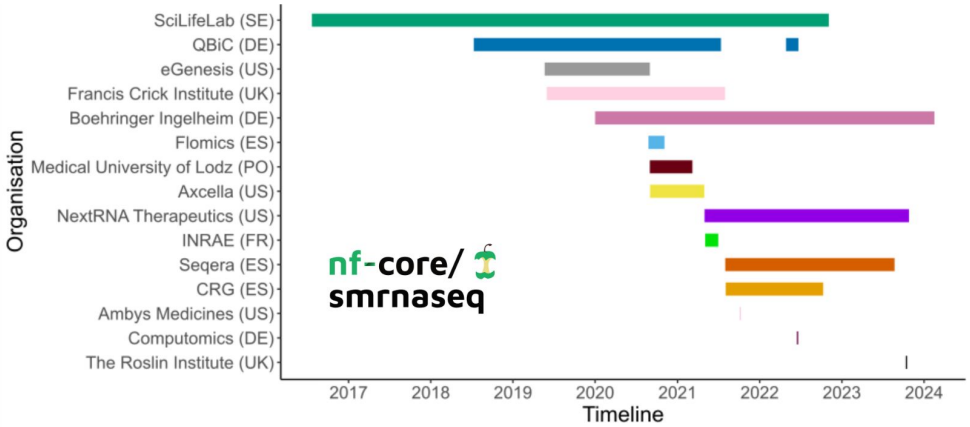
Langer *et al. Genome Biology* *(2025) 26:228*
https://doi.org/10.1186/s13059-025-03673-9



Collective maintenance makes pipelines last

Paper here!

Langer *et al. Genome Biology* (2025) 26:228
https://doi.org/10.1186/s13059-025-03673-9



Current nf-core/proteinfold development team:

Luisa Santus
Emilio Palumbo
Jose Espinosa-Carrasco

Patri Bota

Keiran N. Rowell
Joshua Caley
Thomas Litfin

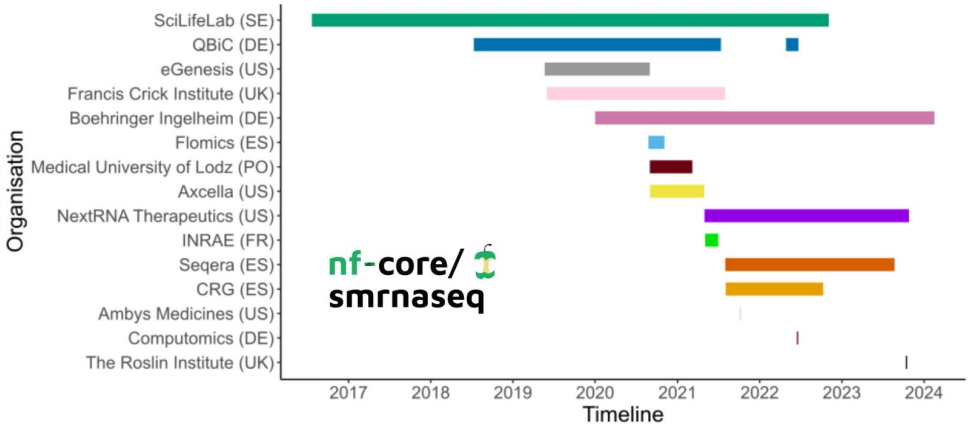
Ziad Al Bkhetan

Martin Steinegger

Collective maintenance makes pipelines last

Paper here!

Langer *et al. Genome Biology* (2025) 26:228
https://doi.org/10.1186/s13059-025-03673-9



Join us:

https://nf-co.re/join

https://nf-co.re/join/slack
#proteinfold
#proteinfold-dev

https://github.com/nf-core/proteinfold

Current development team:

Luisa Santus
Emilio Palumbo
Jose Espinosa-Carrasco

Patri Bota

Keiran N. Rowell
Joshua Caley
Thomas Litfin

Ziad Al Bkhetan

Martin Steinegger

The pipeline is open to ideas and contributions!!!

Paper in preparation

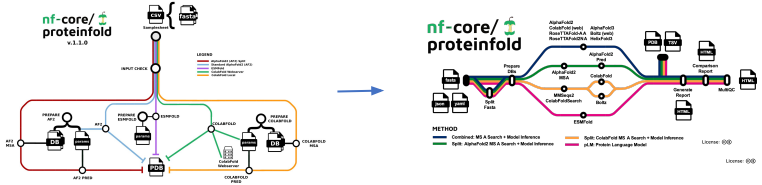
nf-core/proteinfold: a bioinformatics best-practice analysis pipeline for protein 3D structure prediction

Authors and Affiliations:
Athanasios Baltzis[1], Jose Espinosa-Carrasco[1*], Luisa Santus[1], Leila Mansouri[1], Martin Steinegger[3,4], Harshil Patel[5], Toni Hermoso Pulido[1], Julia Ponomarenko[1], Emilio Palumbo[1], Emyr James[1], Patricia Mirela Bota[2], Baldo Oliva[2], Minh Vu[6], Keiran N. Rowell[7], Joshua Caley[7], Nathan Glades[7], Thomas Litfin[7], Georgie Samaha[6,8], Mitchell J O'Brien[6,8], Nigel Ward[9], Katharine A Michie[10], Steven Manos[11], nf-core community[†], Ziad Al Bkhetan[11,*], Cedric Notredame[1,2*]

Abstract

The advances in deep learning frameworks have revolutionised protein studies and contributed to unprecedented accurate predictions of protein structures. Leveraging these advancements, we introduce here a new nf-core pipeline allowing the deployment of three main publicly available resources: AlphaFold2, ColabFold and ESMFold. This pipeline enhances accessibility to these tools by addressing issues related to dependencies on third-party software and databases as well as facilitating the deployment on a variety of platforms including HPC, clouds and personal workstations. nf-core/proteinfold (https://github.com/nf-core/proteinfold), is a Nextflow pipeline developed by the nf-core community according to its guidelines. As such, it supports scalable, portable and reproducible computation. These best-practice guidelines ensure that the pipeline is properly optimised for execution on the major cloud providers as well as HPC infrastructures. We foresee that this development endeavour will have a significant impact on a variety of biological analyses based on protein structures by granting access to an open-source, community-developed resource to obtain protein folds.

Future directions



- Release version 2.0.0
- Submit paper
- Threshold base runs
- Option to benchmark prediction accuracy against input experimental structure
- Option to provide input MSA in *a3m* format (e.g. for Colabfold)
- Extend pTM/iPTM data in report (e.g. region specific selection)
- Expand collaboration with PDBe (hackathon planned for next year)

Acknowledgement

Center for Genomic Regulation

Notredame's lab
Athanasios Baltzis
Leila Mansouri
Luisa Santus
Cedric Notredame

Biocore
Toni Hermoso Pulido
Julia Ponomarenko

SIT
Emilio Palumbo
Emyr James

Seoul National University
Martin Steinegger

S

University of New South Wales

Structural Biology Facility
Keiran N. Rowell
Joshua Caley
Nathan Glades
Thomas Litfin

Australian Biocommons
Ziad Al Bkhetan
Steven Manos

EMBL-EBI
Evangelos Karatzas

Seqera
Evan Floden
Paolo Di Tommaso
Harshil Patel
Jordi Deu-Pons

Communities:

The AWS Open Data Sponsorship Program

Questions?