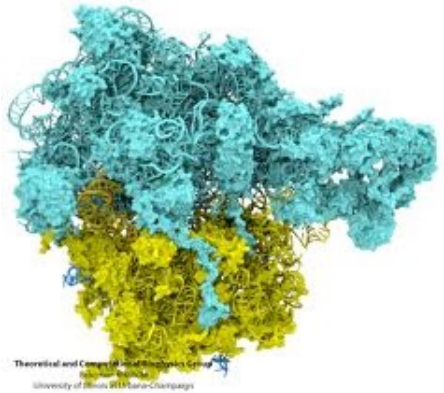
# Scientific literature analysis in computational molecular biology BIO - 468

**Prof. Patrick Barth (patrick.barth@epfl.ch)**
**Prof. Anne-Florence Bitbol (anne-florence.bitbol@epfl.ch)**
**Prof. Matteo Dal Peraro (matteo.dalperaro@epfl.ch)**

**TA: Benedikt Singer (benedikt.singer@epfl.ch)**

**Presenting the course today: Dr. Luciano Abriata (@epfl.ch)**



# About Patrick…

- studied **Physics, Chemistry and Biology** at University of Paris, ENS
- became a **biophysicist** (PhD at CEA Saclay, France, postdoc at UC Berkeley and U Washington, USA)
- associate professor in SV, Institute of Bioengineering (IBI)
- **Director of the EDCB program; steering member of the Center for Intelligent Systems at EPFL**

# About his lab …

- **Laboratory of Protein and Cell Engineering** (LPCE: lpce.epfl.ch),
- we are located at AAB ground floor and AI 2nd floor
- we develop and apply computational methods to study and design biological systems. We validate our predictions and designs using a battery of biophysical and cellular biology approaches.
- **aim**: understanding the determinants of communication in biological systems (e.g. protein allostery, mechanical force propagation) and use them for designing novel protein and cellular functions (e.g. biosensor design, CAR T-cell therapies), AI for de novo protein design

# About Anne-Florence …

- **physicist**, studied at ENS Lyon
- PhD in theoretical soft matter physics (U Paris), postdoc in theoretical biophysics at Princeton U, USA
- CNRS researcher
- assistant professor in SV, Institute of Bioengineering (IBI)

# About her lab …

- **Laboratory of Computational Biology and Theoretical Biophysics**
- located at AAB ground floor
- use theory and computation to study biology
- **aim**: understanding sequence-function relationship in proteins (prediction of protein-protein interactions), protein language models, Evolution on rugged fitness landscapes / Applications to antimicrobial resistance evolution, population expansion and evolution of the gut microbiota

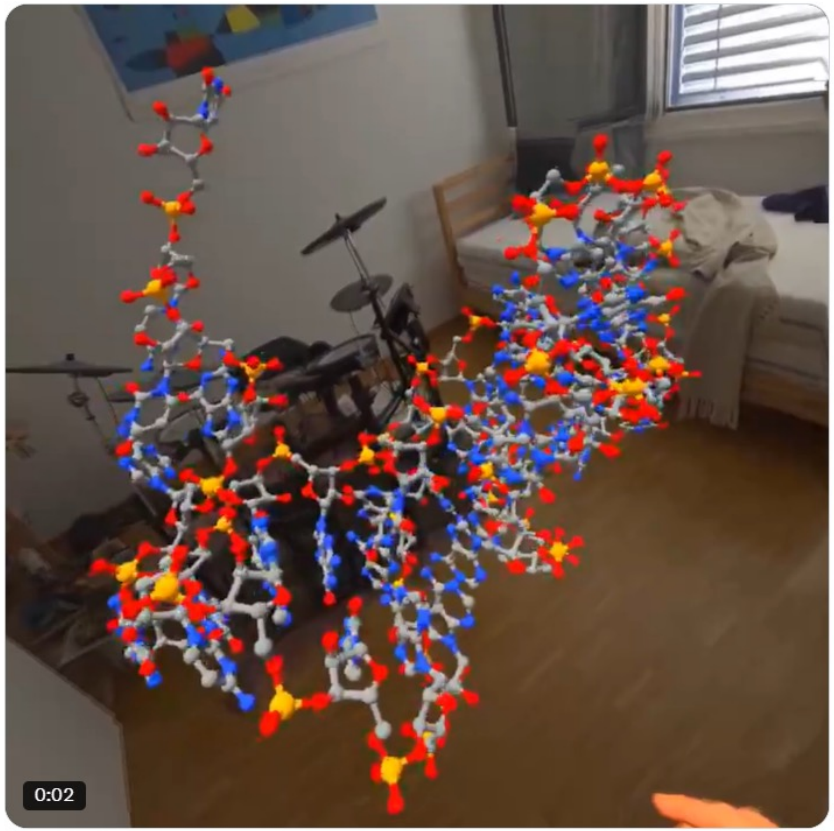
# About Matteo …

- **physicist**, studied at University of Padova, Italy
- became a **biophysicist** (PhD at SISSA in Trieste, postdoc at UPenn in Philadelphia USA)
- associate professor at SV, Institute of Bioengineering (IBI)
- group leader at the Swiss Institute of Bioinformatics (SIB)
- **Co-Director of IBI**

# About his lab …

- **Laboratory of Biomolecular Modeling** (LBM: lbm.epfl.ch),
- located at AAB  ground floor
- use theory and computation (HPC+AI) to study biology at molecular level
- have a wet-lab (AI 2nd floor) with main focus on structural biology (X-ray crystallography, NMR, EM and biophysics)
- **aim**: understanding the physico-chemical principles of biological function (e.g., molecular assembly, crowding) and use them for engineering (e.g., drug and protein design), AI for protein docking, nanopore design

# About Luciano …



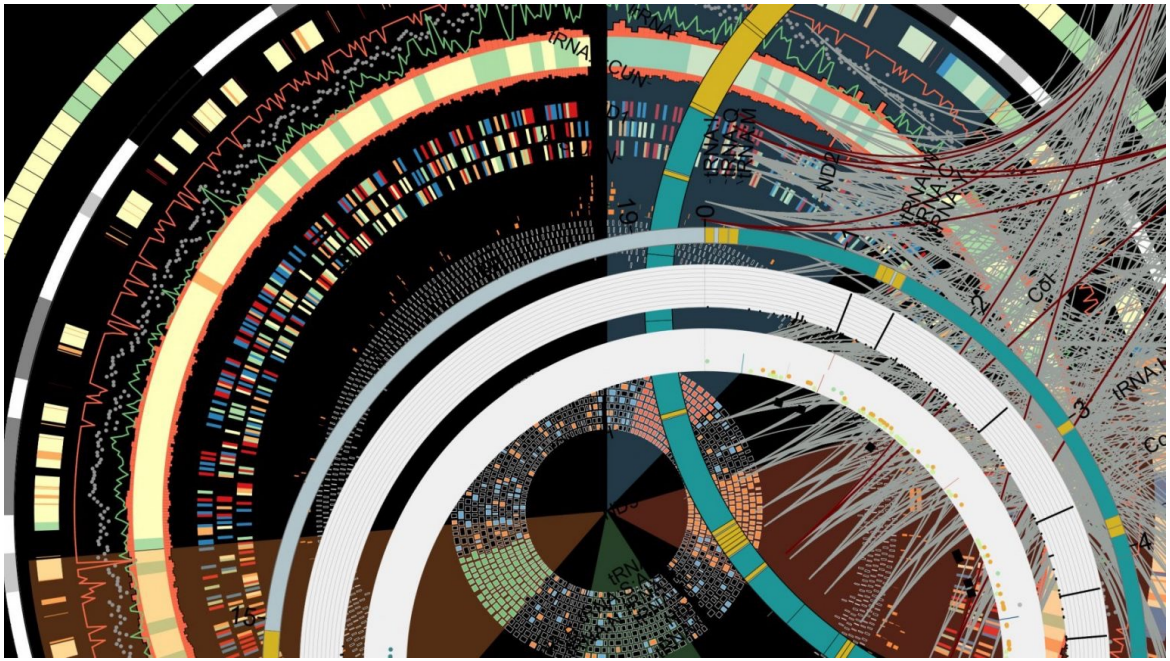

https://x.com/labriataphd/status/1813614441508090306

- **BioNMR & spectroscopies (“wet” PhD on metalloproteins)**
- **Molecular simulations, modeling and bioinformatics**
- **Human-Computer Interactions for molecular graphics and modeling**

**https://x.com/labriataphd**

**https://lucianosphere.medium.com/**

# Life Sciences PhD Programs: EDCB



The program offers students well versed in quantitative subjects and equipped with computational skills the opportunity to address the biological questions associated with processes across a broad range of temporal and spatial scales, often arising from the interpretation of massive amounts of data. EPFL provides a unique environment to form this future generation of multidisciplinary life scientists.

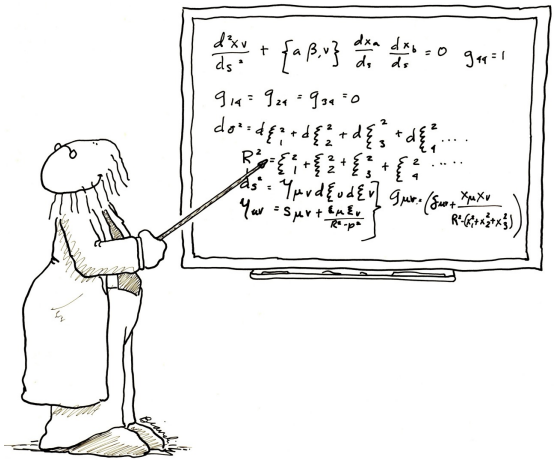
**The program's themes** include:

- Applied mathematics
- Biochemistry
- Bioinformatics
- Biophysics
- Computational neurosciences
- Developmental biology
- Digital epidemiology
- Genomics and functional genomics
- Machine learning
- Microbiology
- Neural data science
- Protein design
- Structural biology
- Systems biology

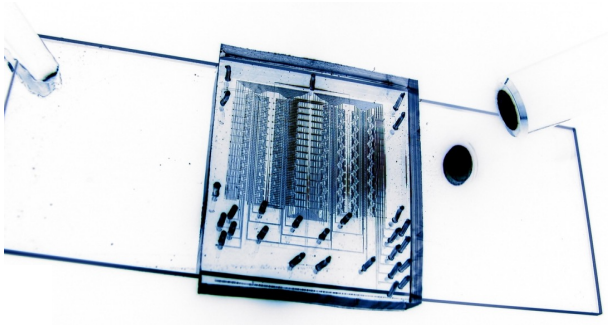
**Ever-growing databases + high-throughput experiments require:**

**Automated analyses / pattern recognition / AI / big data / data crunching**

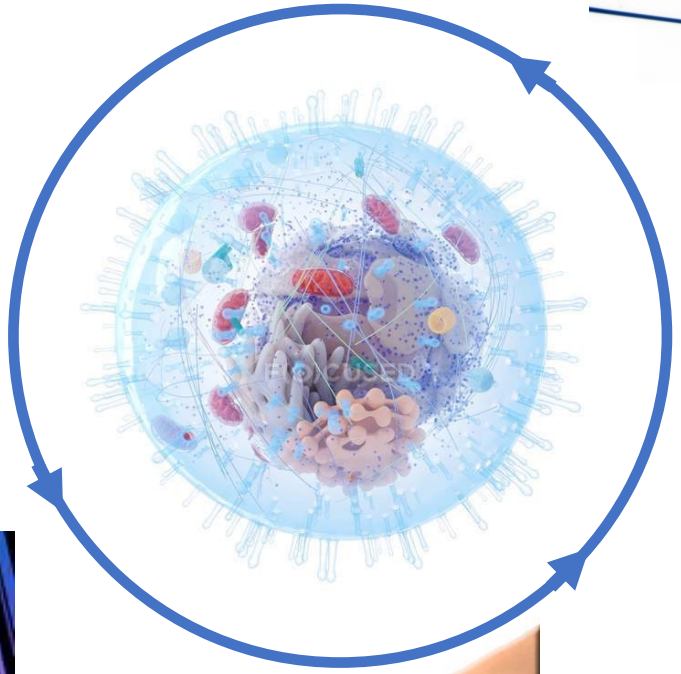
# Quantitative molecular biology

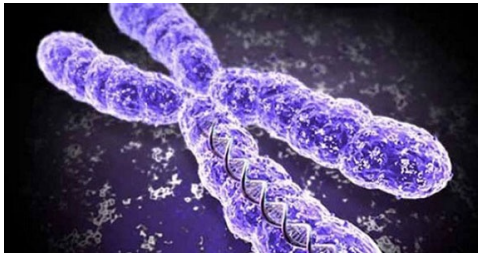


Theory and Computation

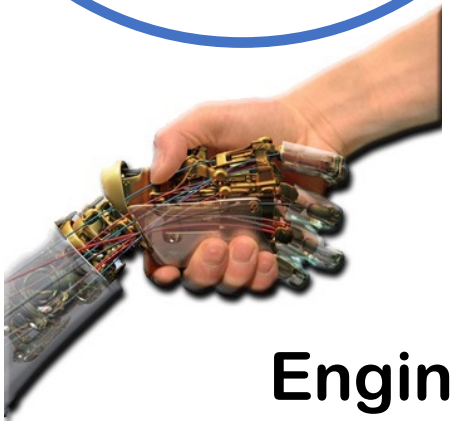


Large scale omics / big data



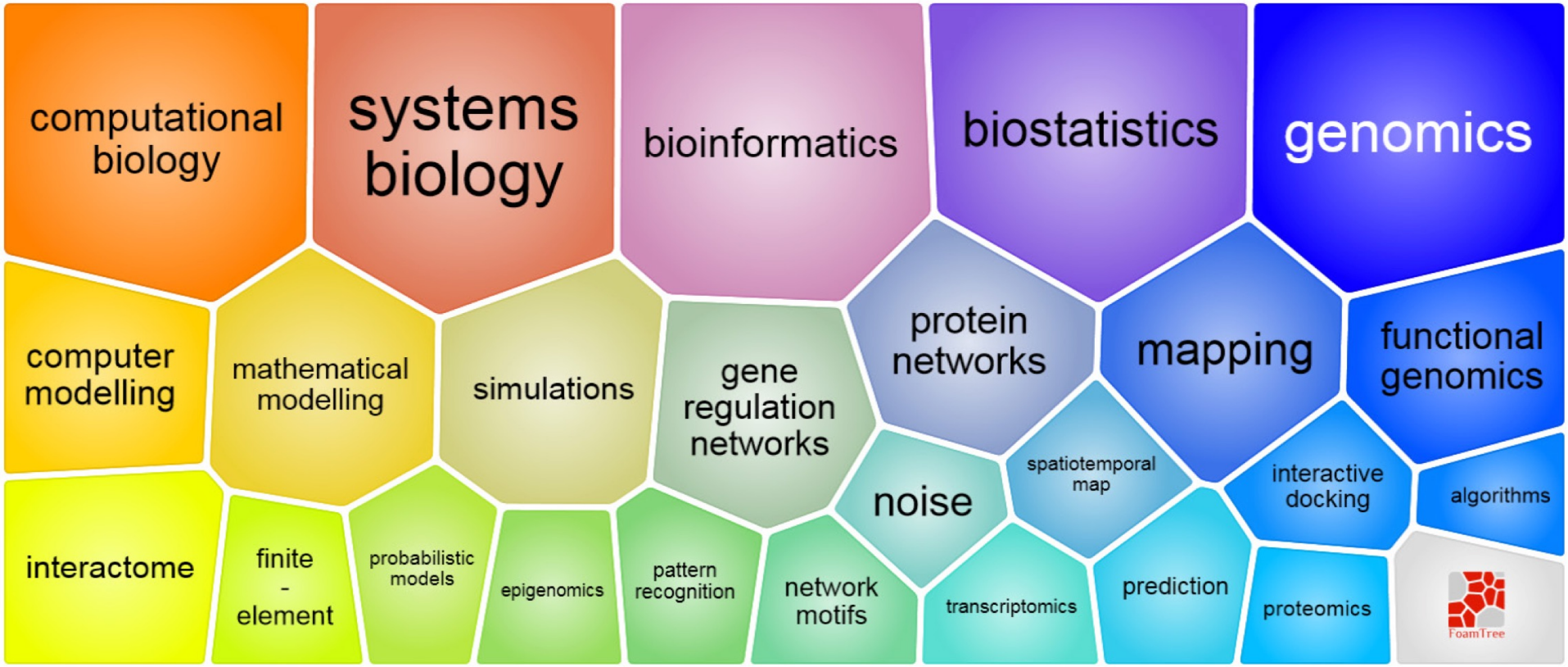


Molecular biology



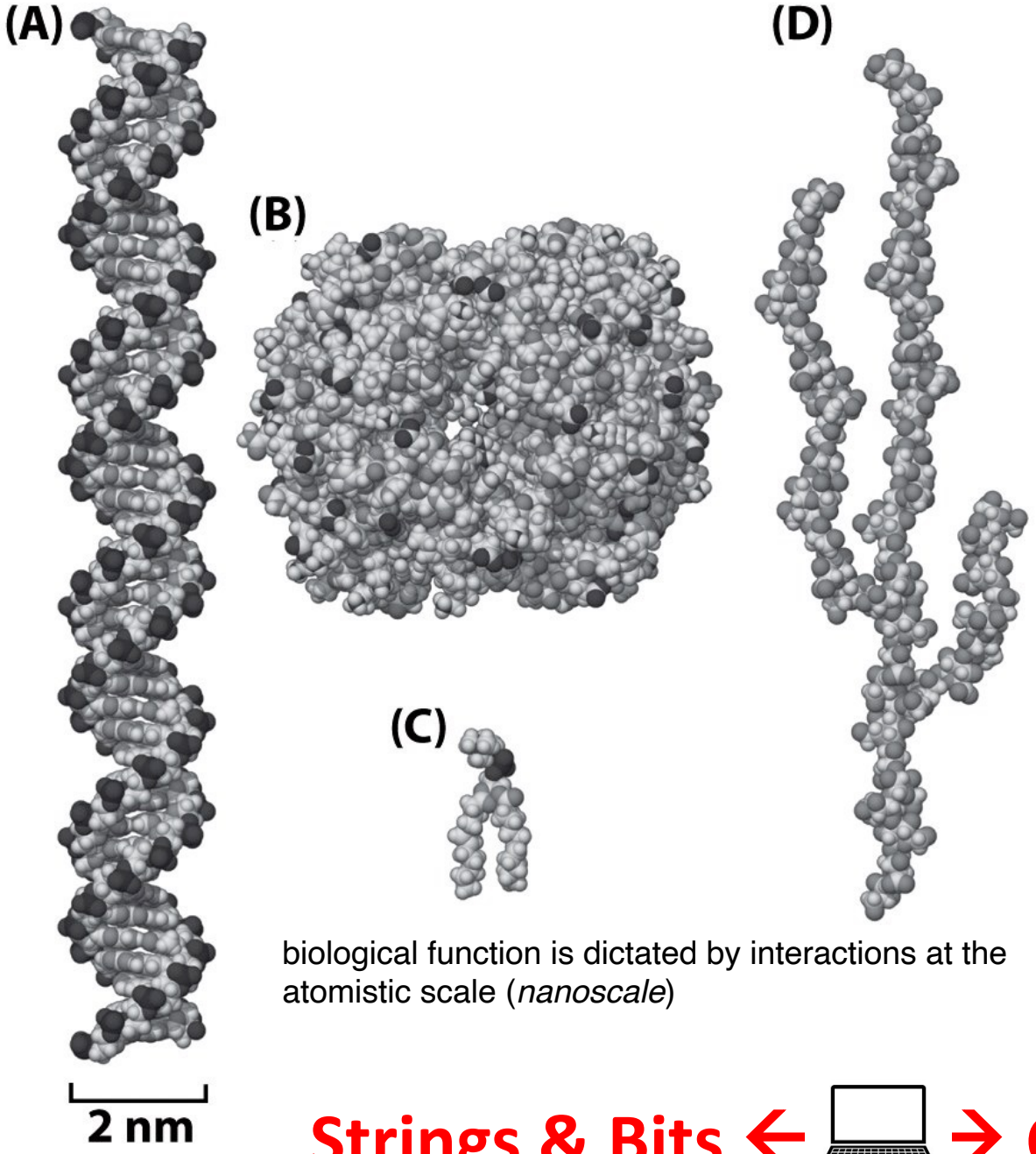
Engineering

# The many facets of computational biology




19-07-2017 I Clusters of most used keywords in ERC projects in the field of computational biology

- different theories
- different resolutions
- different computation
- different data types
- deep learning
- AI …



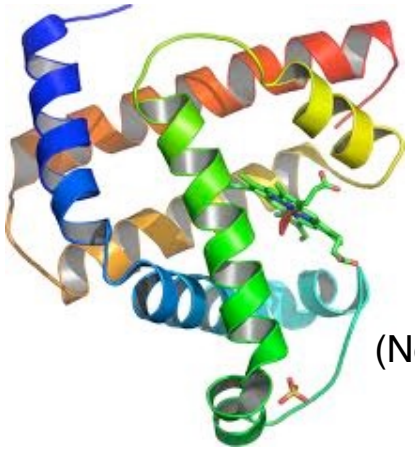

biological function is dictated by interactions at the atomistic scale (*nanoscale*)

# Biological matter

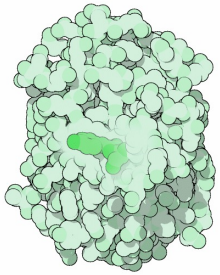
**A.** nucleic acids
**B.** proteins
**C.** lipids
**D.** carbohydrates
…

**Strings & Bits ← → Chemistry & Physics**

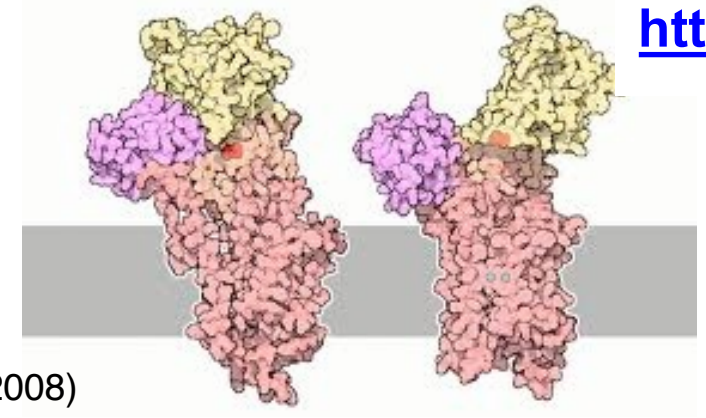
http://www.rcsb.org



**myoglobin**
(Nobel in Chemistry 1962)

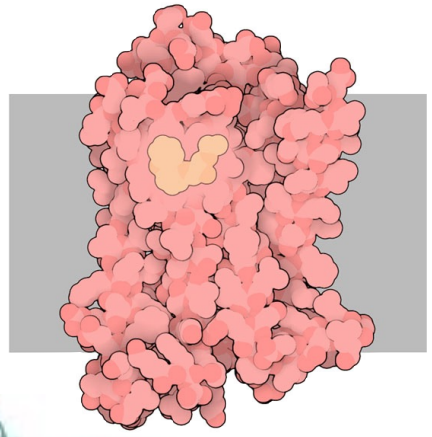


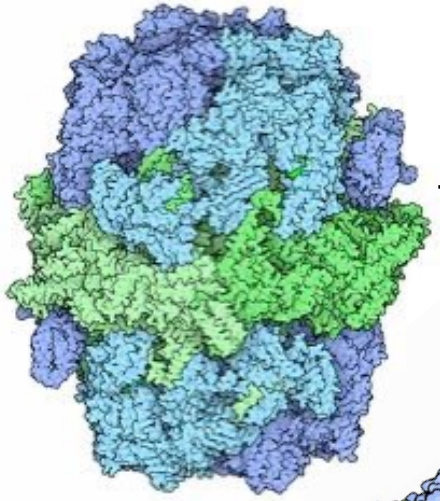
**GFP**
(Nobel in Chemistry 2008)



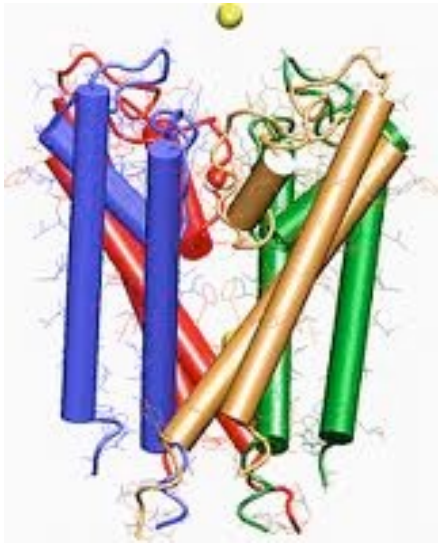
**calcium pump**

**GPCRs**
(Nobel in Chemistry 2012)

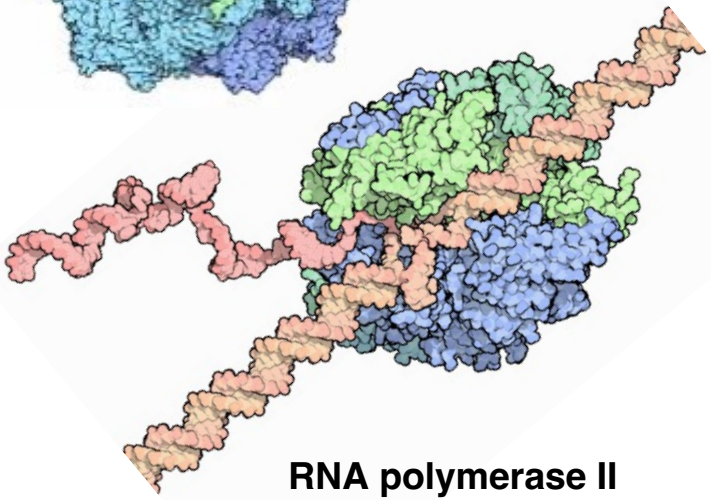




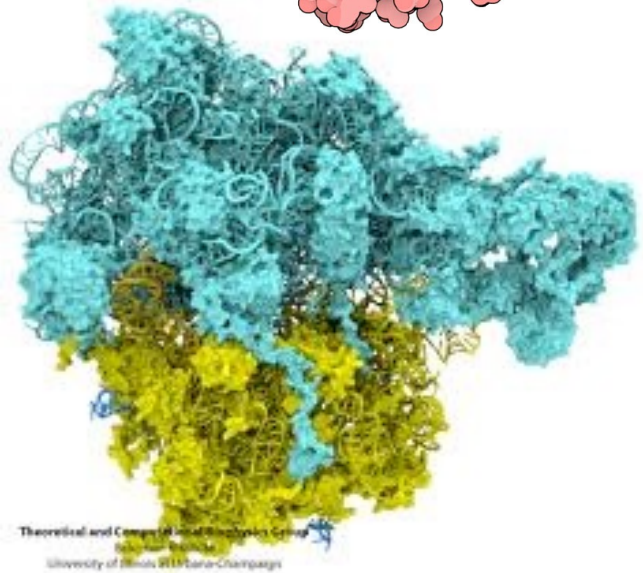
**fatty acid synthase**



**KcsA potassium channel**
(Nobel in Chemistry 2003)

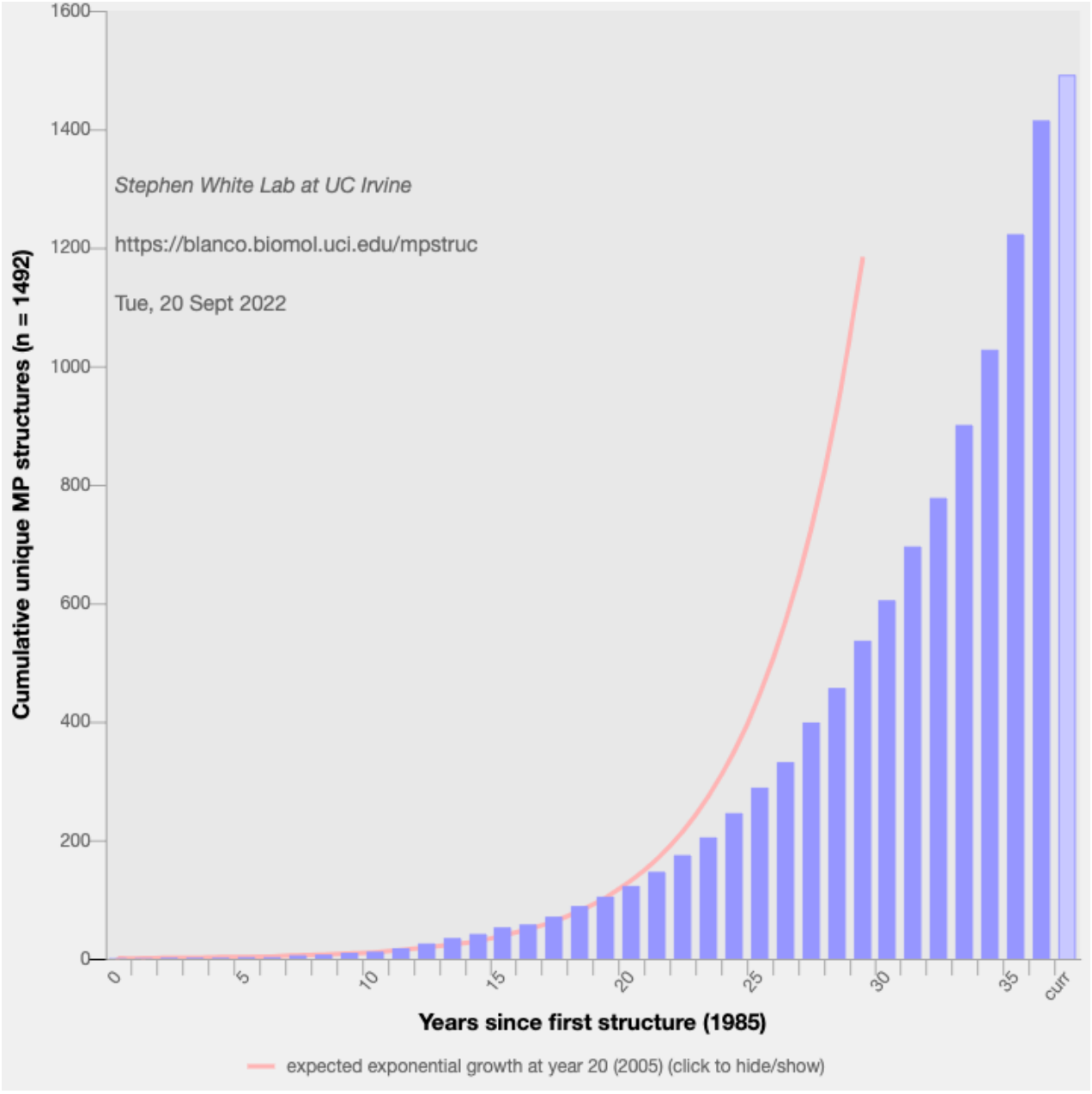


**RNA polymerase II**
(Nobel in Chemistry 2006)



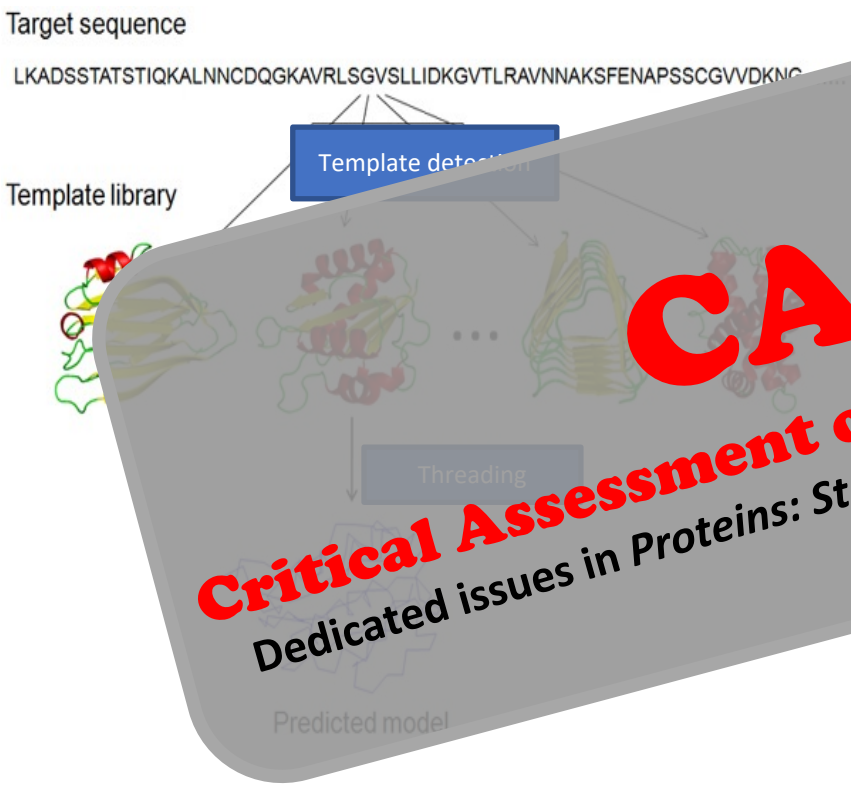
**ribosome**
(Nobel in Chemistry 2009)

# Rapid expansion of the protein structure database



# What if there's no experimental structure ?

## *Homology modeling*




- **Finding structurally similar template in PDB**
- **Correctly aligning sequences**
- **Threading**

## *Template-less prediction*

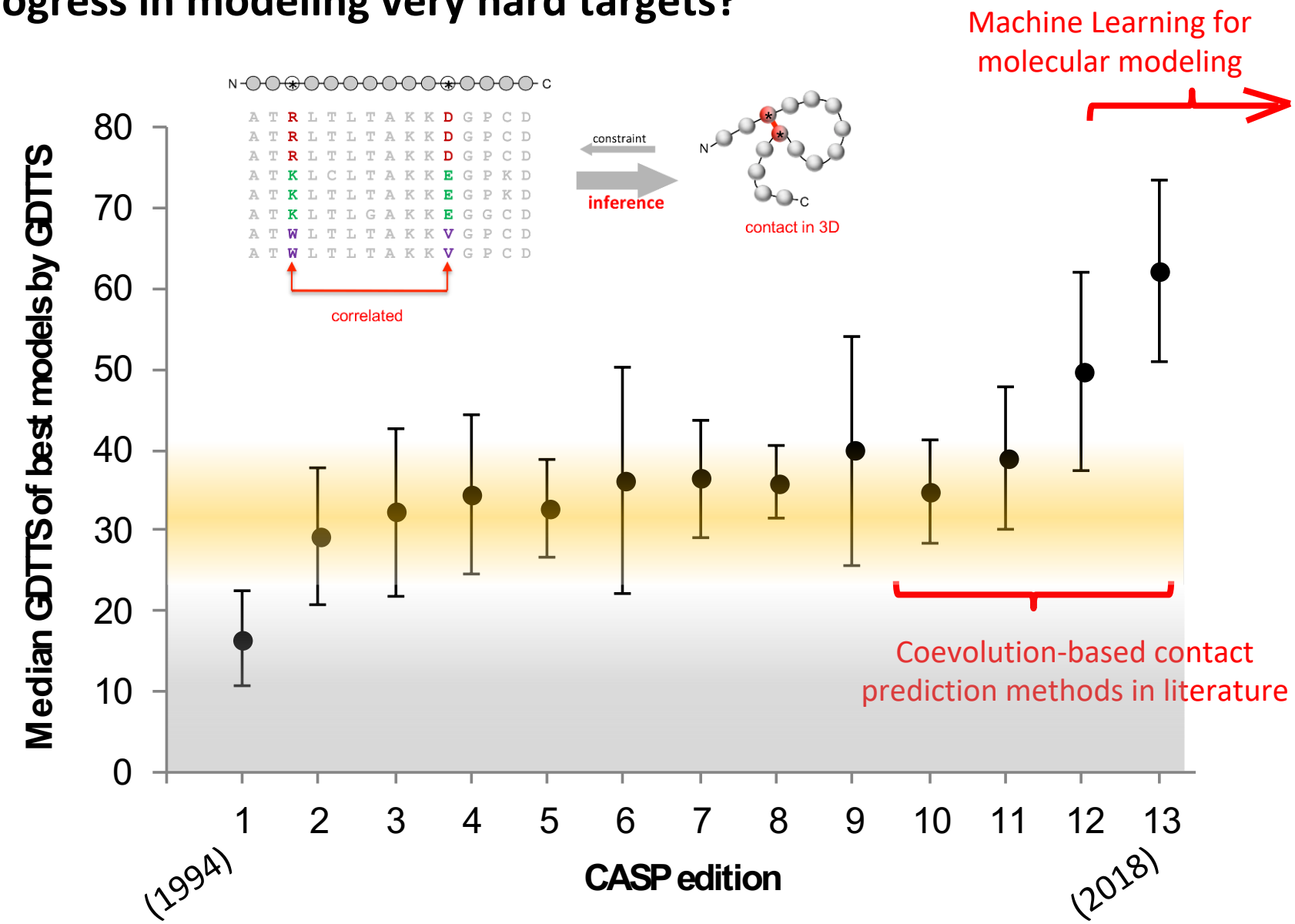
- **Physics- or Data- based conformational potential**
- **Sampling capacity**
- **Scoring to detect folded structure**
- **External data? (experiments, coevolution, etc.)**

**CASP**

**Critical Assessment of Structure Prediction**

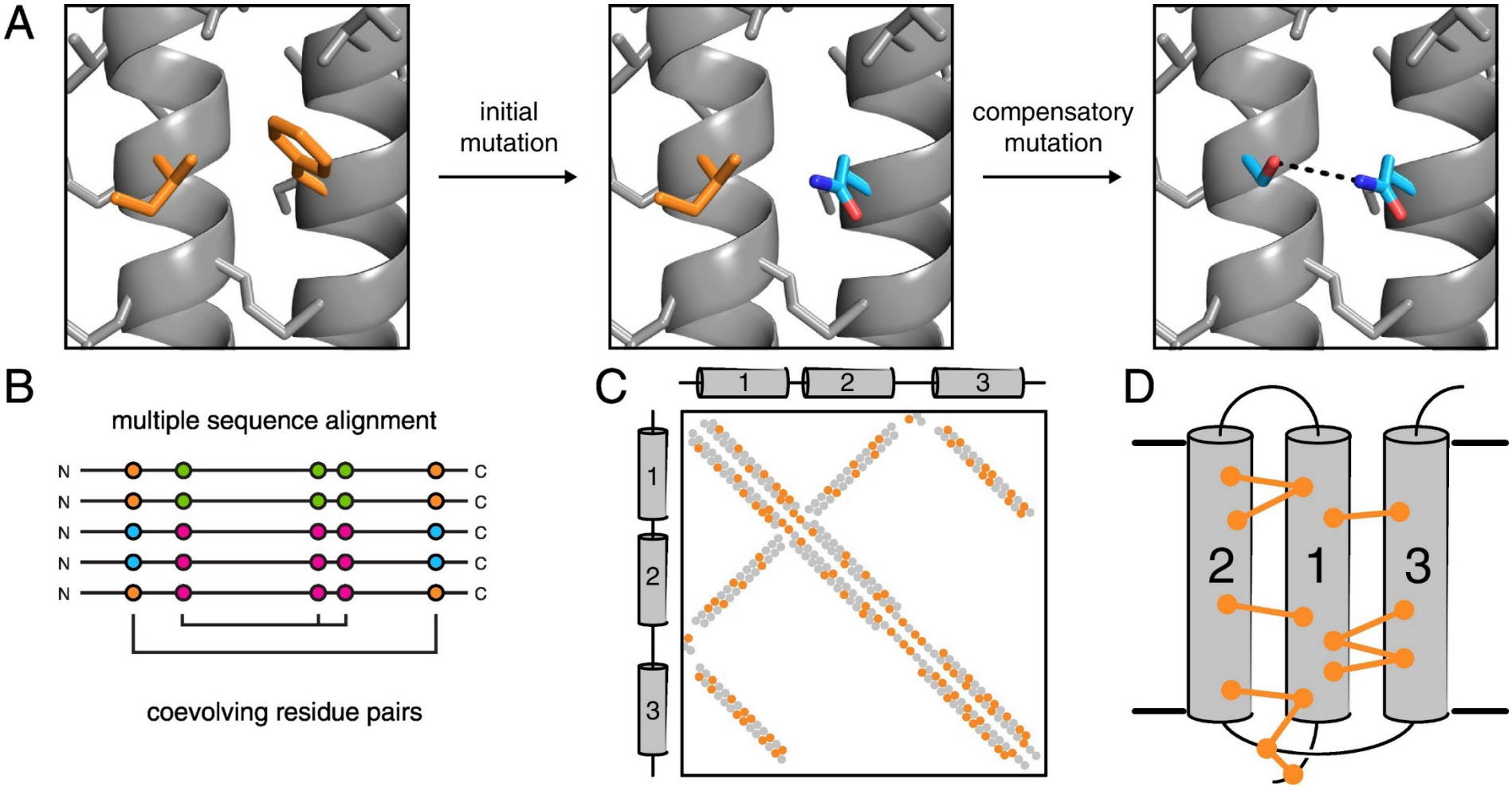
***Dedicated issues in Proteins: Structure, Function and Bioinformatics***

# Progress in modeling very hard targets?

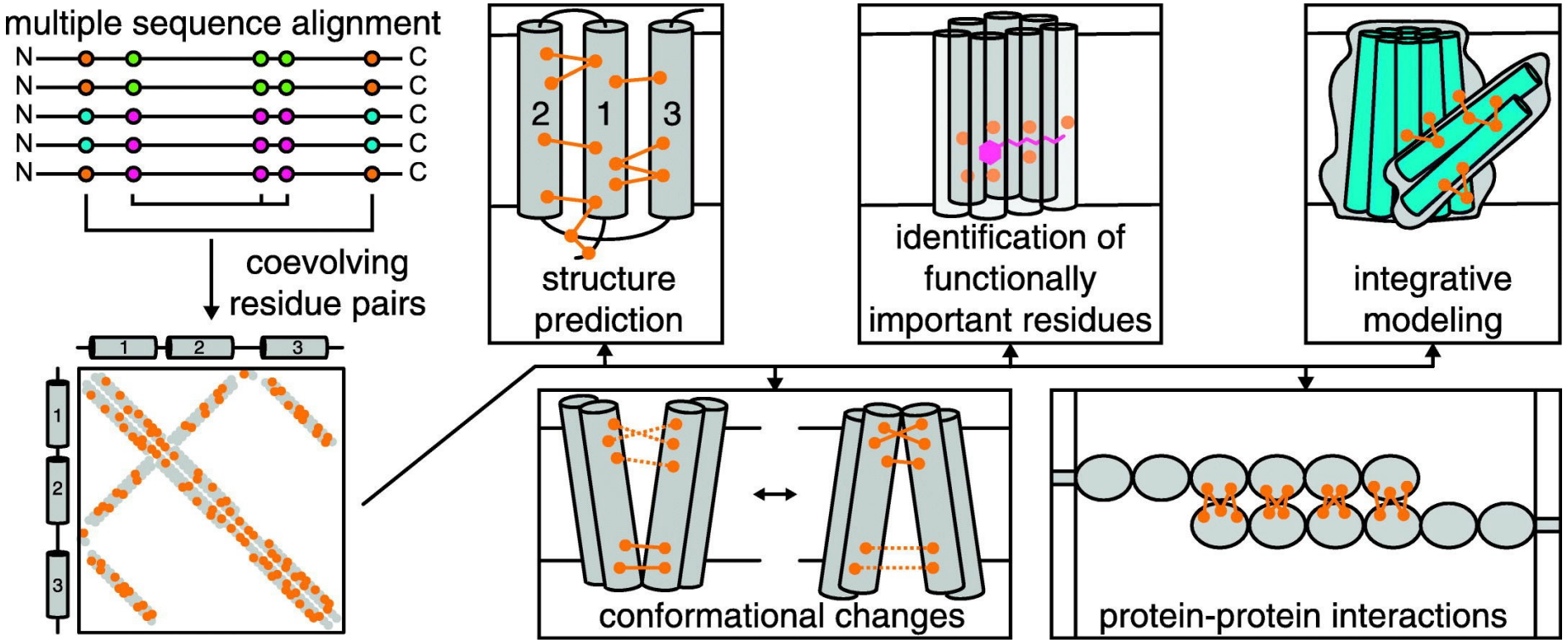



Abriata, Tamo, Dal Peraro, *Proteins* 2019 (official CASP13 assessment)
https://onlinelibrary.wiley.com/doi/10.1002/prot.25787

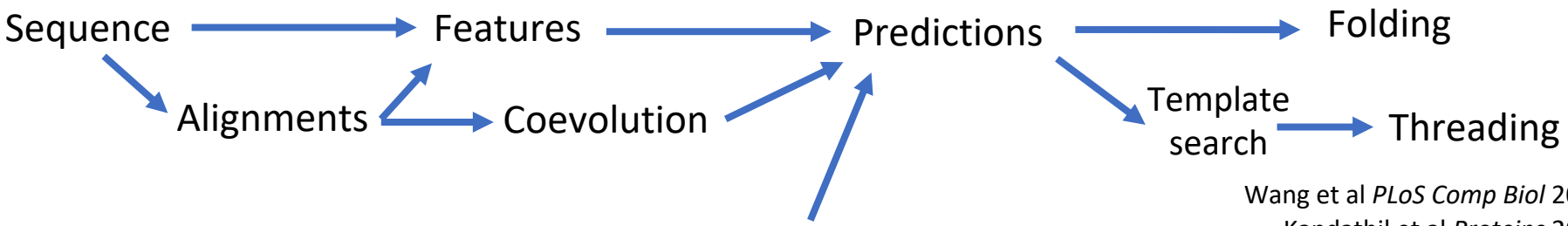
# Accelerated genome sequencing enables the inference of functional couplings in proteins



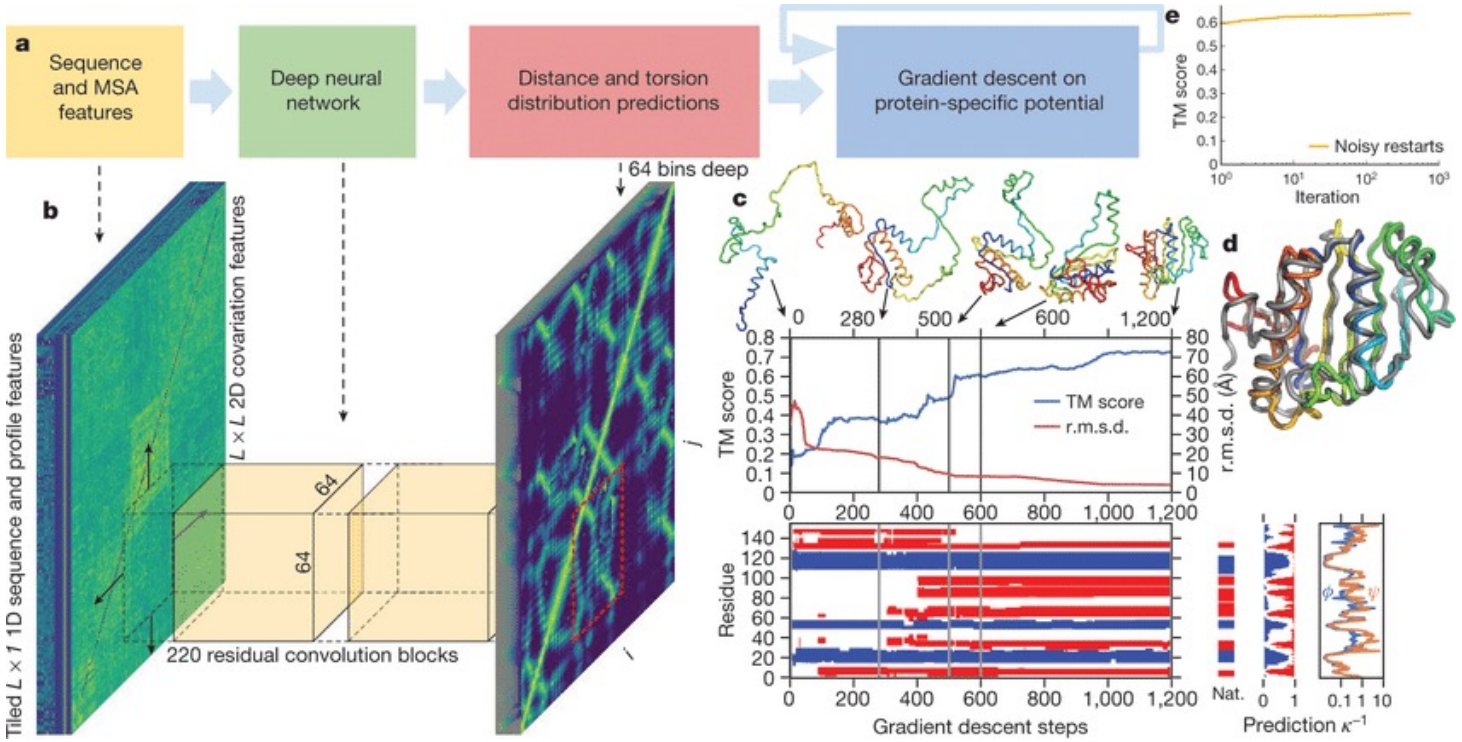
# Accelerated genome sequencing enhances protein structure and function predictions



# Machine learning for Protein structure by the time of AlphaFold 1



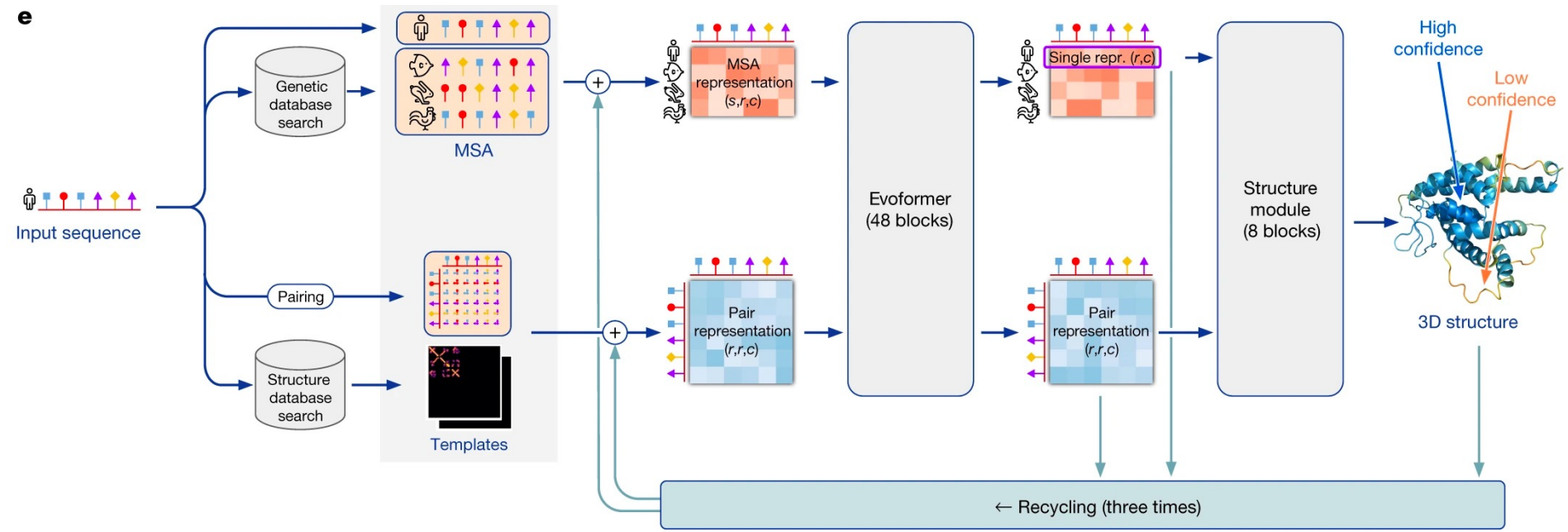

Wang et al *PLoS Comp Biol* 2017
Kandathil et al *Proteins* 2019
Hou et al *Proteins* 2019
Xu *PNAS* 2020
Senior et al *Nature* 2020
Yang et al *PNAS* 2020

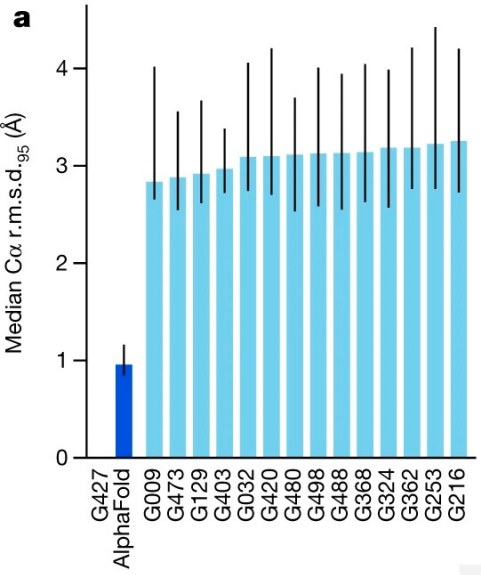


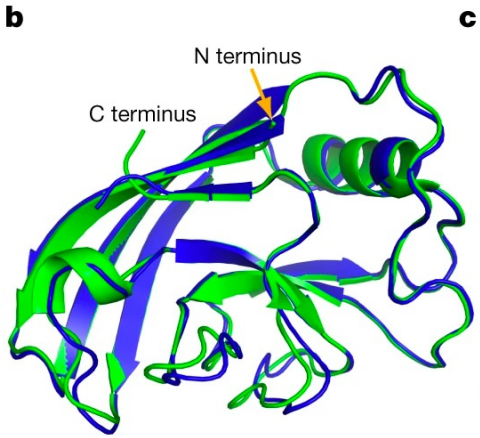

**With AF 1, Deepmind pushed current tech to the max. Not revolutionary yet.**

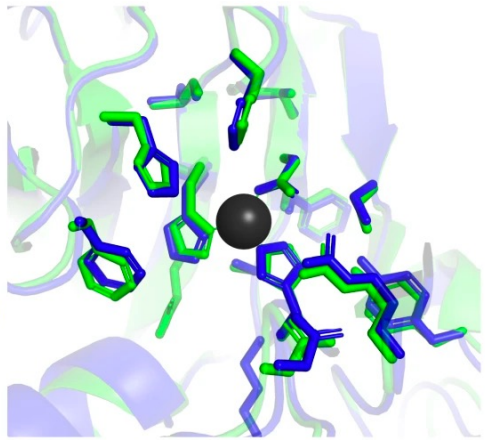
http://labriataphd.altervista.org/

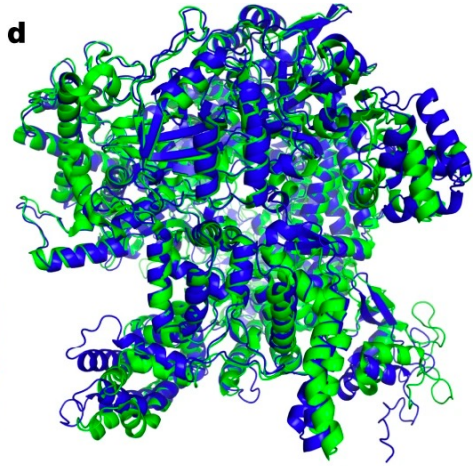
# Alphafold 2: AI applied to protein structure prediction



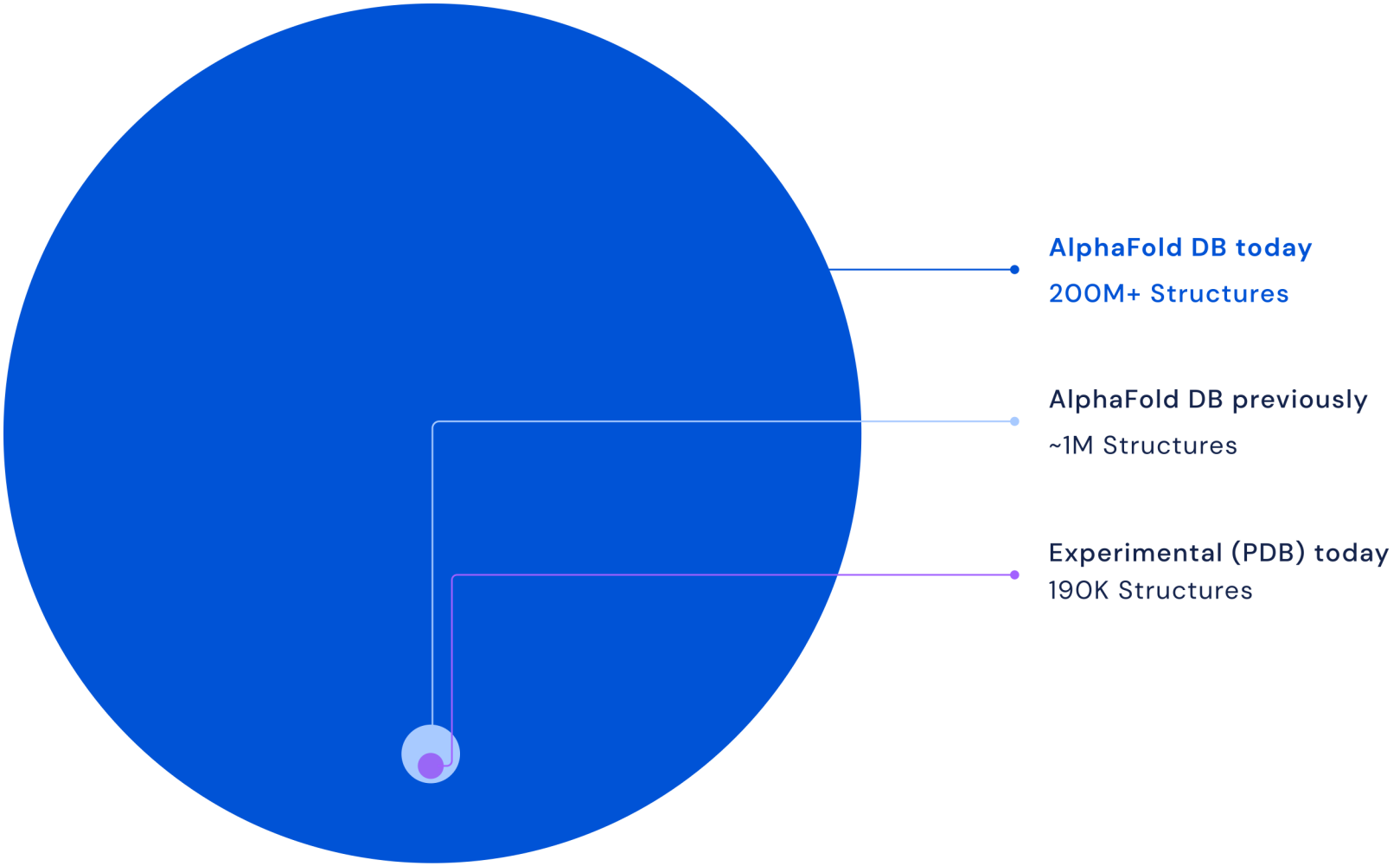



# > 200 millions protein structures predicted by Alphafold



# Computational approaches to structural biology, molecular functions and design

**Methods**

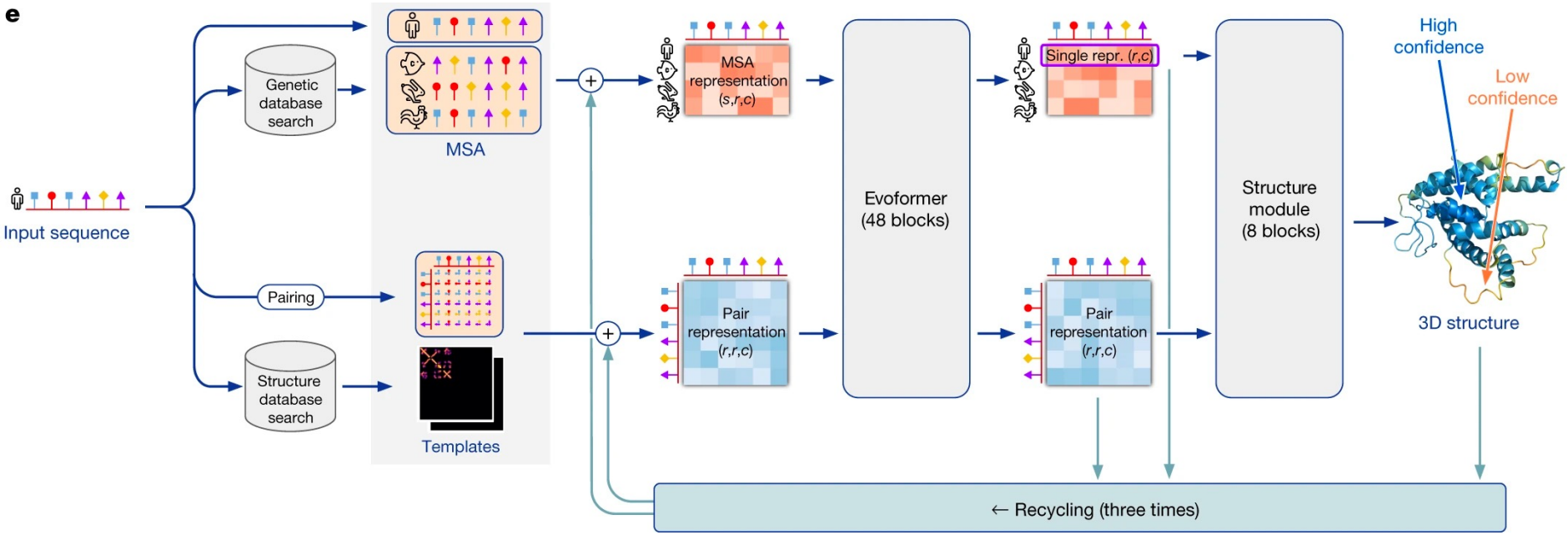
- integrative modeling
- protein structure prediction
- comparative modeling
- molecular simulations
- protein design and engineering
- structural bioinformatics
- virtual screening drug design
- deep learning / neural networks
- protein sequence mining and evolution

**Applications**

- molecular structure and function
- protein folding
- molecular binding
- protein networks and pathways
- drug development
- biosensors
- therapeutics
- …

**Learning by reading** based on a series of papers touching all the different aspects of computational methods applied to structural biology

# Can I reverse Alphafold 2 to design new proteins?




## In theory yes; in practice better AI models came out

## As of today we can even design and predict structures of proteins complexed with various other molecules

# A course with a "Journal Club" format

- **every week a group (2 students) presents a relevant paper (authors):**
  - introduce the field of research and the methods
  - explain the rationale behind the research project
  - discuss the main results and defend their weak points
  - elaborate on the impact of the work
  - prepare a ~30' presentation with slides from the paper or related ones to support their discussion

- **the rest of the class ask constructive critics (reviewers):**
  - assess the results and claims of the paper
  - highlight the strengths and weaknesses, flaws if any
  - analyze impact and alternative applications
  - write a report on the paper that is handled in class

**Grading** is based on both the presentations, the reports, and the engagement on the weekly discussions. TA will support you during the reading in case things are not clear, and will direct you to the right reading.

# A course with a "Journal Club" format

- **Details and papers on moodle:**

  **https://moodle.epfl.ch/enrol/index.php?id=14764**

- **Presentations through permanent zoom:**

  **https://epfl.zoom.us/j/68511447193**

# Questions?

patrick.barth@epfl.ch

anne-florence.bitbol@epfl.ch

matteo.dalperaro@epfl.ch

benedikt.singer@epfl.ch

luciano.abriata@epfl.ch