



FACULTY OF SCIENCE  
Charles University



# Otevřeně o biologických datech

---

Marian Novotný



# Biologie a biomedicína zažívají datovou pohromu



## nature methods

View all journals Search Lo

Explore content About the journal Publish with us

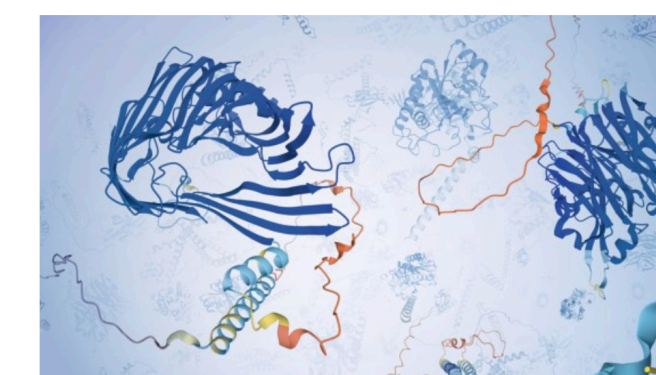
Sign up for alerts RSS

nature > nature methods > focus

FOCUS | 11 JANUARY 2022

### Method of the Year 2021: Protein structure prediction

Protein structure prediction is our Method of the Year 2021, for the remarkable levels of accuracy achieved by deep learning-based methods in predicting the 3D structures of proteins and protein complexes, essentially solving this long-standing challenge.

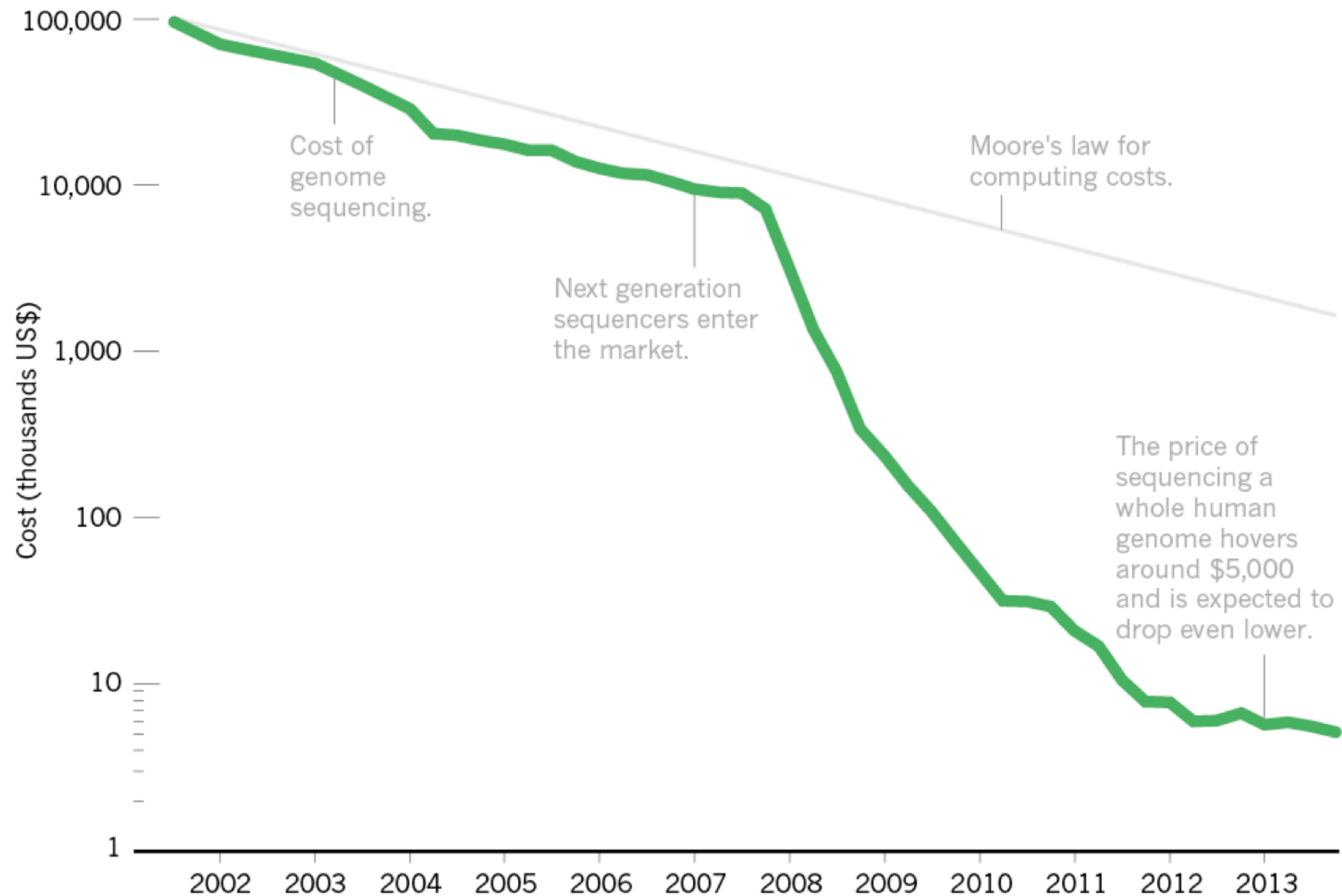




# Cena dat klesá

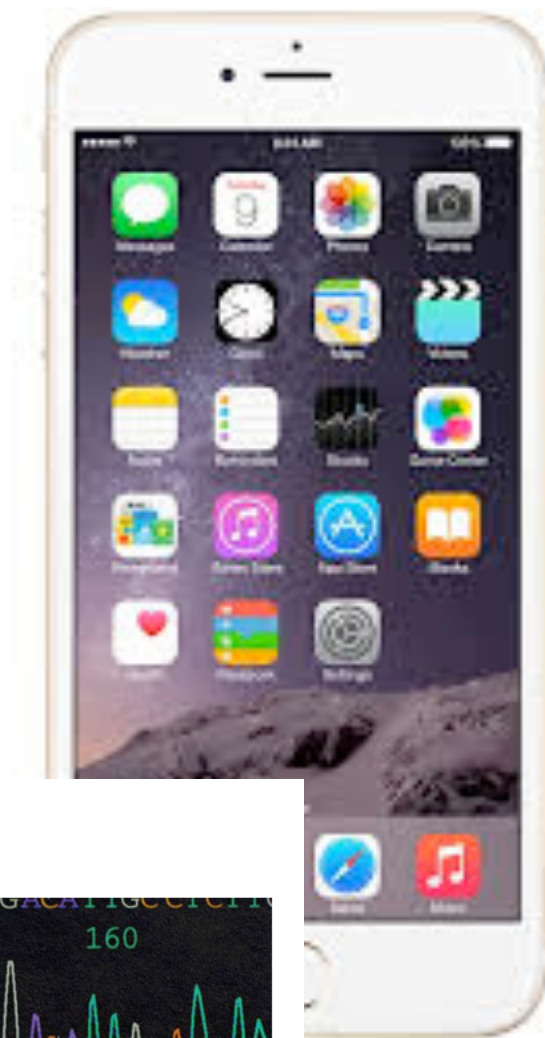
## Falling fast

In the first few years after the end of the Human Genome Project, the cost of genome sequencing roughly followed Moore's law, which predicts exponential declines in computing costs. After 2007, sequencing costs dropped precipitously.





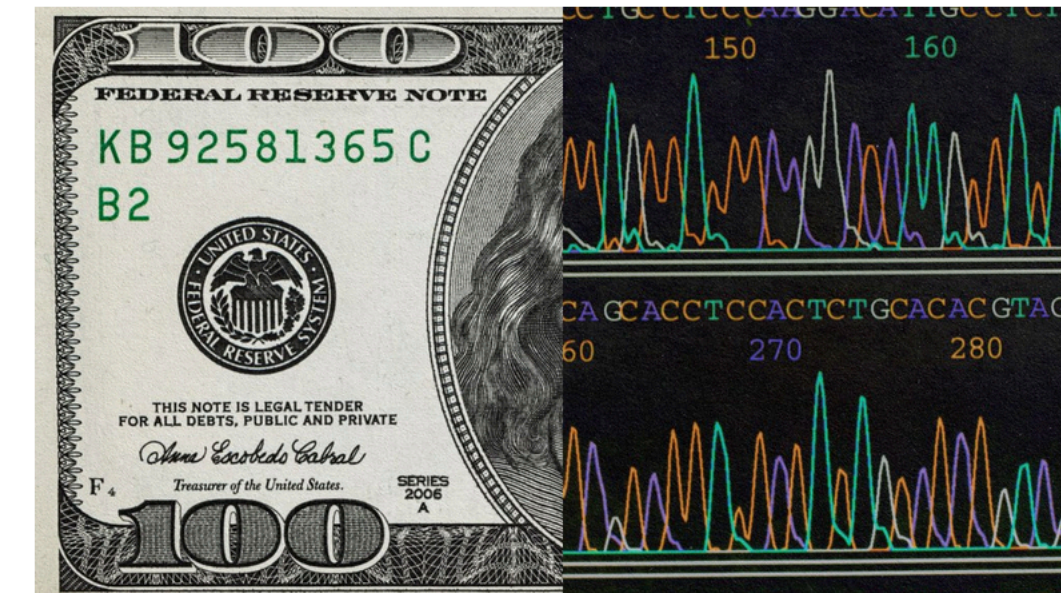
# Cena sekvenování lidského genomu



HEALTH — SEPTEMBER 19, 2023

## What will the \$100 genome mean for our society?

To put things in perspective, the cost of sequencing a single genome in 2012 was around \$10,000.



Annelisa Leinbach / Big Think; Adobe Stock

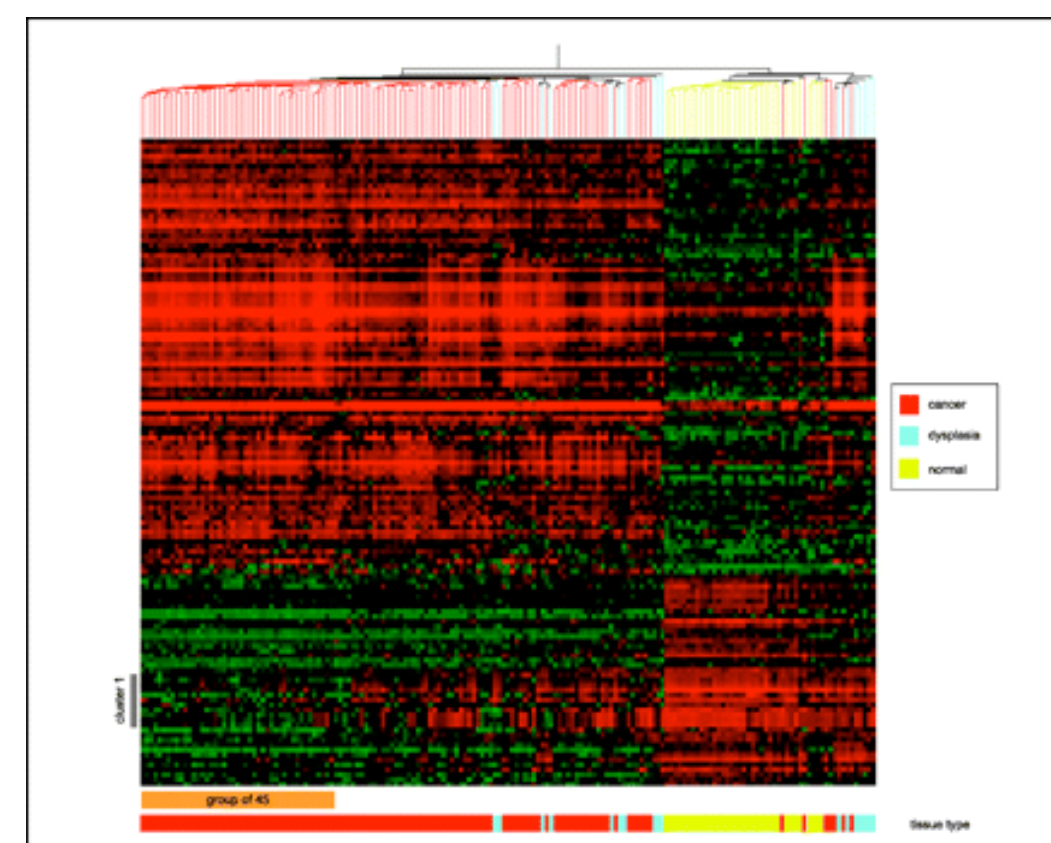
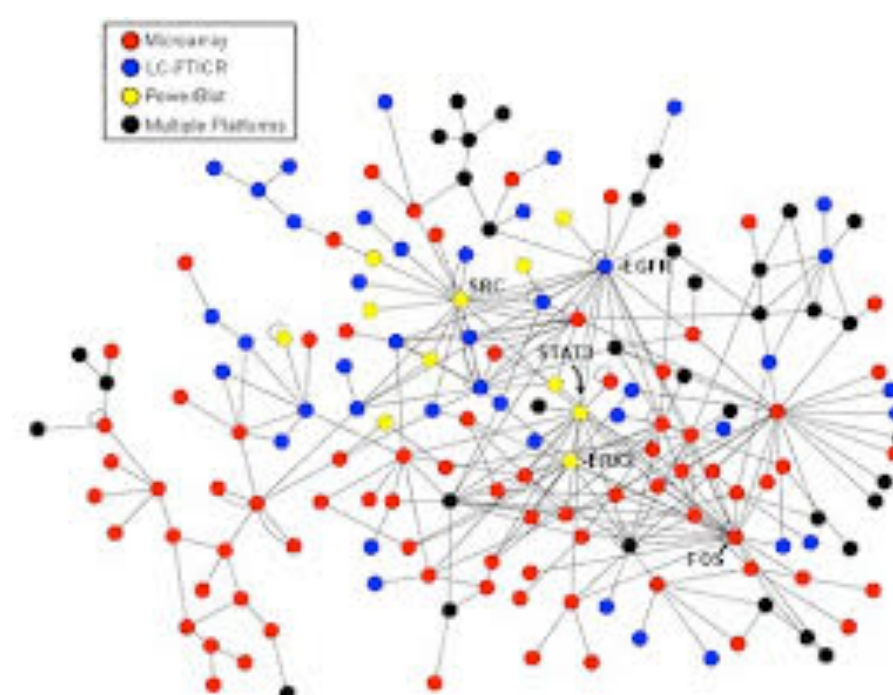
2001

2007

2015

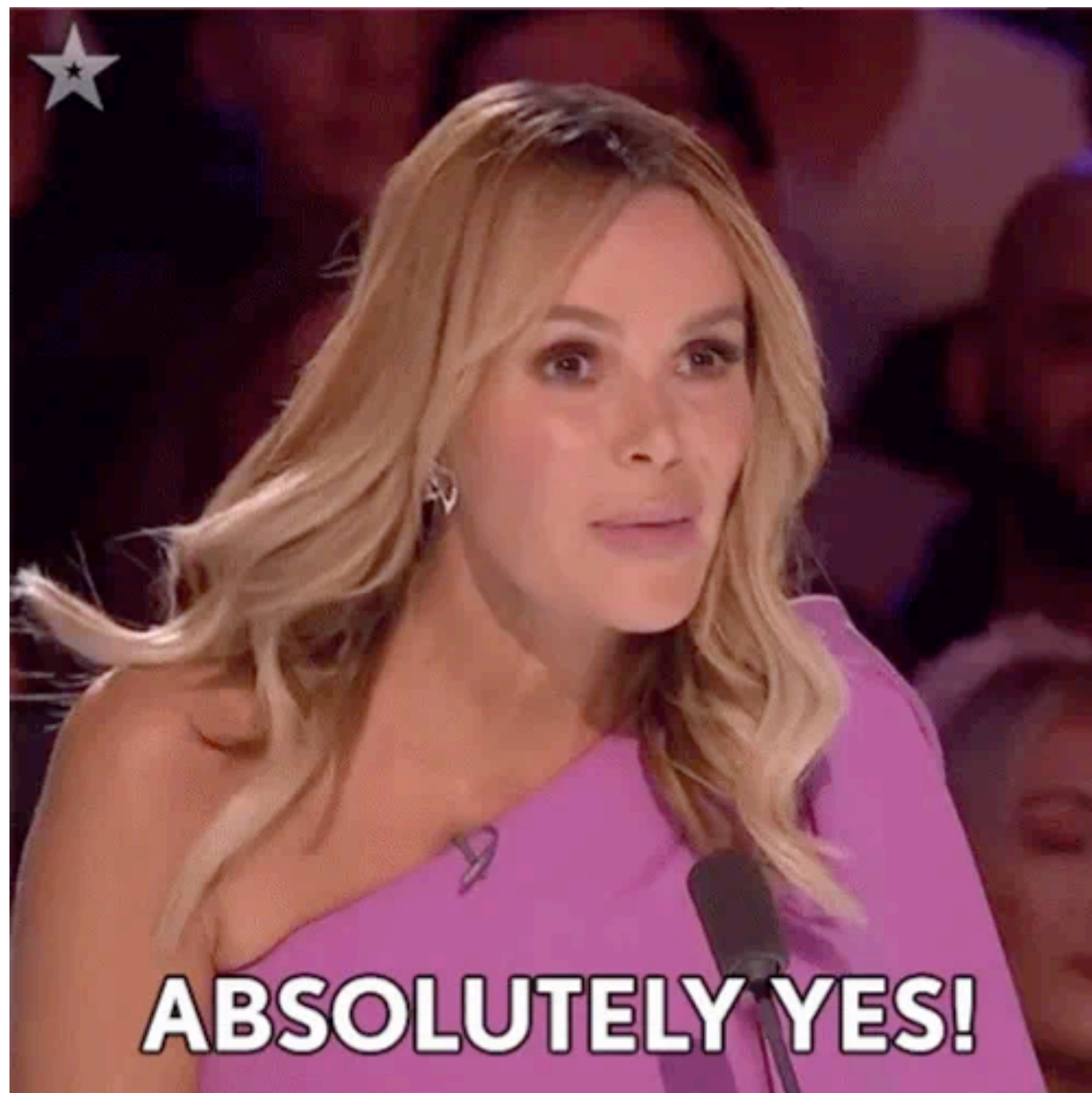
# Co všechno jsou biologická data?

- organismy
- měření, vážení
- sekvence
- 3D struktury
- interakce
- příbuzenské vztahy
- genomické mapování
- expresní profily
- ....





# Jsou biologická data dostupná?



## EMBL-EBI data resources and tools

EMBL's European Bioinformatics Institute maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Showing 1 - 15 out of 69 results

Sort by: Featured



[AlphaFold DB](#)

Database for protein structure predictions for numerous species

DATA RESOURCE CC-BY



# Jsou data volně dostupná?

400 PB of raw storage (2021)



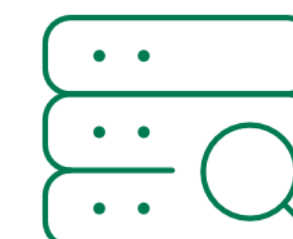
*“Scientists build on the shoulders of giants. Most often, those shoulders are data. Having these millions of structure predictions will change the face of biology. This is useful for medicine, agriculture, biotech, everything – it’s just fantastic.”*

**Janet Thornton, Director Emeritus,  
EMBL-EBI**

## OPENLY ACCESSIBLE DATA RESOURCES

EMBL-EBI maintains the world’s most comprehensive range of freely available and up-to-date molecular data resources. Developed in collaboration with scientists worldwide, these open databases, tools, and software can be accessed by anyone around the world.

**107** million



Requests to our data resource  
websites on an average day

**41** million



Unique IP addresses



Data jsou (někdy) veřejná i když je generuje soukromá společnost

AlphaFold Protein Structure Database

[Home](#) [About](#) [FAQs](#) [Downloads](#) [API](#)

# AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#)

[See search help](#) →





- V případě experimentálních strukturních dat jsou při publikaci vyžadována (už nějakou dobu) i depozice primárních experimentálních dat

Select all entries on this page

2src    CRYSTAL STRUCTURE OF HUMAN TYROSINE-PROTEIN KINASE C-SRC, IN COMPLEX WITH AMP-PNP

Xu W, Doshi A, Lei M, Eck MJ, Harrison SC  
*Mol Cell* (1999) [PMID: [10360179](#)]

Source organism: [Homo sapiens](#)

Assembly composition: protein only structure

Bound ligands: [ANP](#)

Modified residues: [PTR](#)

Assembly name: Proto-oncogene tyrosine-protein kinase Src (Preferred) [search this complex](#)

PDBe complex ID: PDB-CPX-146379 (Preferred) [search this ID](#)

PDBe-KB: [P12931](#)

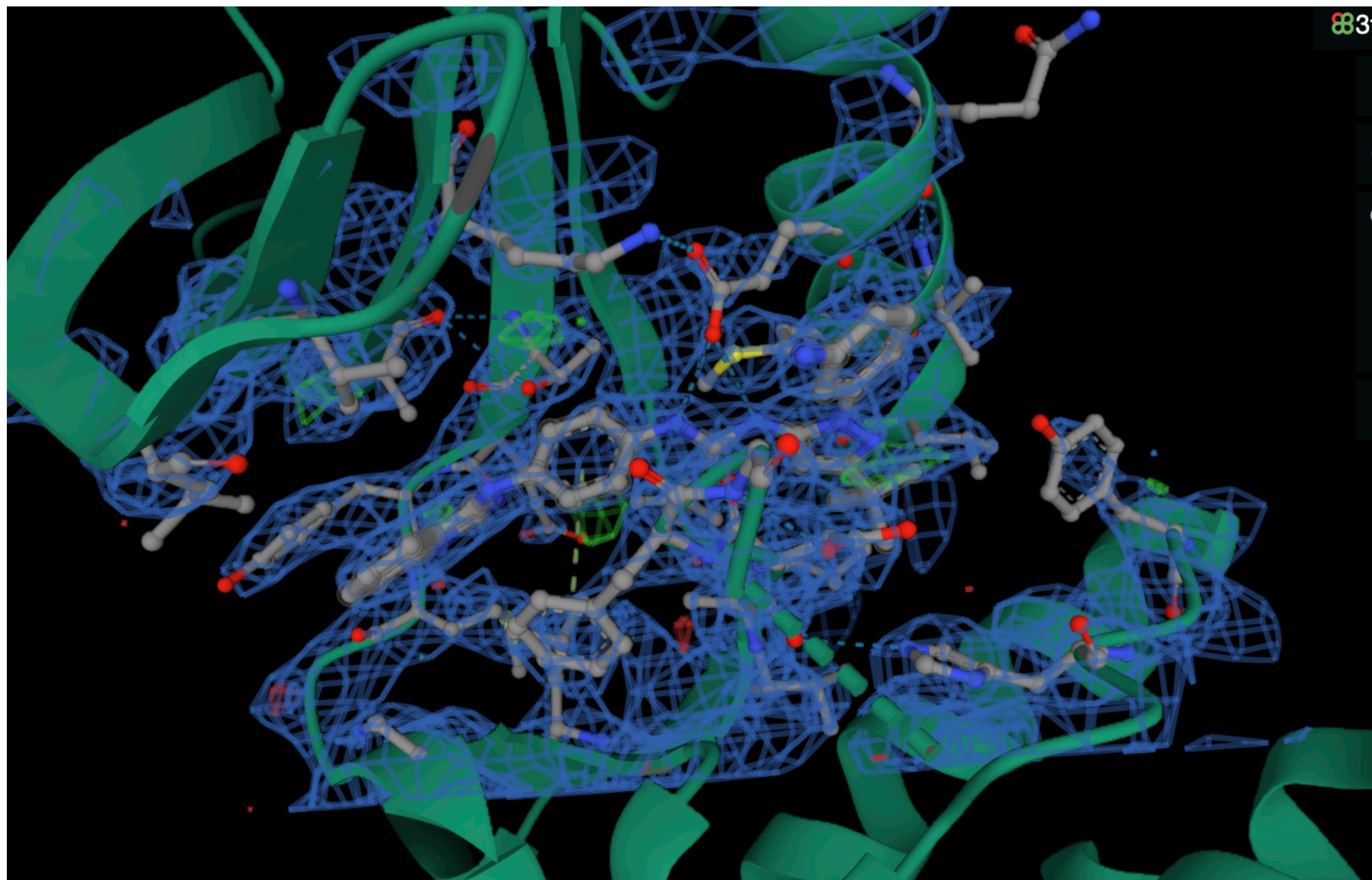
[3D Visualisation](#)    [Download files](#)

**X-ray diffraction**  
**1.5Å resolution**  
**Released:** 22 Jul 1999  
**DOI:** [10.2210/pdb2src](#)

Model geometry   
Fit model/data



# Validace na otevřených datech je možná





# Kdy data nejsou dostupná ?

- Když je generují soukromé společnosti (3D struktury proteinů s léky)
- Když jde o citlivá (lidská) data - rozvoj personalizované medicíny

## European institutes commit to data access across borders

Institutes in Finland, Germany, Norway, Spain, and Sweden became the first five nodes of the Federated European Genome-phenome Archive (Federated EGA). [→](#)

The Federated EGA is one of the largest international networks for discovery and access to sensitive human data – jointly delivered by EMBL-EBI and the Centre for Genomic Regulation in Spain.





“Store locally, access globally”

## Federated EGA vision

Federated EGA strives to support the discovery of and secure access to human data globally, while respecting national data protection regulations, with the goal of accelerating disease research and understanding and improving human health.





# Data má nejenom EBI

## What we offer



### Guidelines

Guidelines and best practices to help you manage life science data, run training courses, develop software and more.



### Web portals

Find the right software, training courses, standards and more in our interlinked portals to life science resources.



### All resources

Find compute services, databases, and the full list of resources ELIXIR coordinates.



### Partnerships with Industry and SMEs

Join events and projects that bring the private and public sectors together.



### Opportunities to work together

Join a scientific group in ELIXIR or partner with us to apply for EC funding.



### For ELIXIR members

If you work at an institute that is part of ELIXIR, then remember to take advantage of the benefits ELIXIR offers.



FACULTY OF SCIENCE  
Charles University



# ELIXIR CZ

**elixir** [Home](#) [About Us](#) [Services](#) [Projects](#) [Events](#) [News](#) [Nodes](#)  [Log in](#)



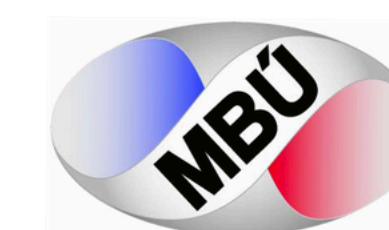
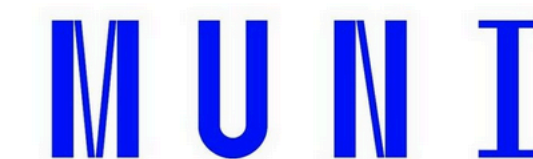
## Czech National Infrastructure for Biological Data

We are dedicated to organization, storage, sharing and facilitation of interoperability of life-science data for further processing and analysis. We respond to the needs of national scientific community, but we are also a proud member of the European Life-Science Infrastructure for Biological Information which brings together life science resources from across Europe.

[Read more about ELIXIR CZ »](#)

**5<sup>th</sup> Hands-on Computational Enzyme Design**  
**12–14 February 2024, ONLINE**

Register <https://...>



How can we help you ?



Enzyme Design Course



## Shrnutí

---

- biologických dat přibývá bezprecedentní rychlostí
- biologická data jsou dlouhodobě volně dostupná
- dostupná nejsou citlivá data, ale i zde jsou pokusy sdílet alespoň meta data
- dostupnost primárních dat ve veřejných databázích umožňuje validaci dat
- vznikají infrastruktury, které se starají o dostupnost a užitečnost nashromážděných biologických dat



FACULTY OF SCIENCE  
Charles University







FACULTY OF SCIENCE  
Charles University





FACULTY OF SCIENCE  
Charles University





FACULTY OF SCIENCE  
Charles University



---

•Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.