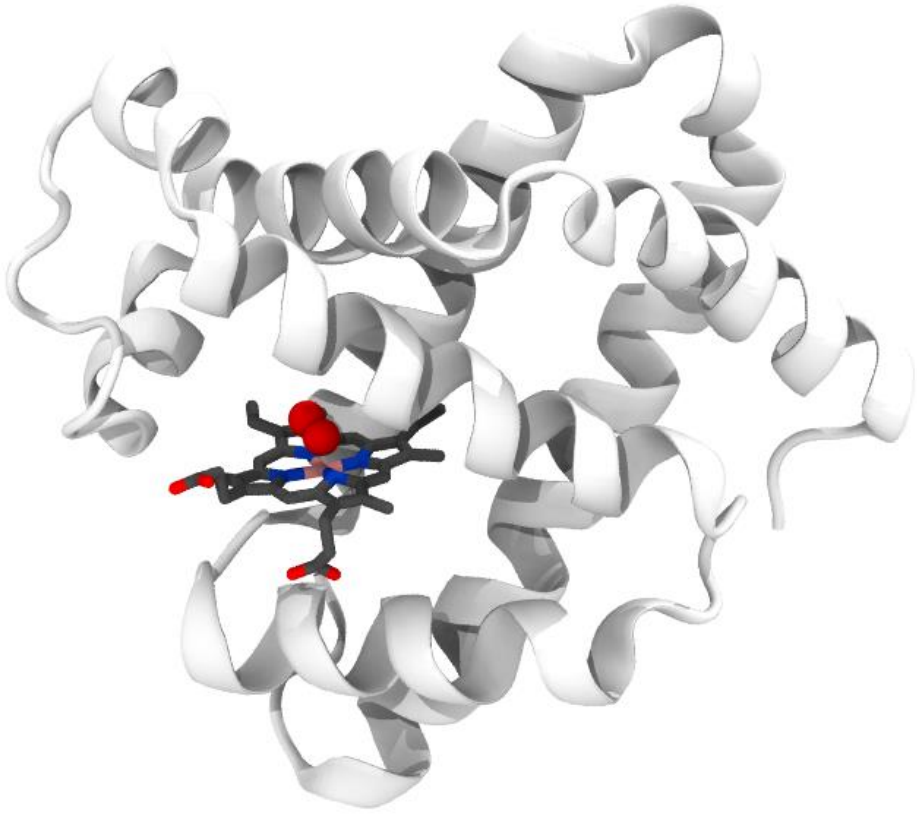
Molecular Biophysics

Protein structure and dynamics

Summary of last lesson

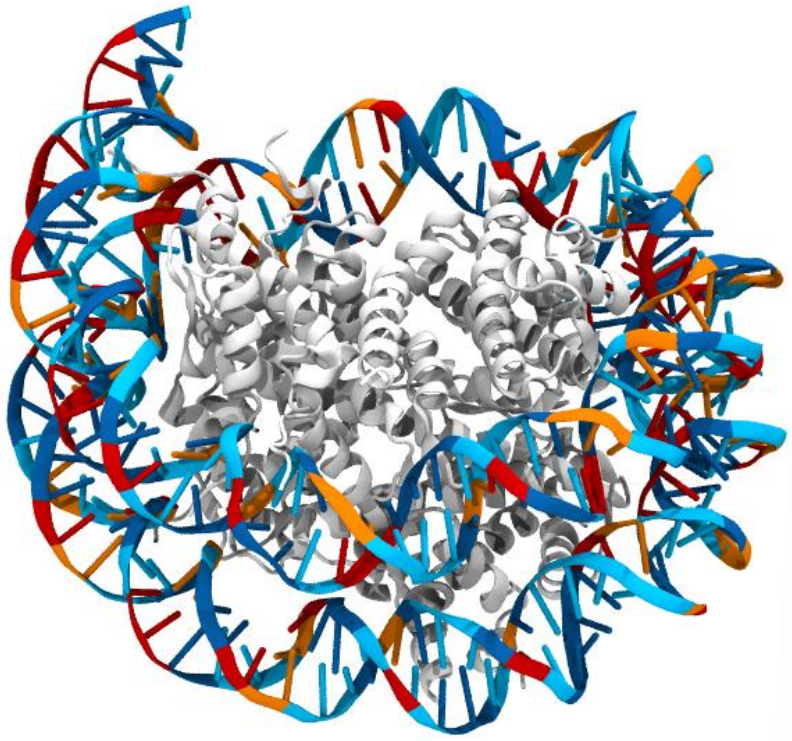
- Proteins are amino acids polymers
- Proteins fold in three dimensional structures determined by their amino acids sequence
- Protein folding is driven by entropy (hydrophobic collapse)
- Folding is not a random process

The structure determines the function



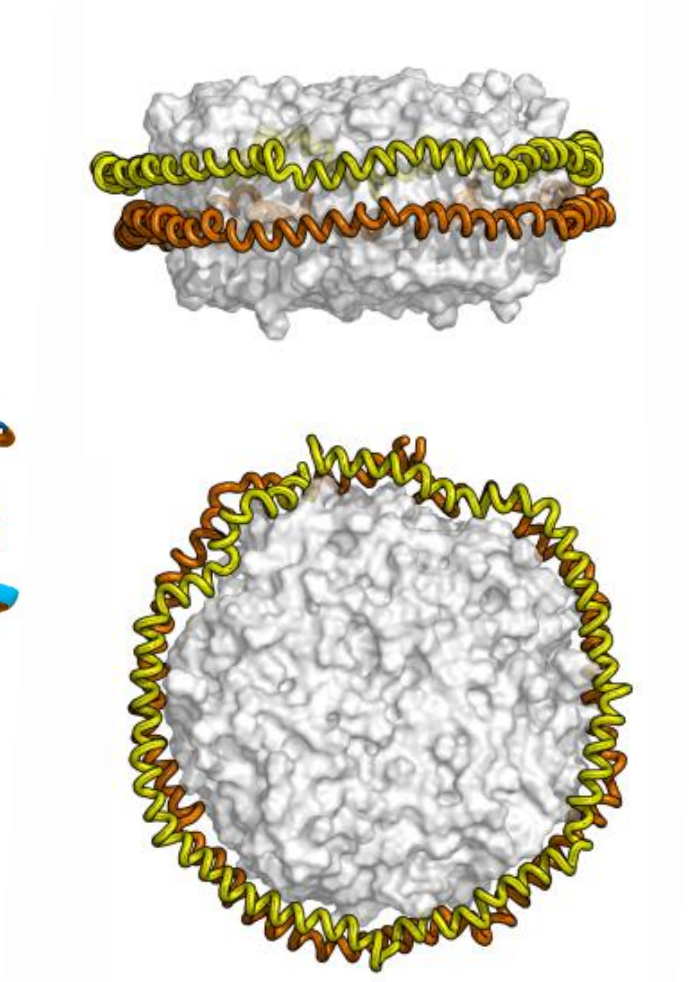
Myoglobin (PDB: 1MBO)

J.C. Kendrew et al., *A three-dimensional Model of the Myoglobin Molecule obtained by X-Ray Analysis,* Nature, 1958



Nucleosome (PDB: 5CPI)

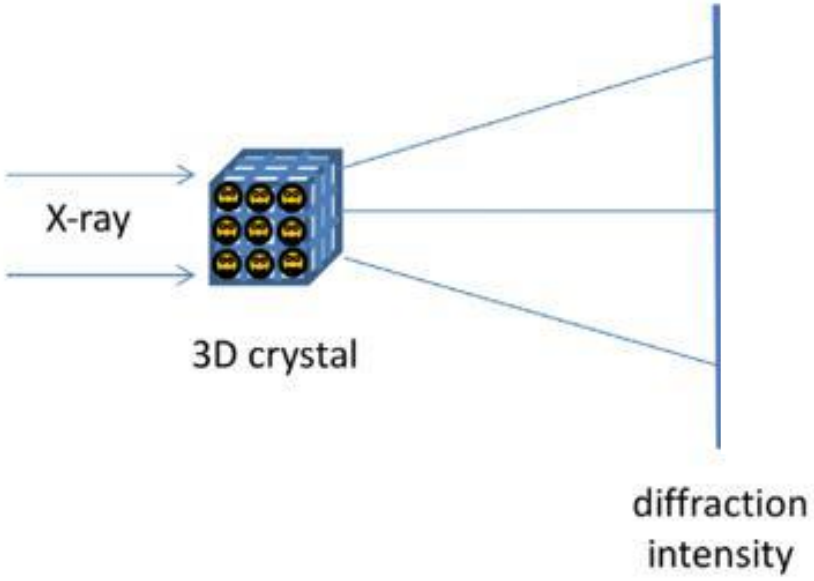
K. Luger et al., *Crystal Structure of the nucleosome core particle at 2.8 A resolution,* Nature, 1997

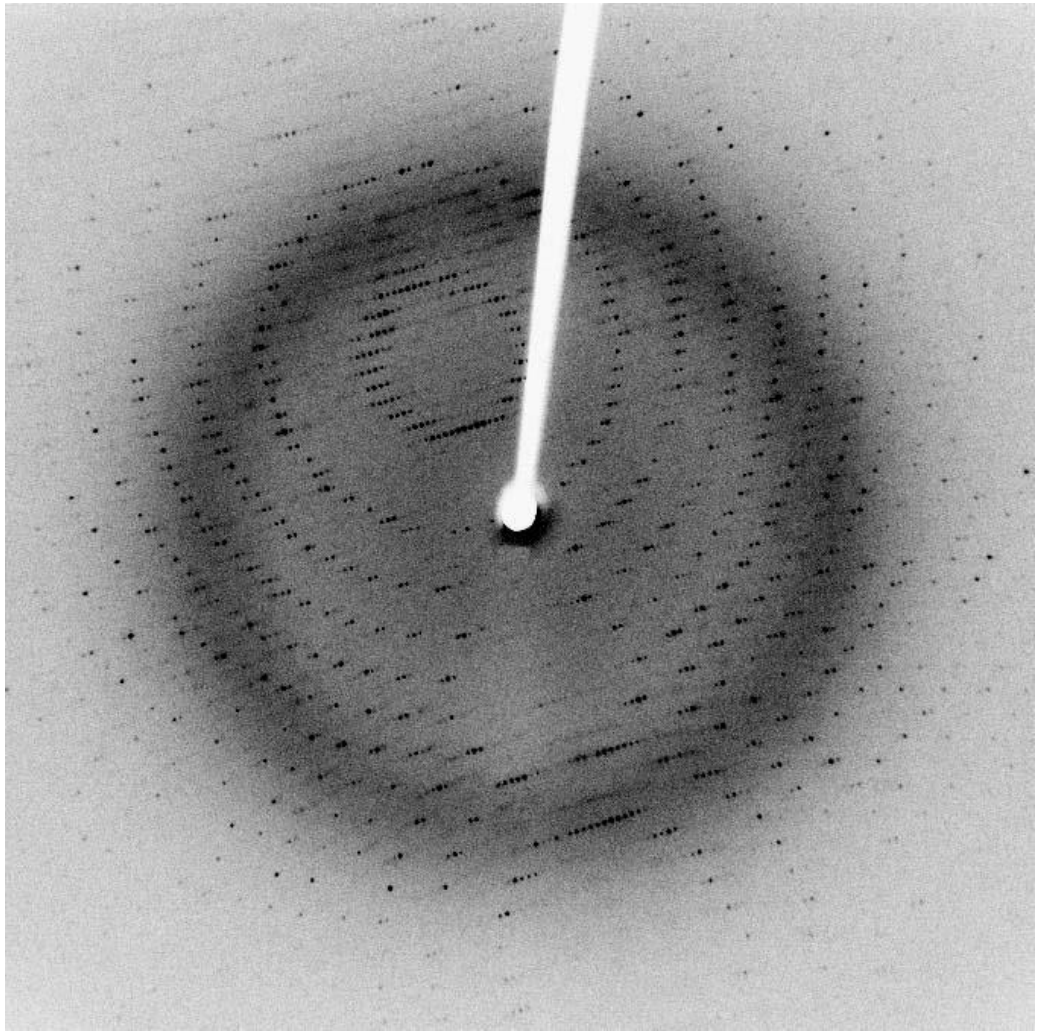


Lipoprotein (PDB: 1AV1)

D.W. Bohrani et al., *Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation*, PNAS, 1997

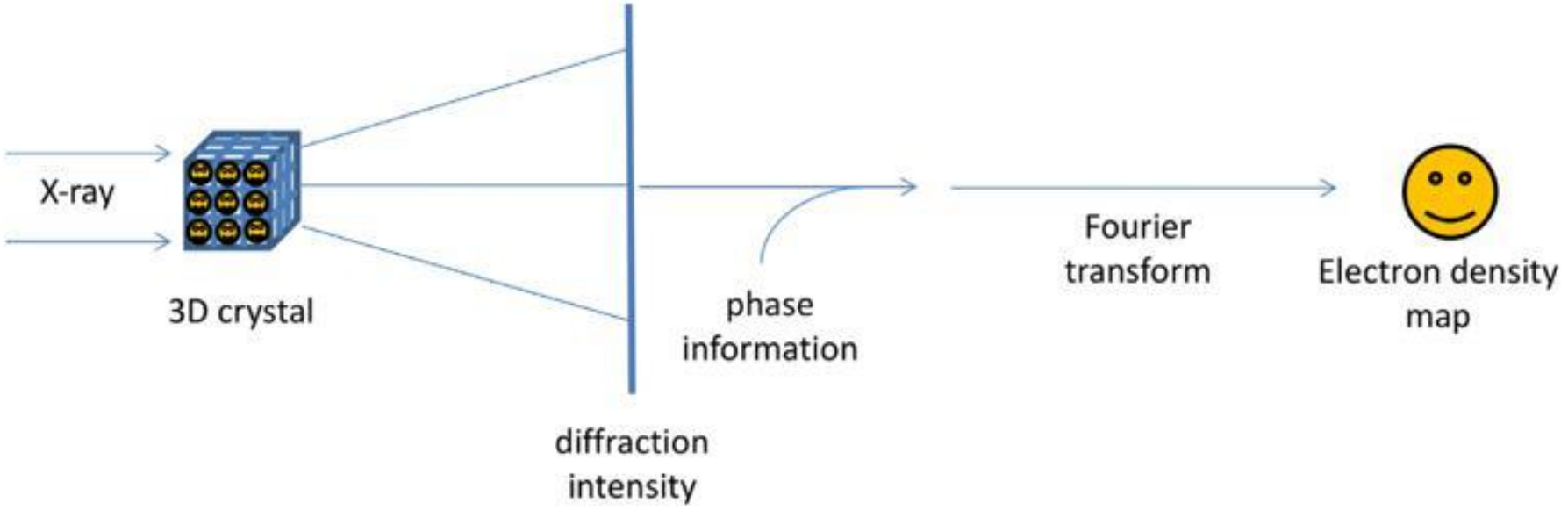
Structure determination: X-ray crystallography

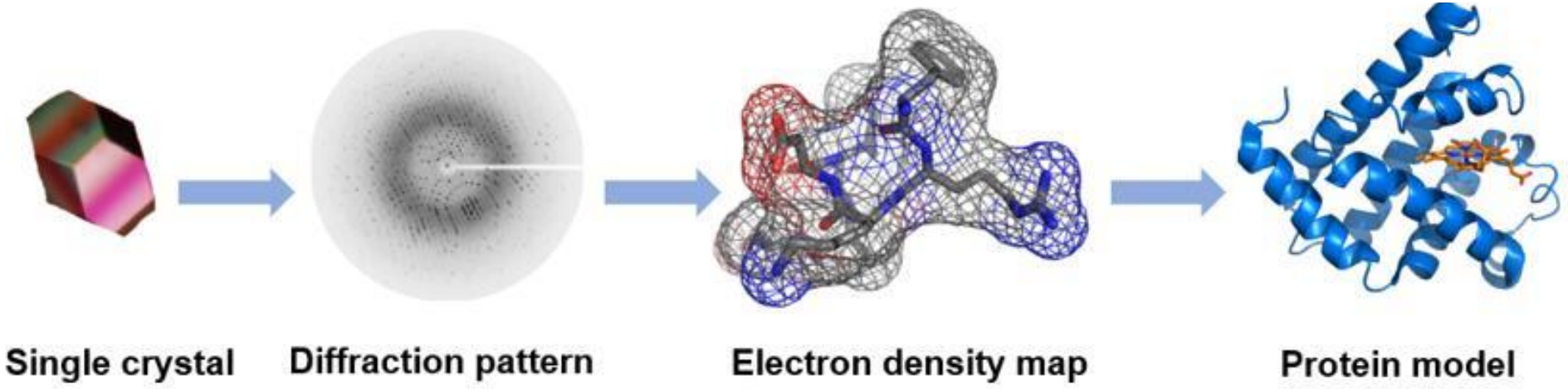




www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm

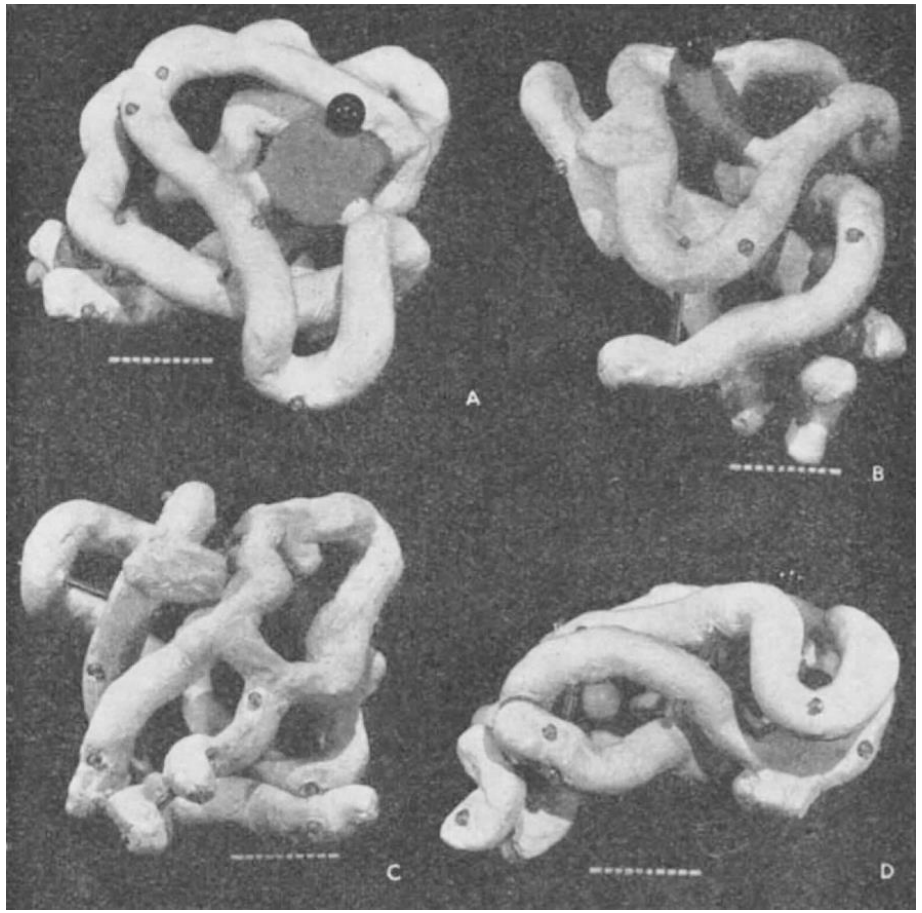
Structure determination: X-ray crystallography

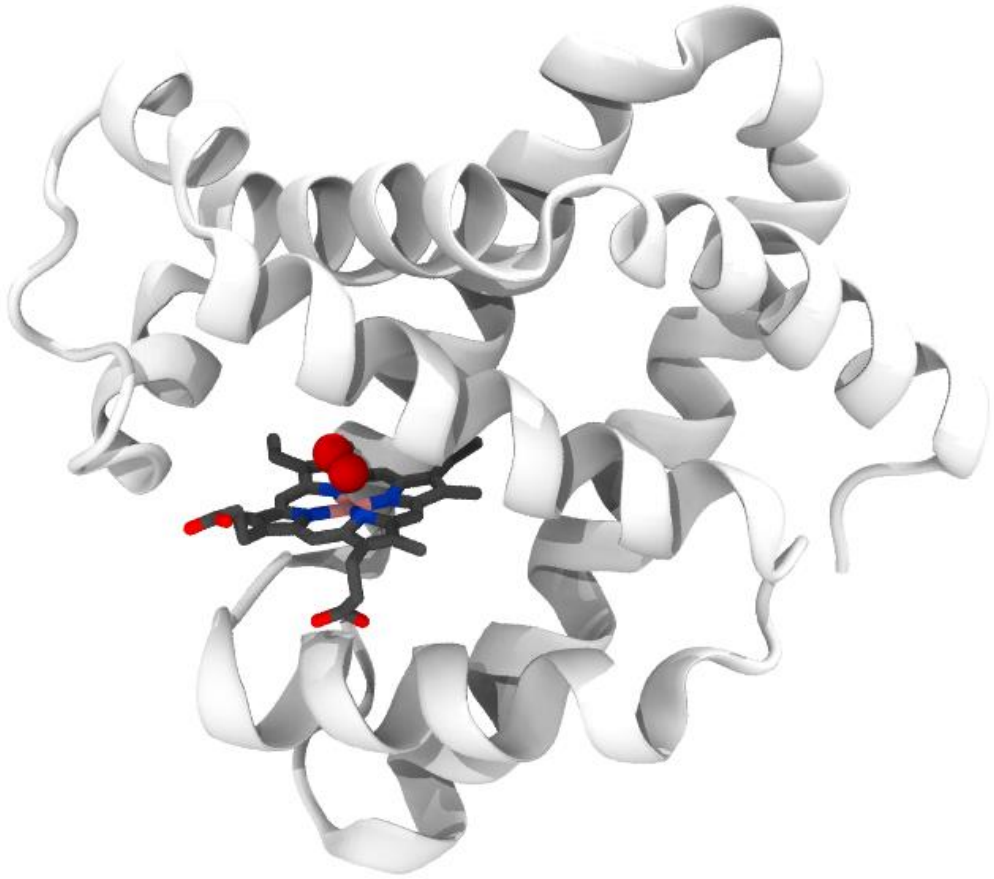




www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm

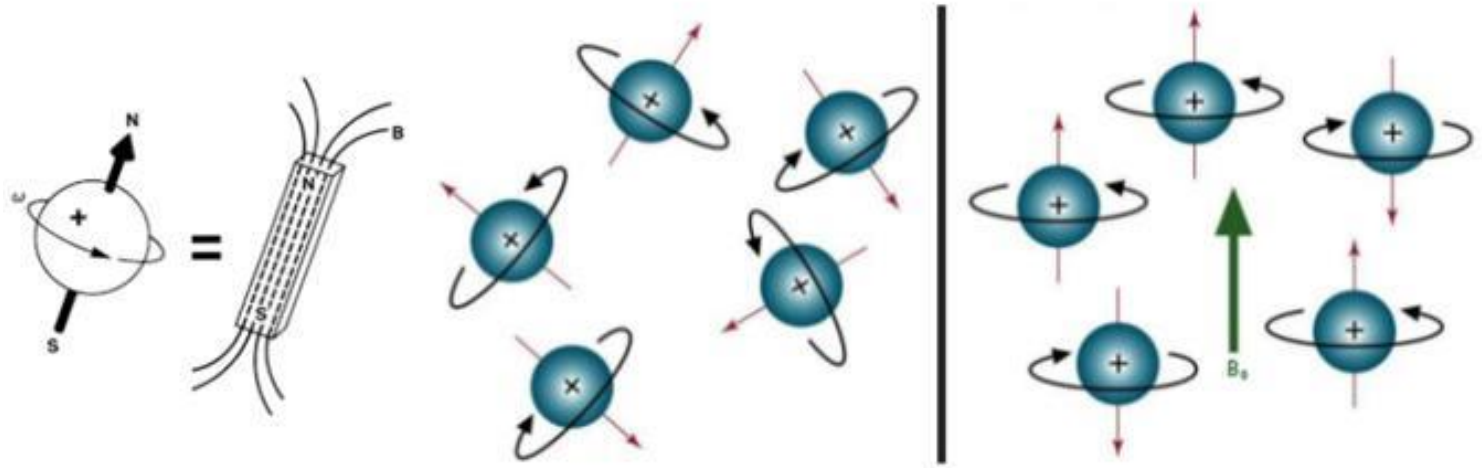
Structure determination: X-ray crystallography

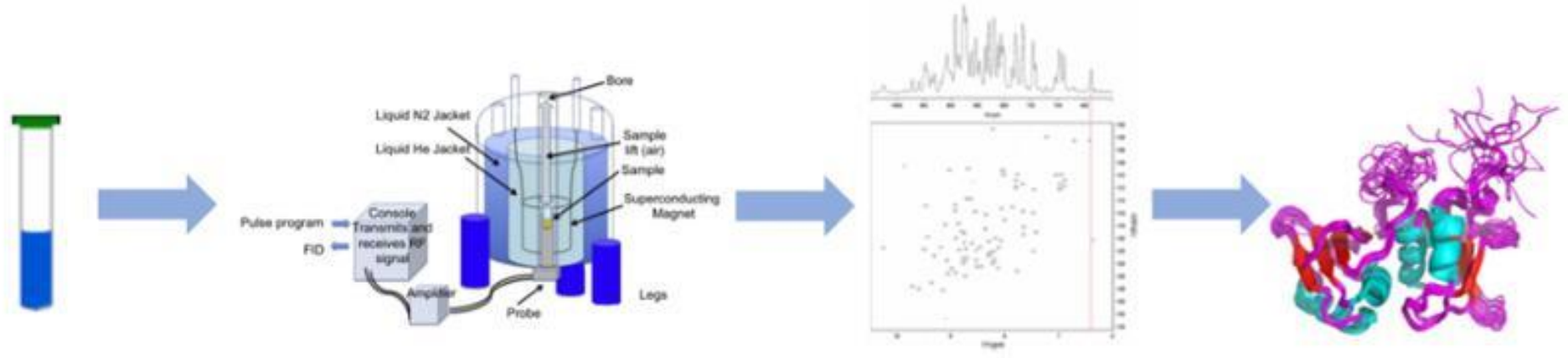




J.C. Kendrew et al., *A three-dimensional Model of the Myoglobin Molecule obtained by X-Ray Analysis,* Nature, 1958

Structure determination: Nuclear Magnetic Resonance (NMR)





www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm

Structure determination: Electron Microscopy (EM)

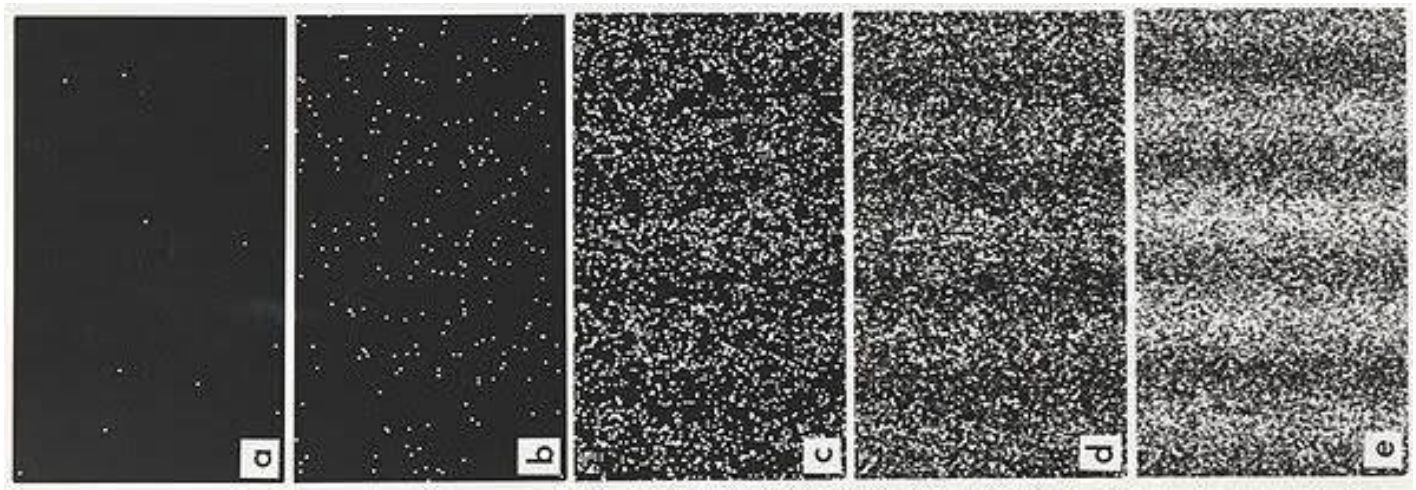
- Particles as «waves that transfers energy and momentum»

$$\lambda = \frac{h}{p}$$

λ : wavelength
p : momentum
h : Planck constant

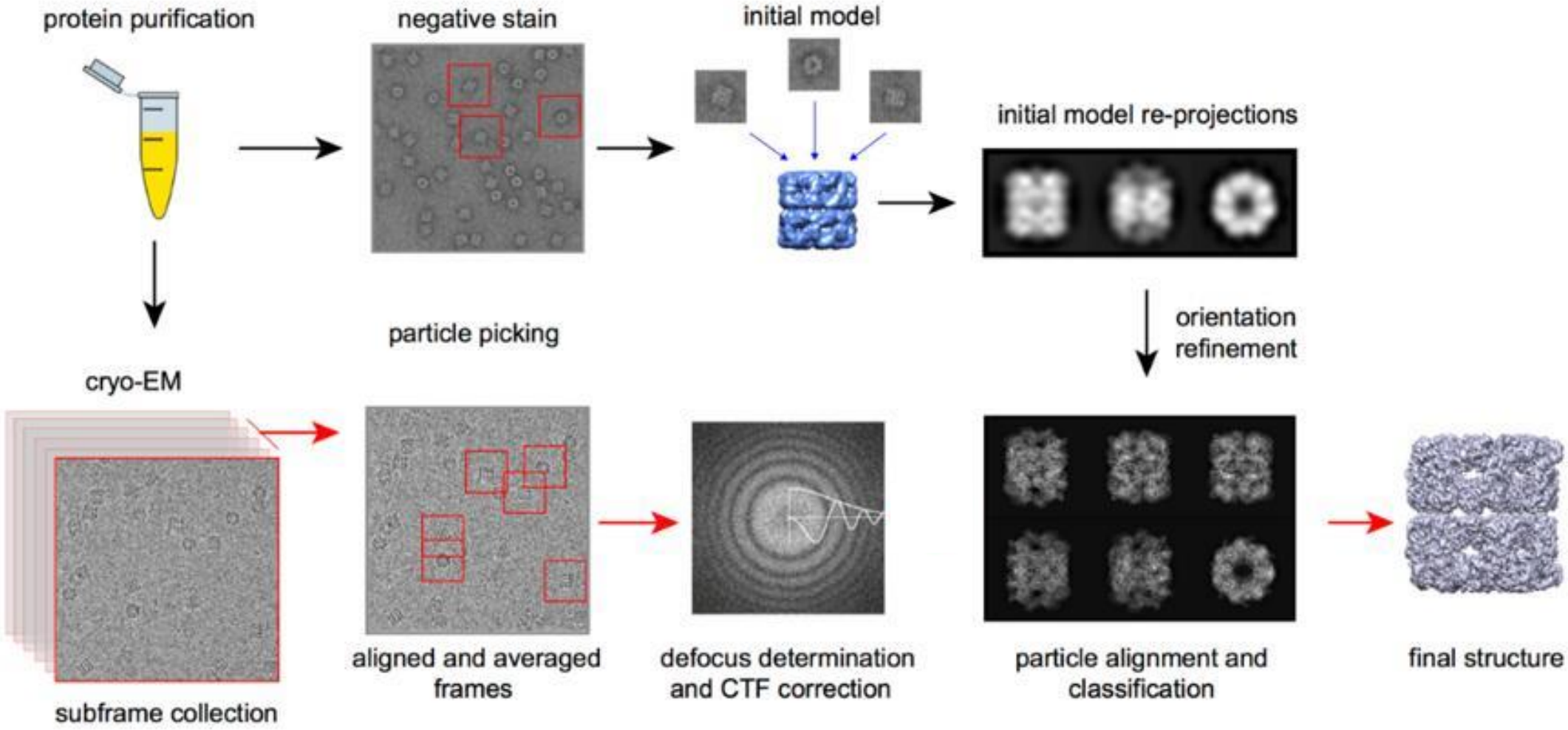
Louis De Broglie
1892-1987

- Davisson–Germer experiment: electrons diffract too!



A. Tonomura *et al.*, *Demonstration of single-electron buildup of an interference pattern*, American Journal of Physics, 1989

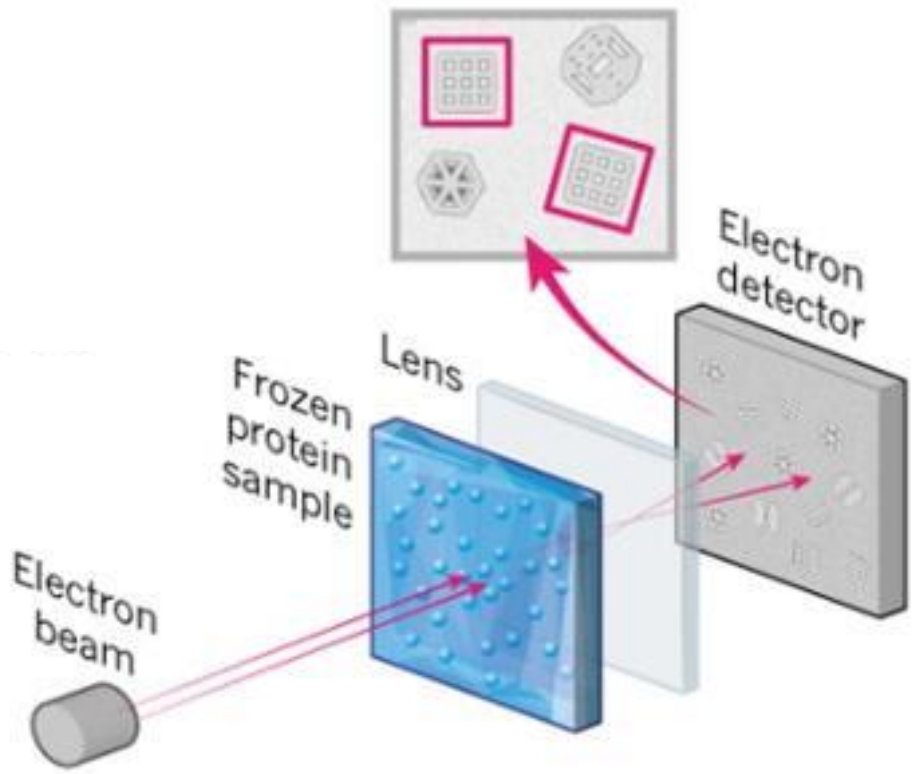
Structure determination: Electron Microscopy (EM)

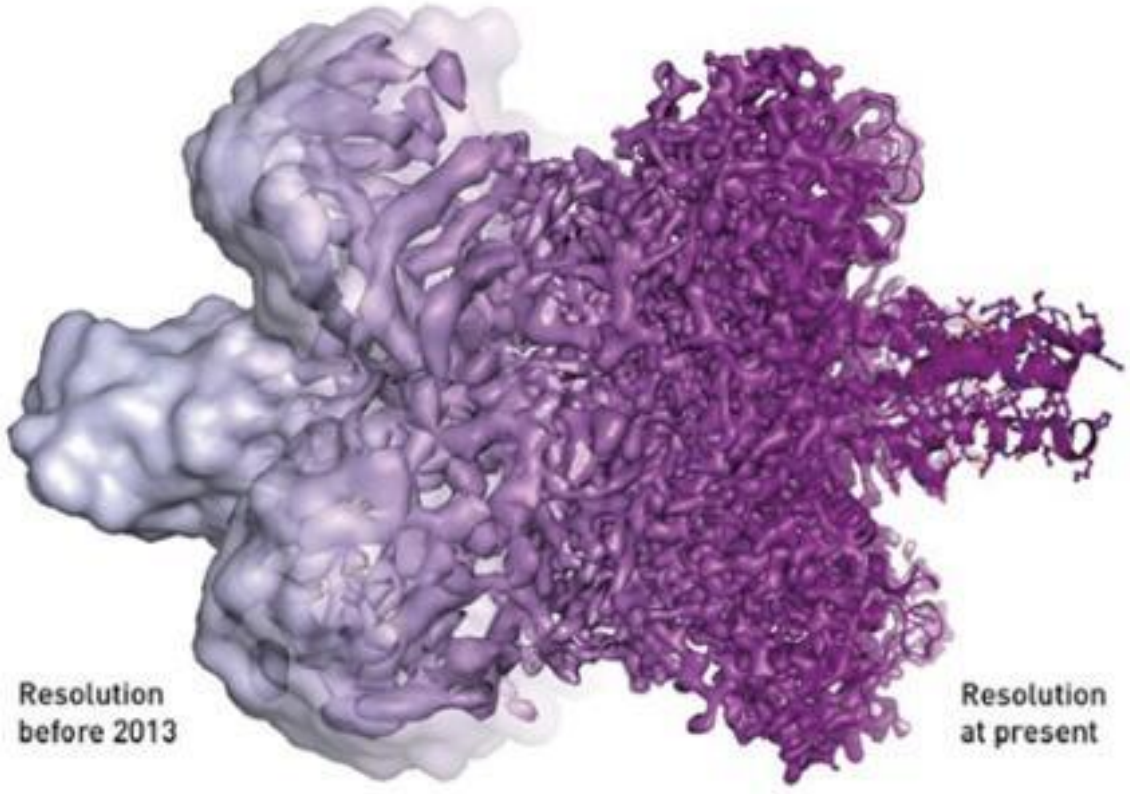


www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm

Structure determination: Electron Microscopy (EM)

Resolution Revolution





www.creative-biostructure.com/comparison-of-crystallography-nmr-and-em_6.htm

[Extra] experiments: pros and cons

X-ray crystallography

- High resolution
- Broad molecular weight range
- protein must crystallize
- crystal must diffract
- provides static information

NMR

- Structure studied in solution
- Need high sample concentration and purity
- Requires sophisticated data analysis
- Limited to small proteins (<50 kDa)

cryo-EM

- Easy sample preparation
- Structure in its native state
- Limited to large proteins (>100 kDa)
- High resolution is not guaranteed
- Equipment is very expensive

The Protein Data Bank (PDB)

- Molecular structures are deposited in the Protein Data Bank (PDB)
  - 1971: foundation of PDB at Brookhaven National Laboratory
  - 2003: wwPDB founded, now with four deposition centres

- Each molecule assigned a unique 4-characters code (e.g. 1MBO, 1AV1, …)

H.Berman, K.Henrick and H. Nakamura, *Announcing the worldwide Protein Data Bank.* Nature Structural & Molecular Biology, 2003

wwPDB consortium , *Protein Data Bank: the single global archive for 3D macromolecular structure data*, Nucleic Acids Research, 2019

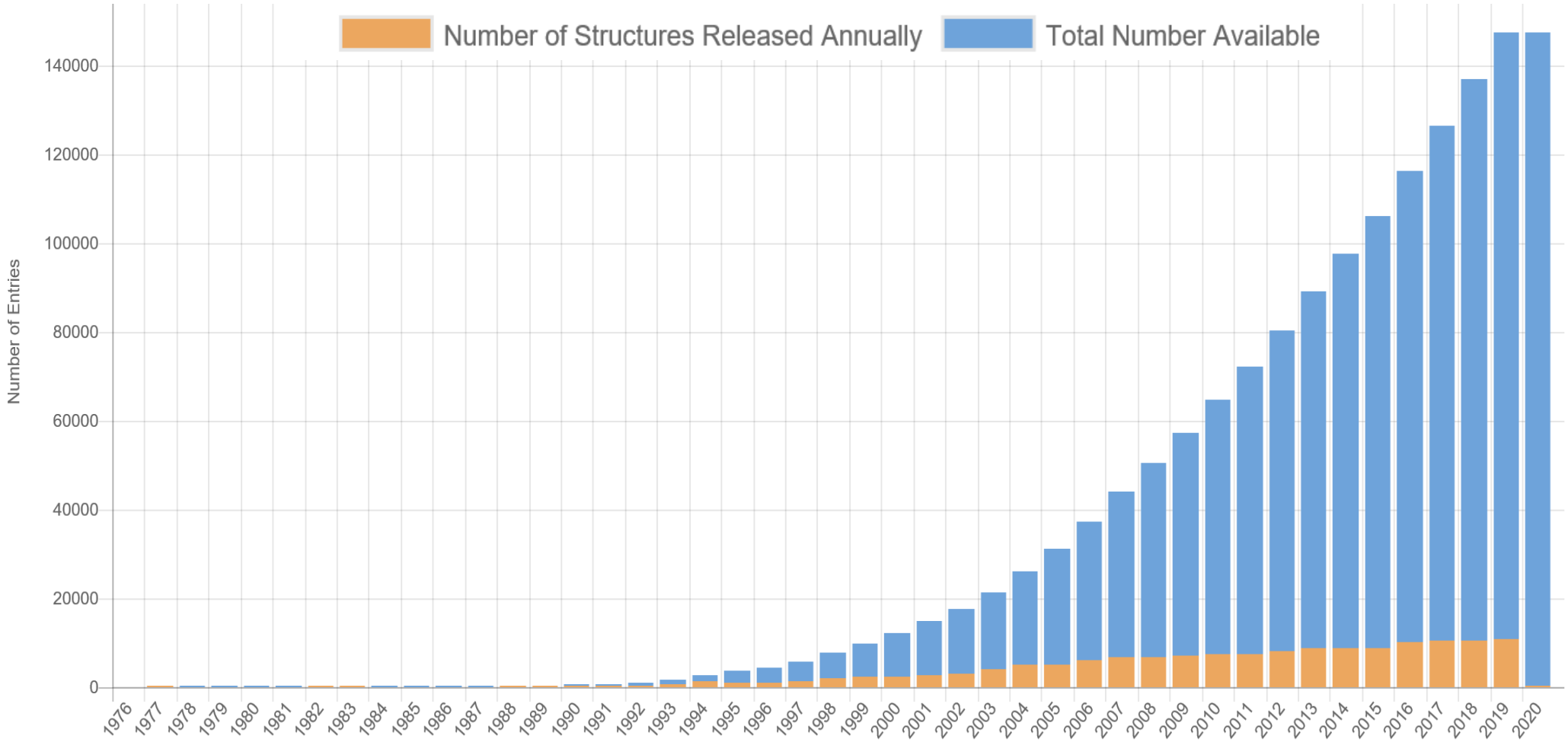
PDB files – storing molecular data

```
ATOM      1  N   GLU A   2       5.955  77.192  41.900  1.00 45.66           N
ATOM      2  CA  GLU A   2       5.061  76.430  40.975  1.00 45.66           C
ATOM      3  C   GLU A   2       5.843  75.299  40.293  1.00 44.55           C
ATOM      4  O   GLU A   2       7.062  75.489  40.068  1.00 45.05           O
ATOM      5  CB  GLU A   2       4.405  77.378  39.971  1.00 46.52           C
ATOM      6  CG  GLU A   2       3.572  78.477  40.600  1.00 47.63           C
ATOM      7  CD  GLU A   2       4.276  79.825  40.670  1.00 48.40           C
ATOM      8  OE1 GLU A   2       5.057  80.088  41.613  1.00 48.51           O
ATOM      9  OE2 GLU A   2       4.017  80.669  39.761  1.00 48.90           O
ATOM     10  N   PRO A   3       5.159  74.174  40.008  1.00 42.59           N
ATOM     11  CA  PRO A   3       5.766  73.007  39.378  1.00 40.68           C
ATOM     12  C   PRO A   3       6.196  73.159  37.927  1.00 38.44           C
ATOM     13  O   PRO A   3       5.432  73.752  37.156  1.00 38.77           O
ATOM     14  CB  PRO A   3       4.727  71.885  39.479  1.00 40.84           C
ATOM     15  CG  PRO A   3       3.619  72.468  40.290  1.00 41.61           C
ATOM     16  CD  PRO A   3       3.722  73.967  40.321  1.00 41.92           C
ATOM     17  N   VAL A   4       7.389  72.692  37.542  1.00 35.36           N
ATOM     18  CA  VAL A   4       7.857  72.786  36.159  1.00 32.16           C
ATOM     19  C   VAL A   4       7.558  71.409  35.549  1.00 30.87           C
ATOM     20  O   VAL A   4       7.869  70.408  36.244  1.00 31.57           O
ATOM     21  CB  VAL A   4       9.346  73.071  35.972  1.00 31.22           C
ATOM     22  CG1 VAL A   4       9.757  73.005  34.520  1.00 30.94           C
ATOM     23  CG2 VAL A   4       9.696  74.481  36.366  1.00 31.15           C
ATOM     24  N   TYR A   5       6.992  71.318  34.349  1.00 28.03           N
ATOM     25  CA  TYR A   5       6.736  69.993  33.805  1.00 25.19           C
```

ATOM ID name resname chain resid x y z occupancy β-factor atomtype

www.wwpdb.org/documentation/file-format

The Protein Data Bank (PDB)



[Extra] Protein fold classification: SCOP

Manual classification. New version (SCOPe) combines manual and automatic methods. Hierarchy:

1. Class: secondary structure composition (α, β, α/β, α+β, α/β)
2. Fold: compare secondary structure elements type, orientation and order
3. Superfamily: structural homology indicates common ancestor
4. Family: sequence identity indicates common ancestor
5. Domain: independent folding unit

scop.mrc-lmb.cam.ac.uk

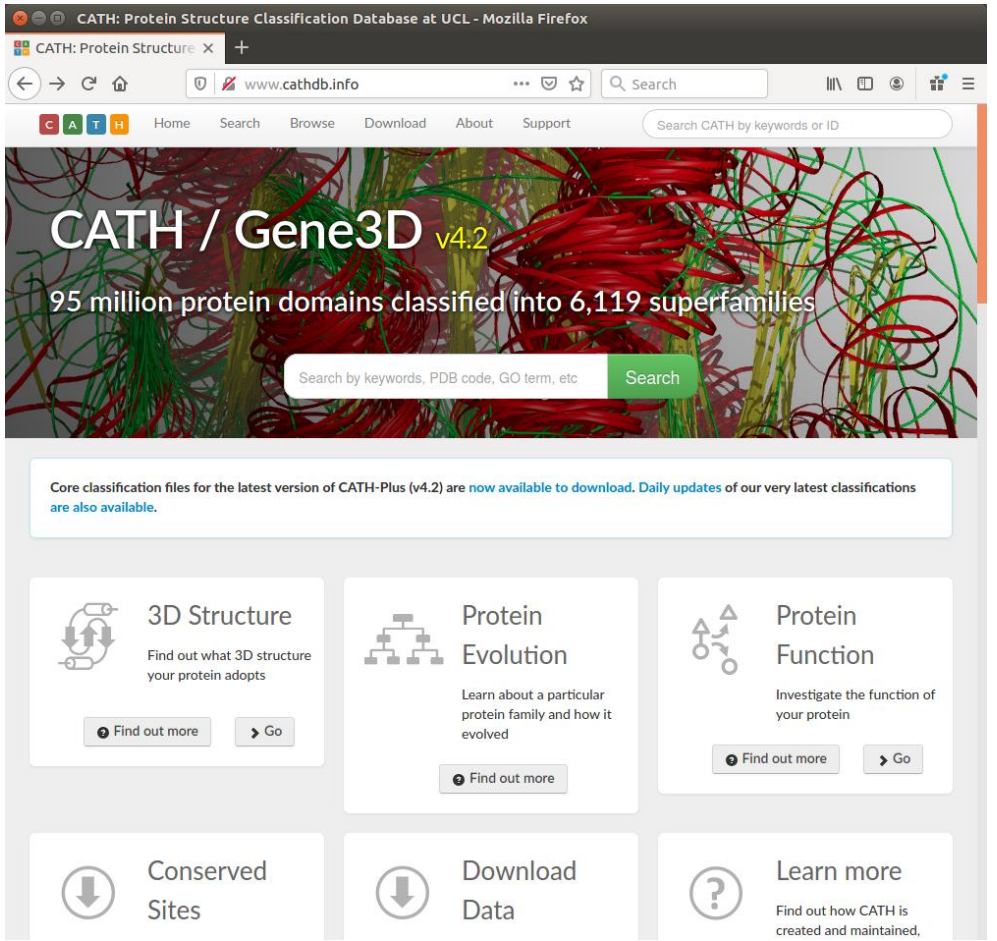
Murzin et al., *SCOP: a structural classification of proteins database for the investigation of sequences and structures, J.Mol.Biol., 1995*

J.M. Chandonia et al., *SCOPe: classification of large macromolcular structures in the structural classification of proteins-extended database,* Nucleic Acids Research, 2019

[Extra] Protein fold classification: CATH

Automatic classification of protein structures. Hierarchy:

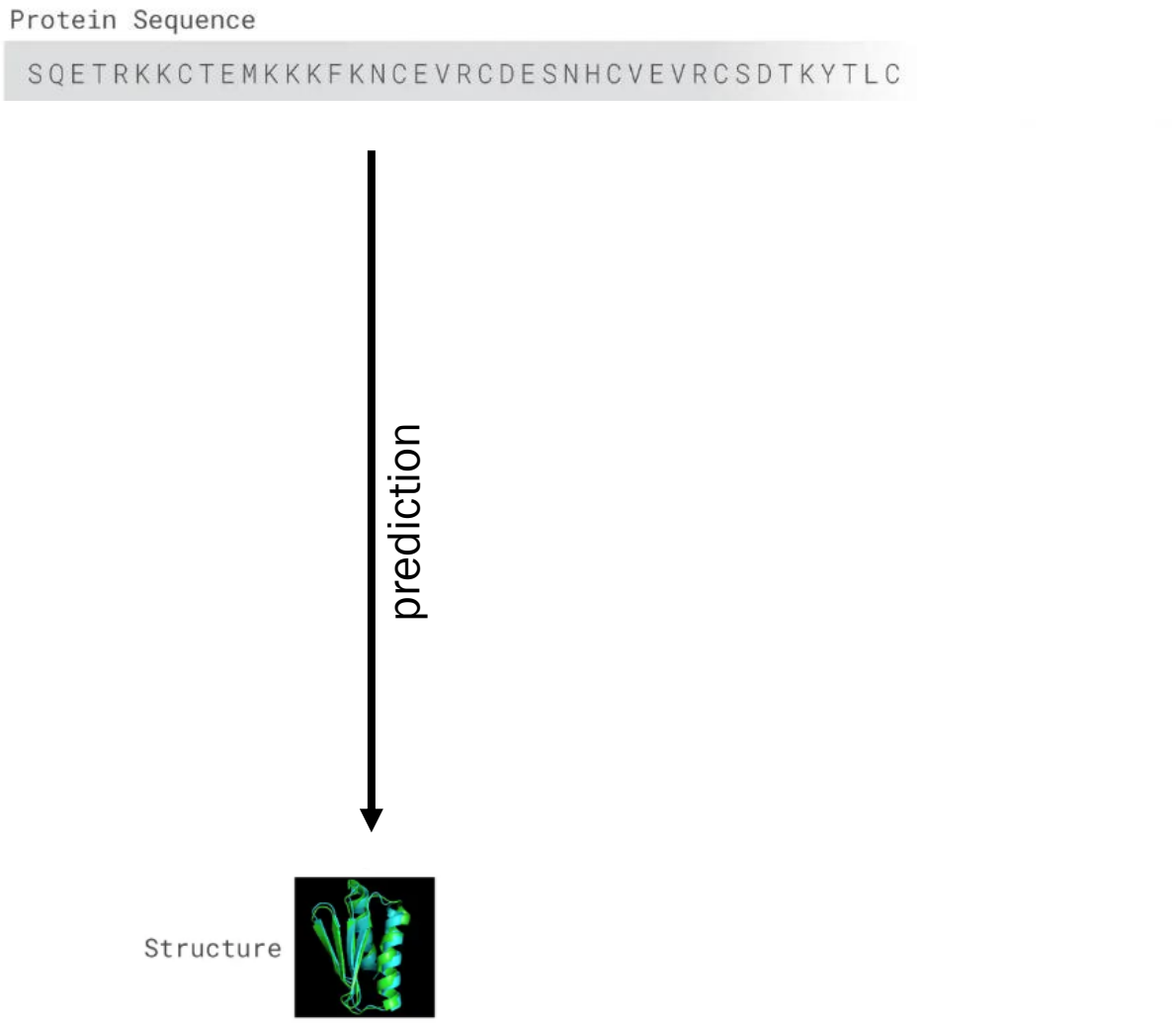
1. Class: secondary structure composition (α, β, α+β)
2. Architecture: compare secondary structure elements type, orientation and order
3. Topology: connectivity of secondary structure elements
4. Homologous Superfamily: structures shaing a common ancestor (using primary sequence)



www.cathdb.info

C. A. Orengo et al., *CATH – a hierarchic classification of protein domain structures*, Structure, 1997

Protein fold prediction



CASP, since 1994 biennial competition on protein fold prediction: predictioncenter.org

- *ab initio*
- template-based
- data-assisted
- contact prediction
- refinement

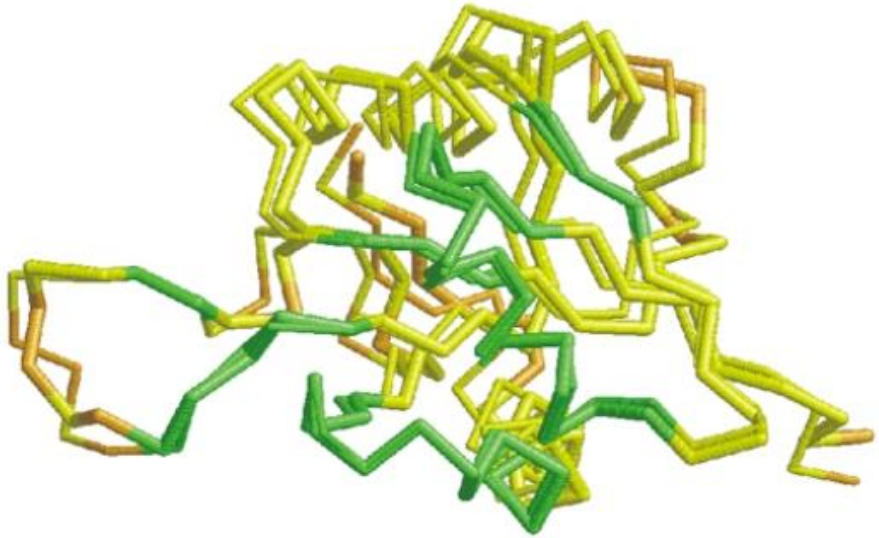
Protein fold prediction

Protein Sequence

SQETRKKCTEMKKKFKNCEVRCDESNHCVEVRCSDTKYTLC

prediction

Structure



- Quality assessment: RMSD is unsuitable
- Use global Distance Test total score (**GDS_TS**)
  - calculate largest subset of amino acids in model aligning to target with RMSD smaller than given cutoff.
  - Report average using 1, 2, 4 and 8 Å cutoffs

A. Zemla, *LGA: a method for finding 3D similarities in protein structures*, Nucleic Acids Research, 2003

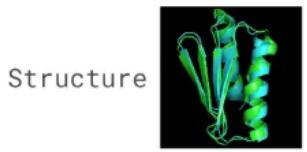
Protein fold prediction

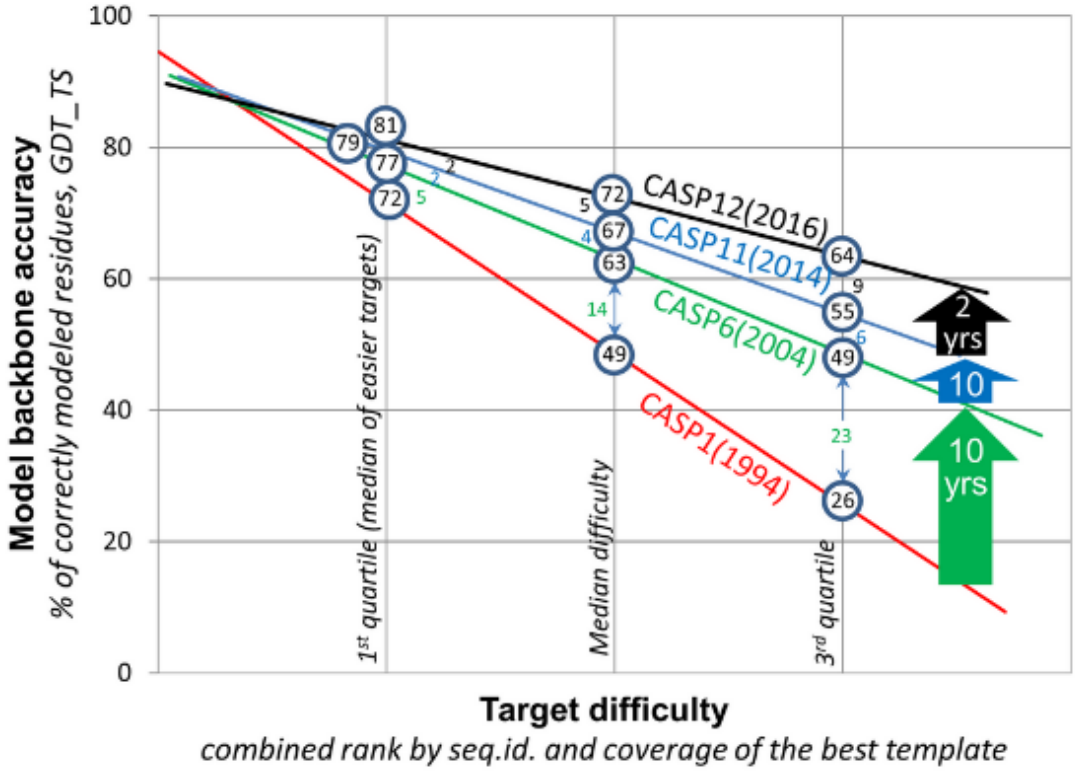
Protein Sequence

SQETRKKCTEMKKKFKNCEVRCDESNHCVEVRCSDTKYTLC

prediction

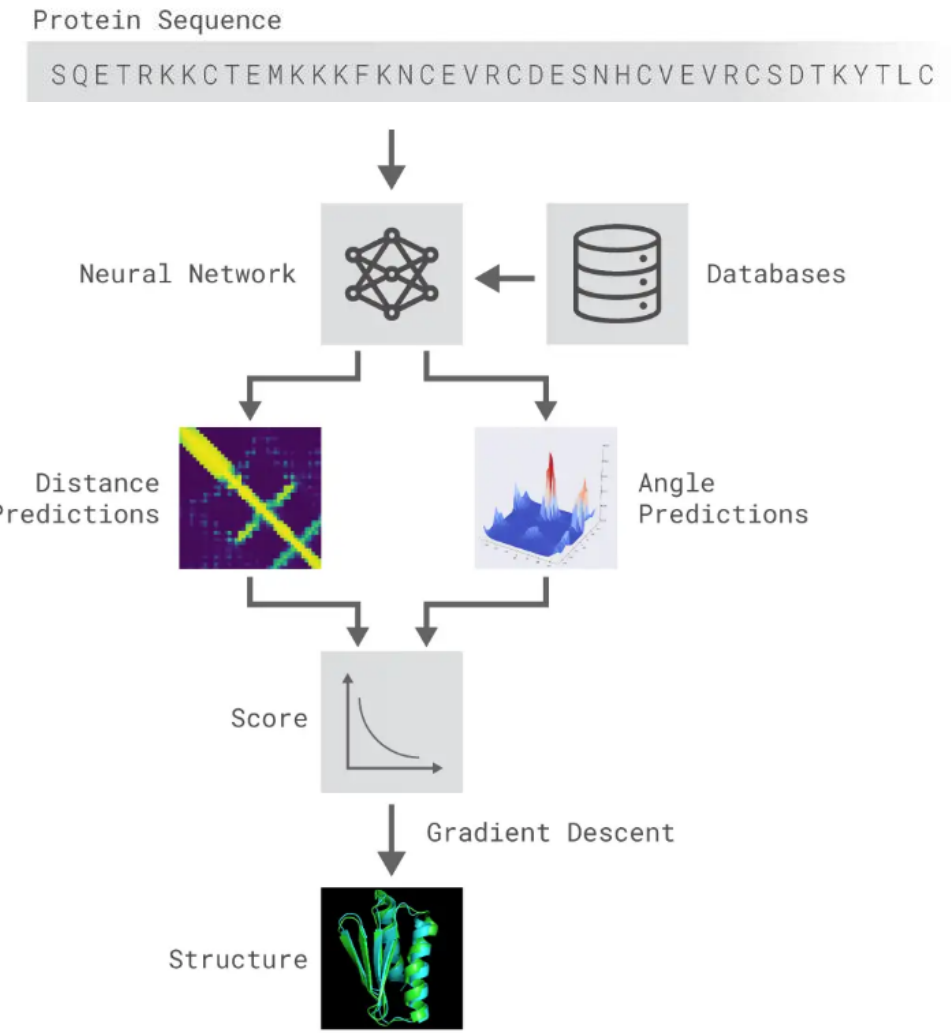
Structure





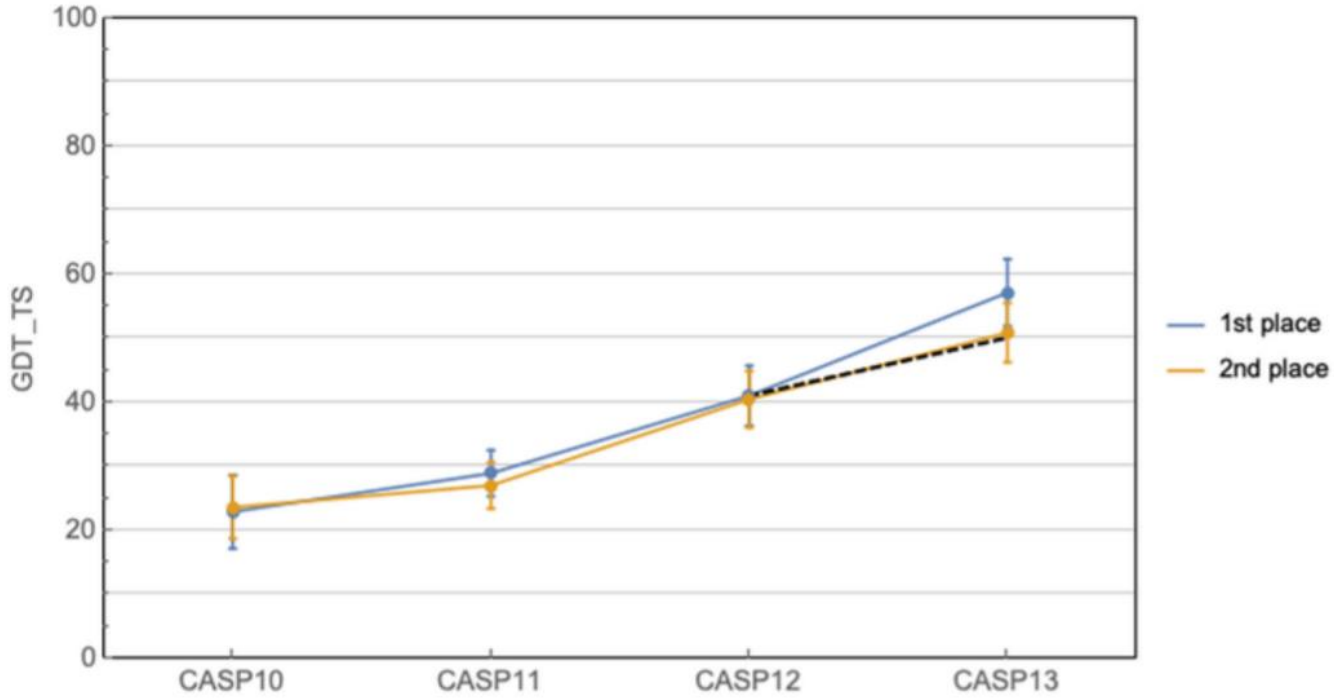
- Use global Distance Test total score (**GDS_TS**)
  - calculate largest subset of amino acids in model aligning to target with RMSD smaller than given cutoff.
  - Report average using 1, 2, 4 and 8 Å cutoffs

Protein fold prediction



A.W. Senior et al., *Improved protein structure prediction using potentials from deep learning*, Nature, 2020
deepmind.com/blog/article/alphafold

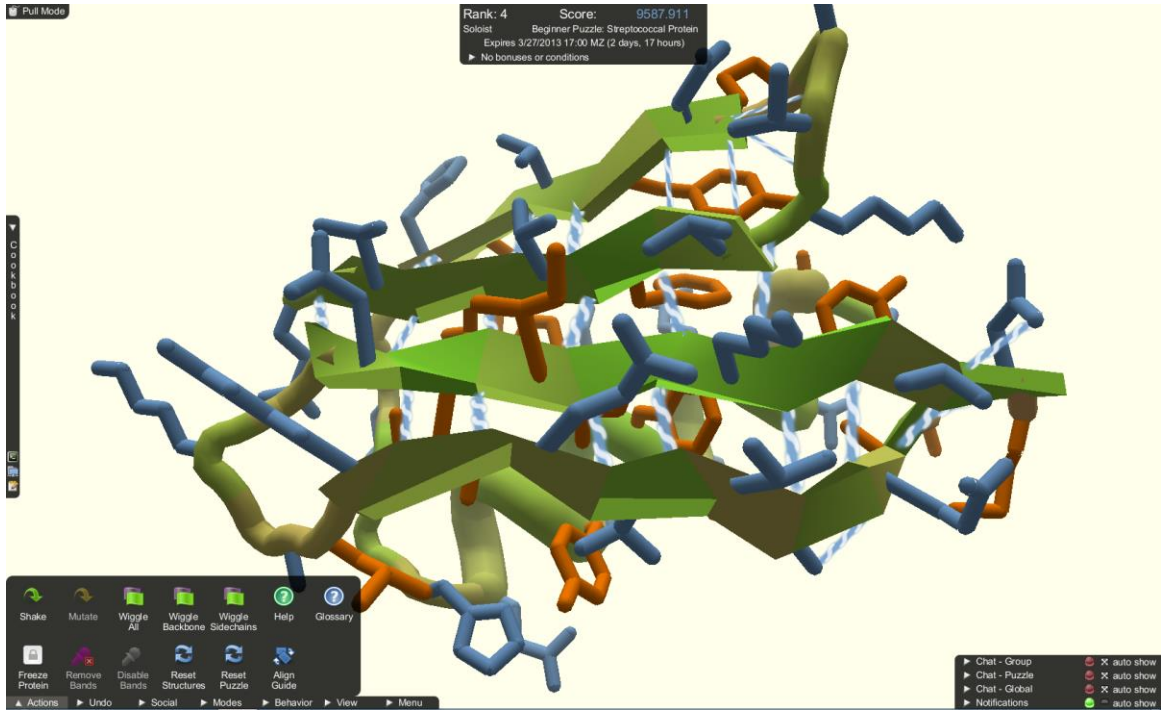
2018 (CASP13) *ab initio* competition won by *AlphaFold*, from Google's DeepMind (best in 25/43 cases)

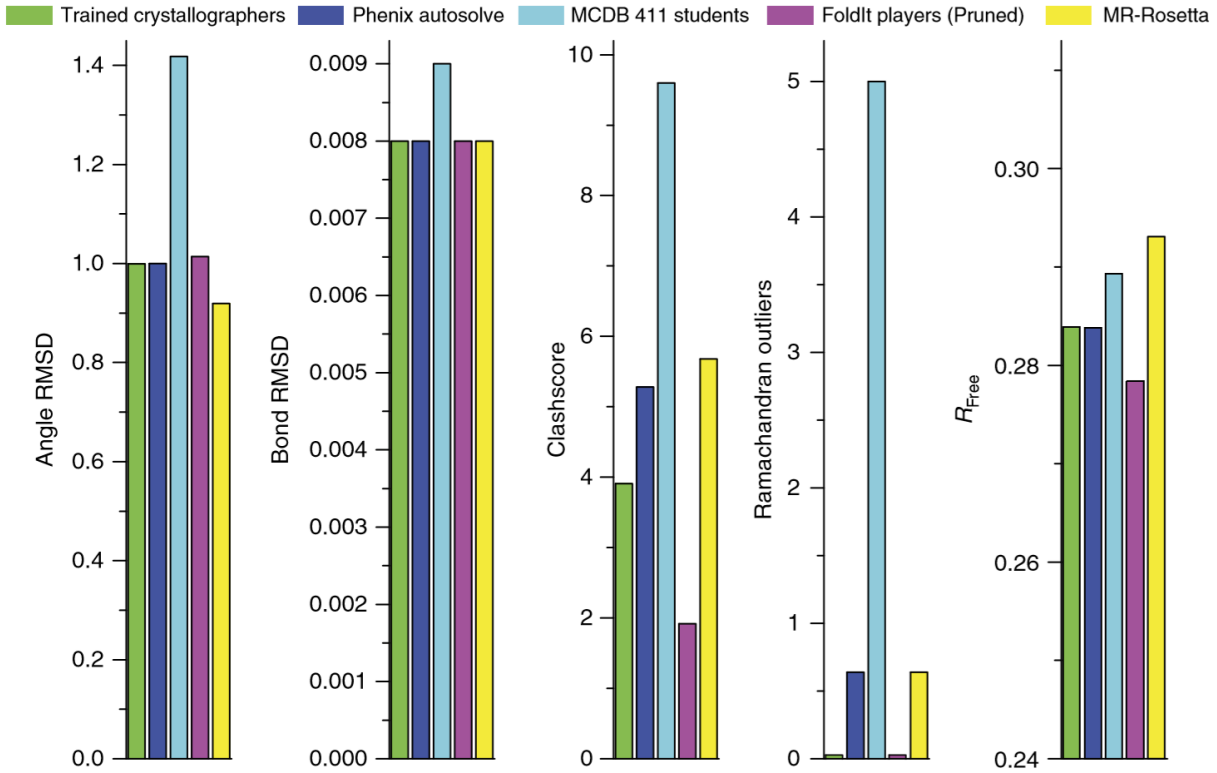


M. AlQuraishi, *AlphaFold at CASP13*, Bioinformatics, 2019
moalquraishi.wordpress.com/2018/12/09/alphafold-casp13-what-just-happened

Foldit

Protein folding videogame. Players can predict protein folds, and design new foldable proteins sequences.





fold.it

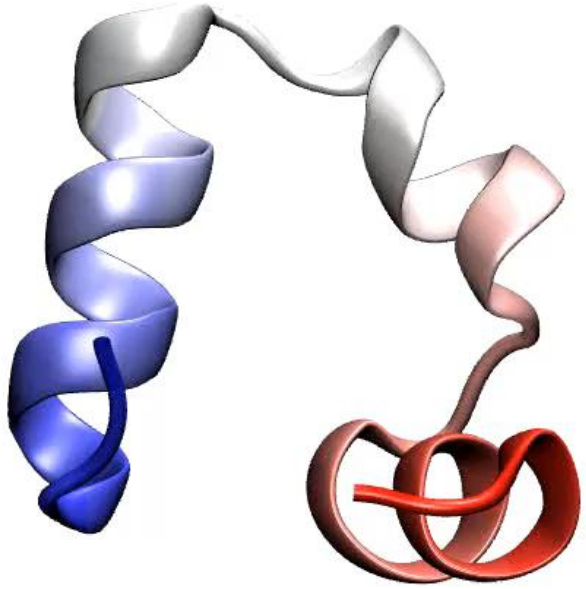
Koeptnick et al., *De novo protein design by citizen scientists*, Nature, 2019

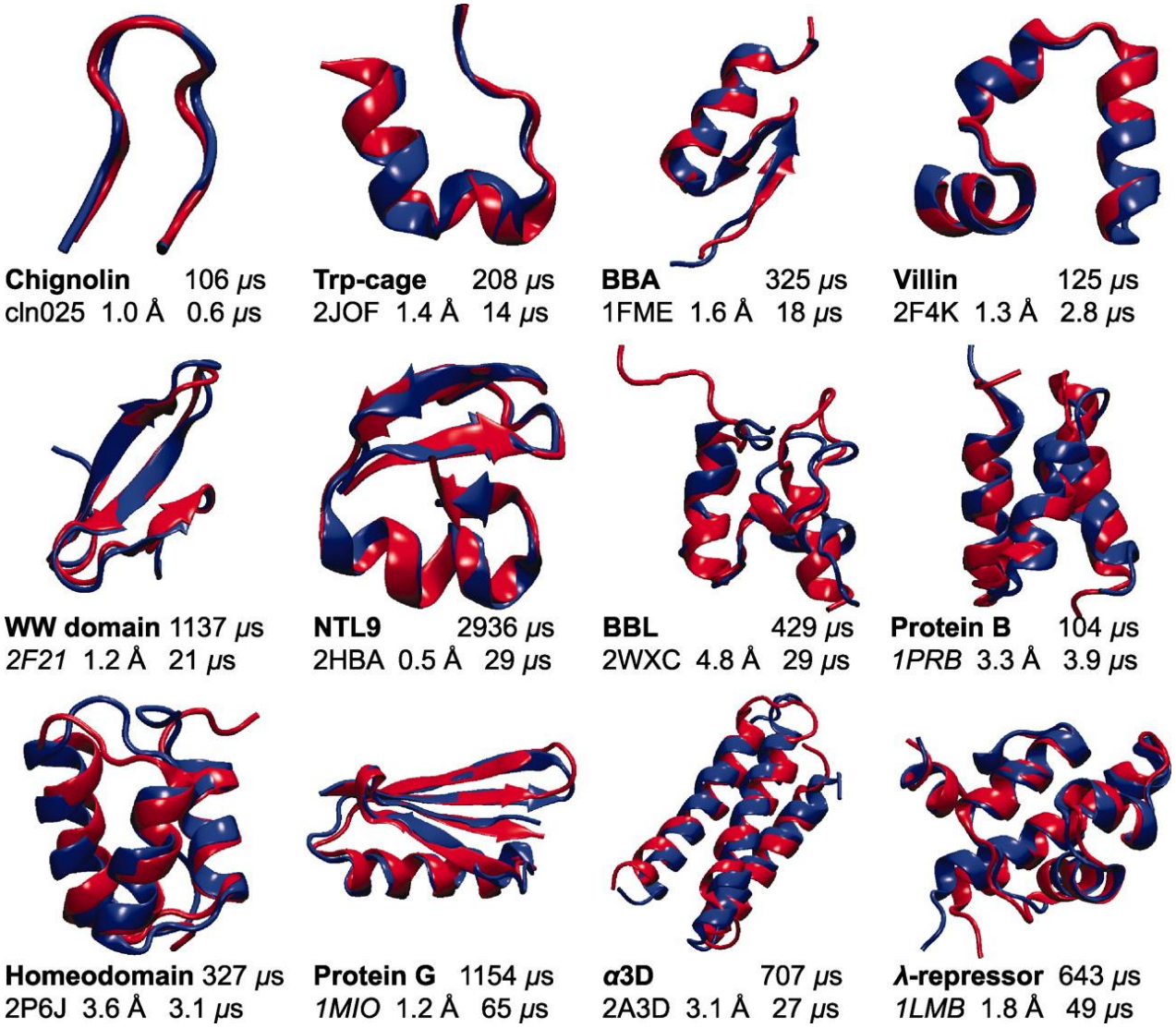
F. Khatib et AL., *Crystal Structure of a monomeric retroviral protease solved by protein folding game players*, Nature Struct.& Mol. Biol., 2011

S. Cooper et al., *The challenge of designing scientific discovery games*, FDG'10: Proceedings of the Fifth International Conferece on the Foundation of Digital Games, 2010

Watching proteins fold: simulation

- Following experimentally the *folding pathway* of a protein is difficult
- Folding of small fast-folding (<100 µs) proteins can be studied via simulation



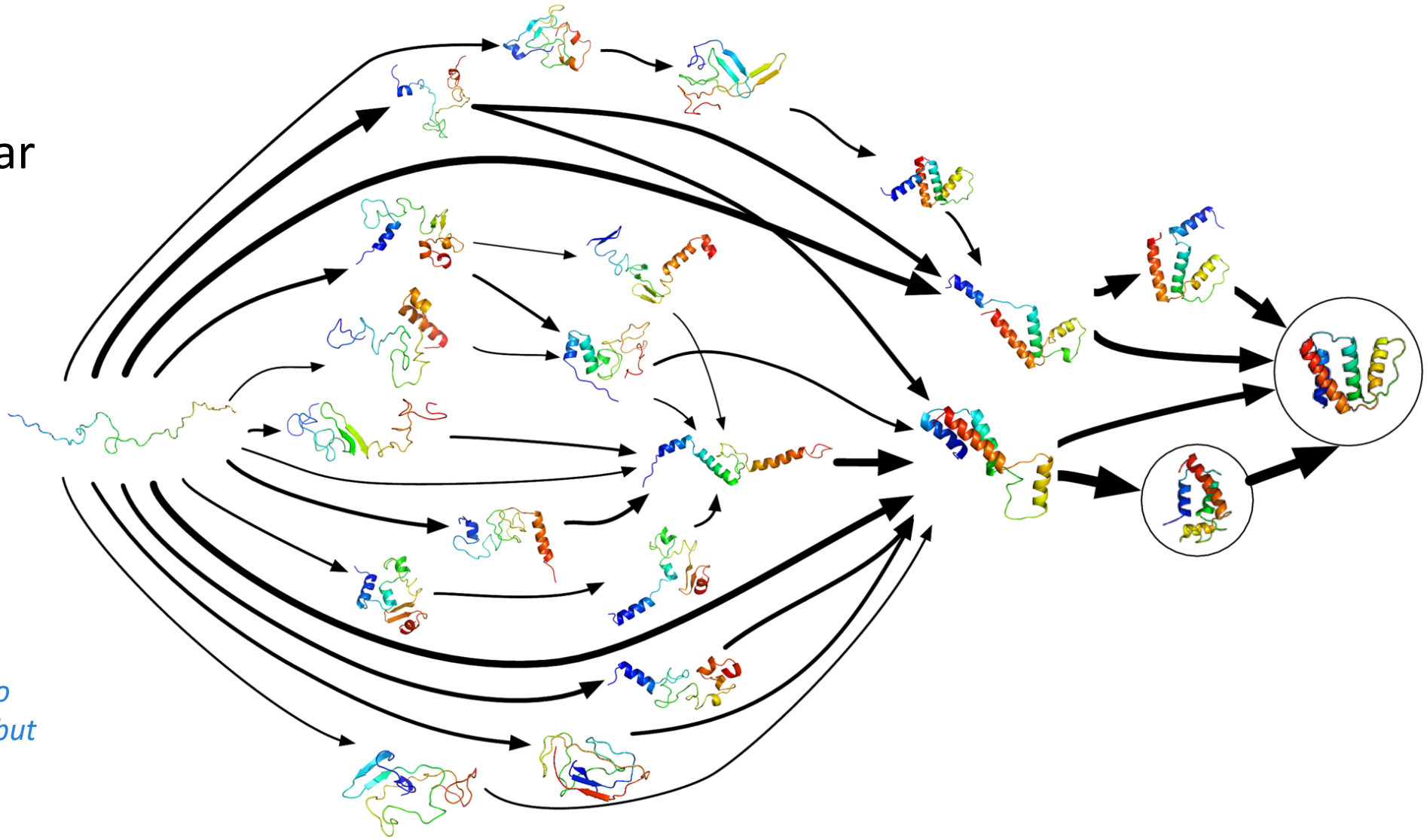


P. Freddolino and K. Schulten. *Common structural transitions in explicit-solvent simulations of villin headpiece folding*. Biophysical J., 2009

K. Lindorff-Larsen et al. *How fast-folding proteins fold*. Science, 2011

Folding@Home

Combine distribute computing, Molecular simulation and Markov State modelling to predict protein folding pathways.

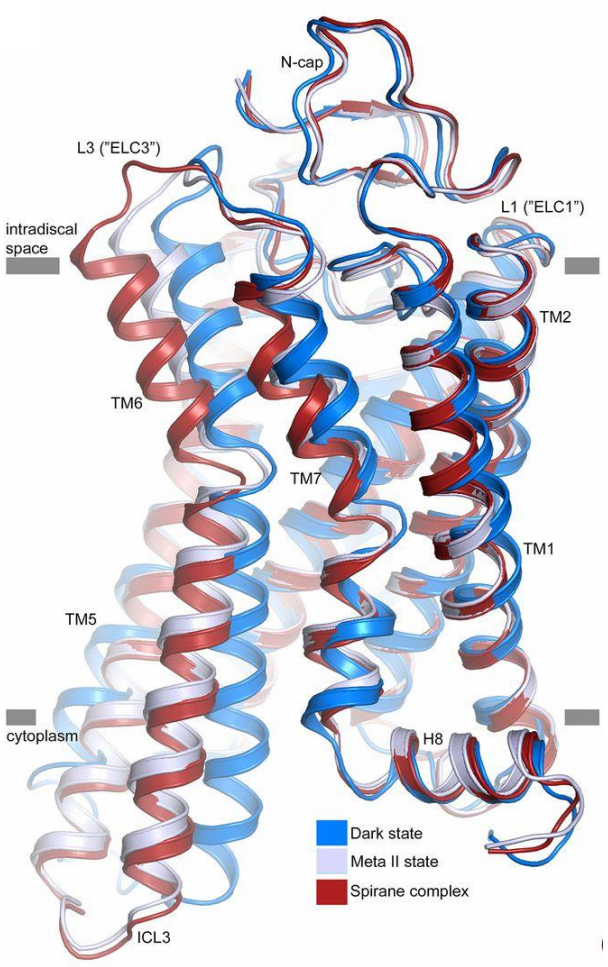


www.foldingathome.org

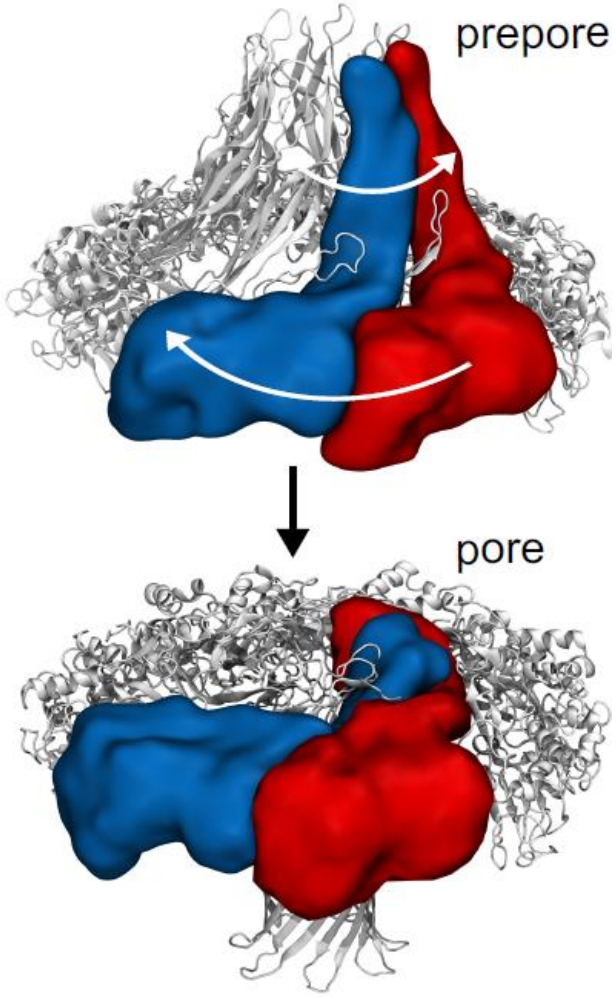
V. S. Pande, K. Beauchamp, G. R. Bowman, *Everything you wanted to know about Markov State Models but were afraid to ask*. *Methods*, 2010

Protein dynamics

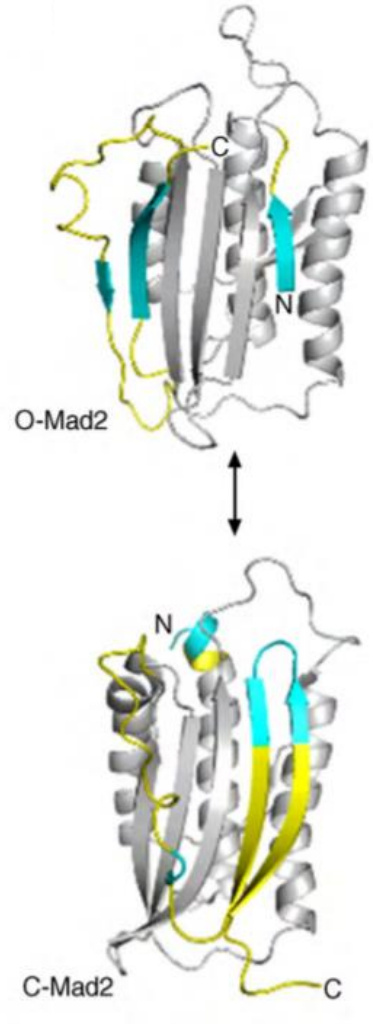
sequence + interaction with environment = conformational space



D. Mattle et al., *Ligand channel in pharmacologically stabilized rhodopsin*, PNAS, 2018



M.T. Degiacomi et al., *Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism*, Nat. Chem. Biol., 2013

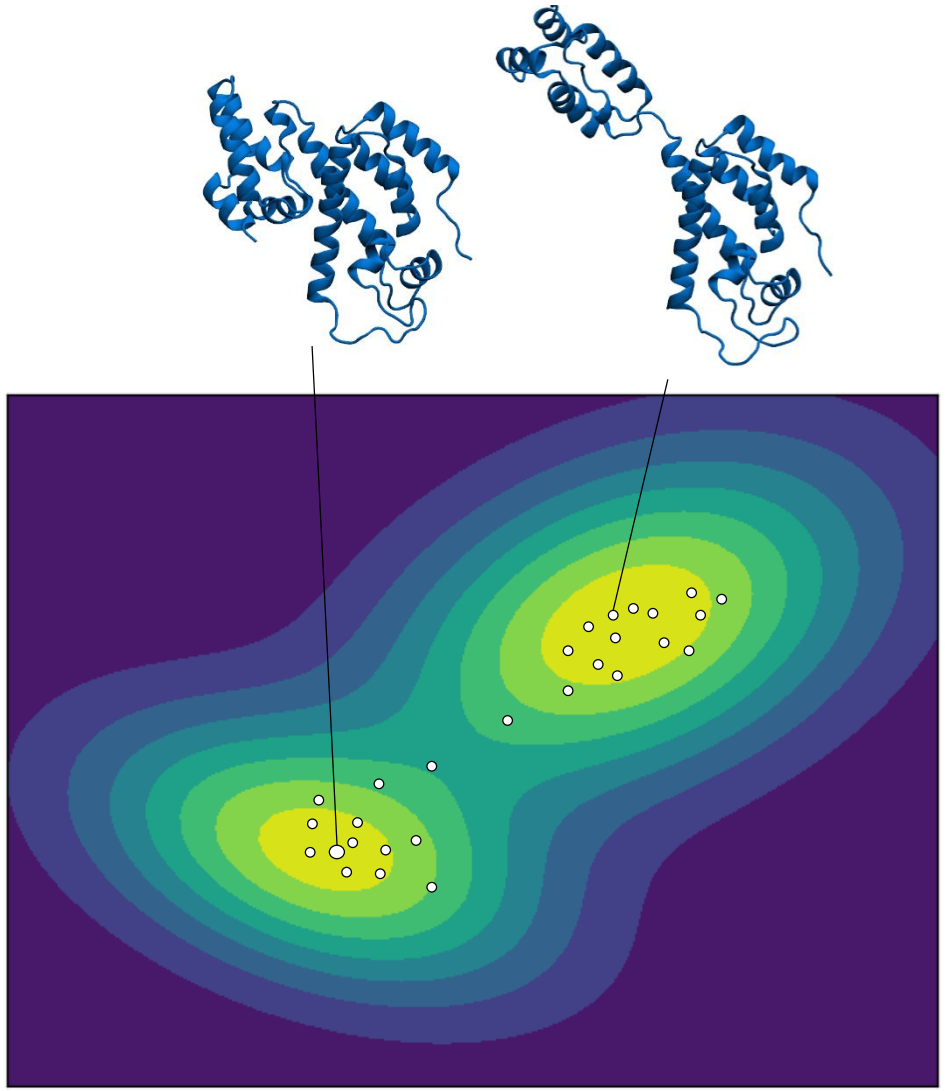


P.N. Bryan and J. Orban, *Proteins that switch folds*, Curr. Op. Struct. Biol., 2010

Protein conformational space

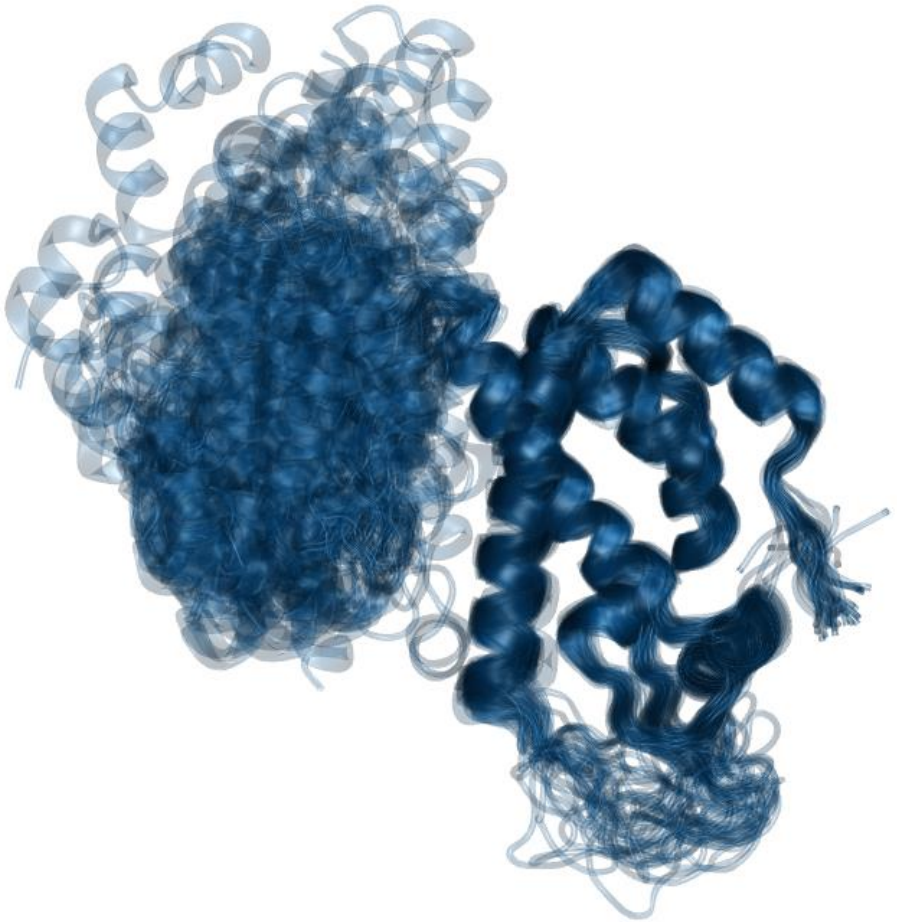
Energetically favourable conformations are explored more often than unfavourable ones

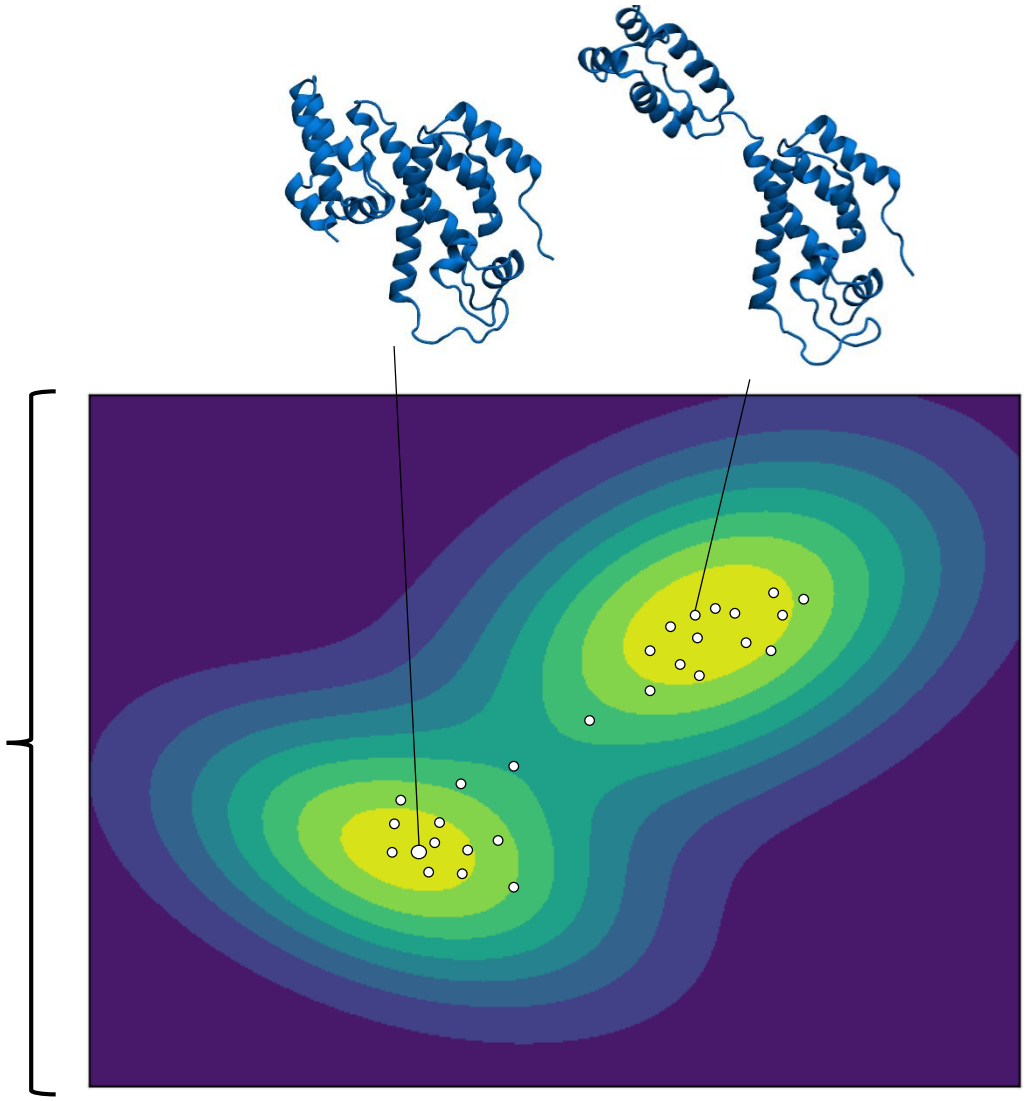
The higher the energy barrier between two states, the longer the time needed to observe a transition



Protein conformational space

$$p_i \propto e^{-\varepsilon_i / kT}$$

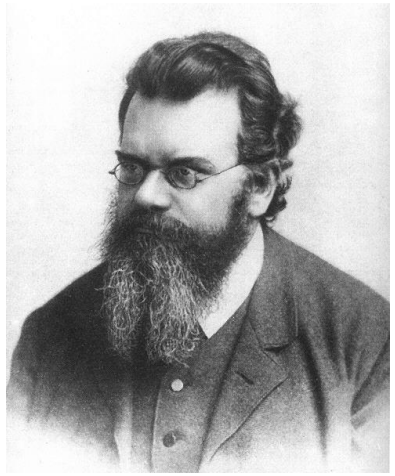




Protein conformational space

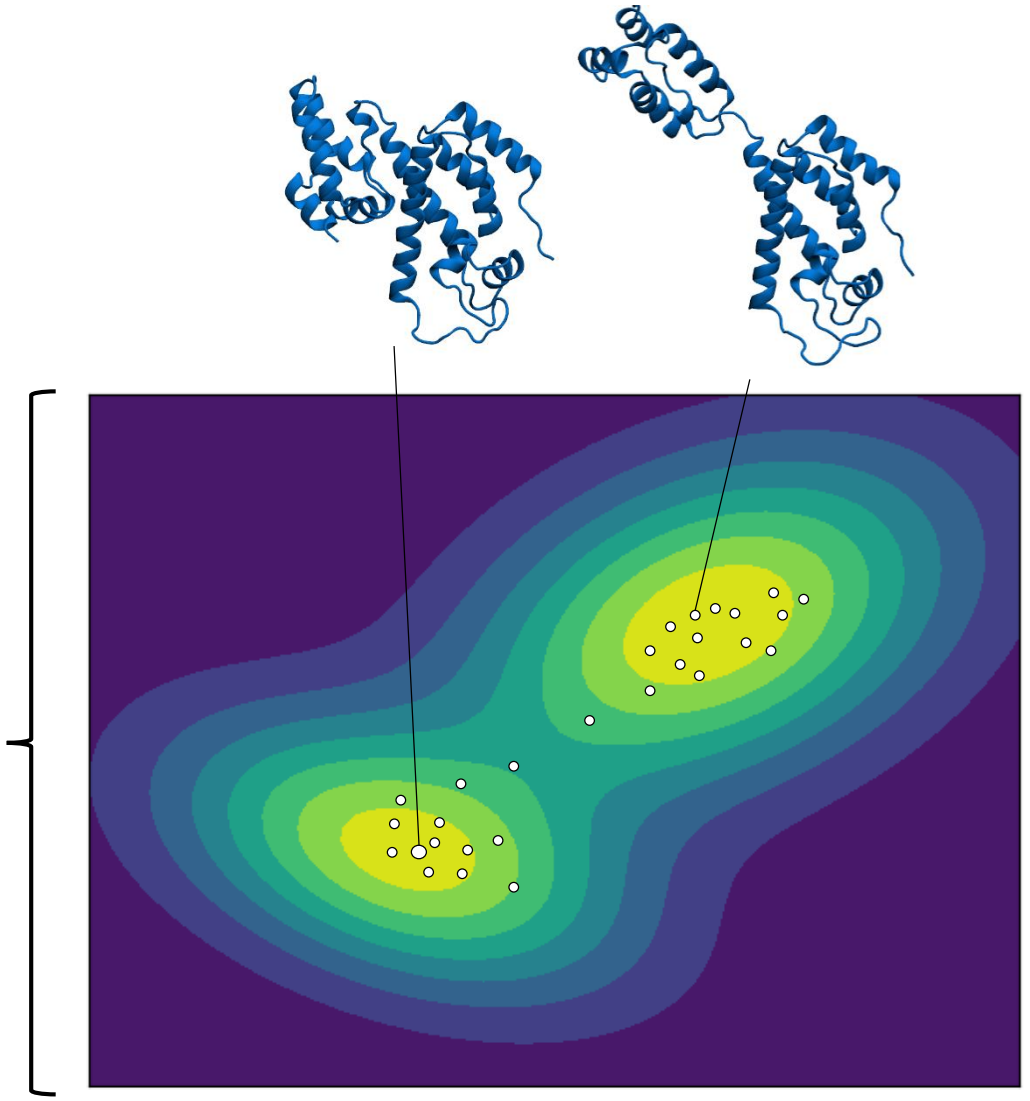
$$p_i \propto e^{-\varepsilon_i / kT}$$

The Boltzmann distribution



$$p_i = \frac{1}{Q} e^{-\varepsilon_i / kT}$$

$$Q = \sum_{i=1}^{M} e^{-\varepsilon_i / k_B T}$$

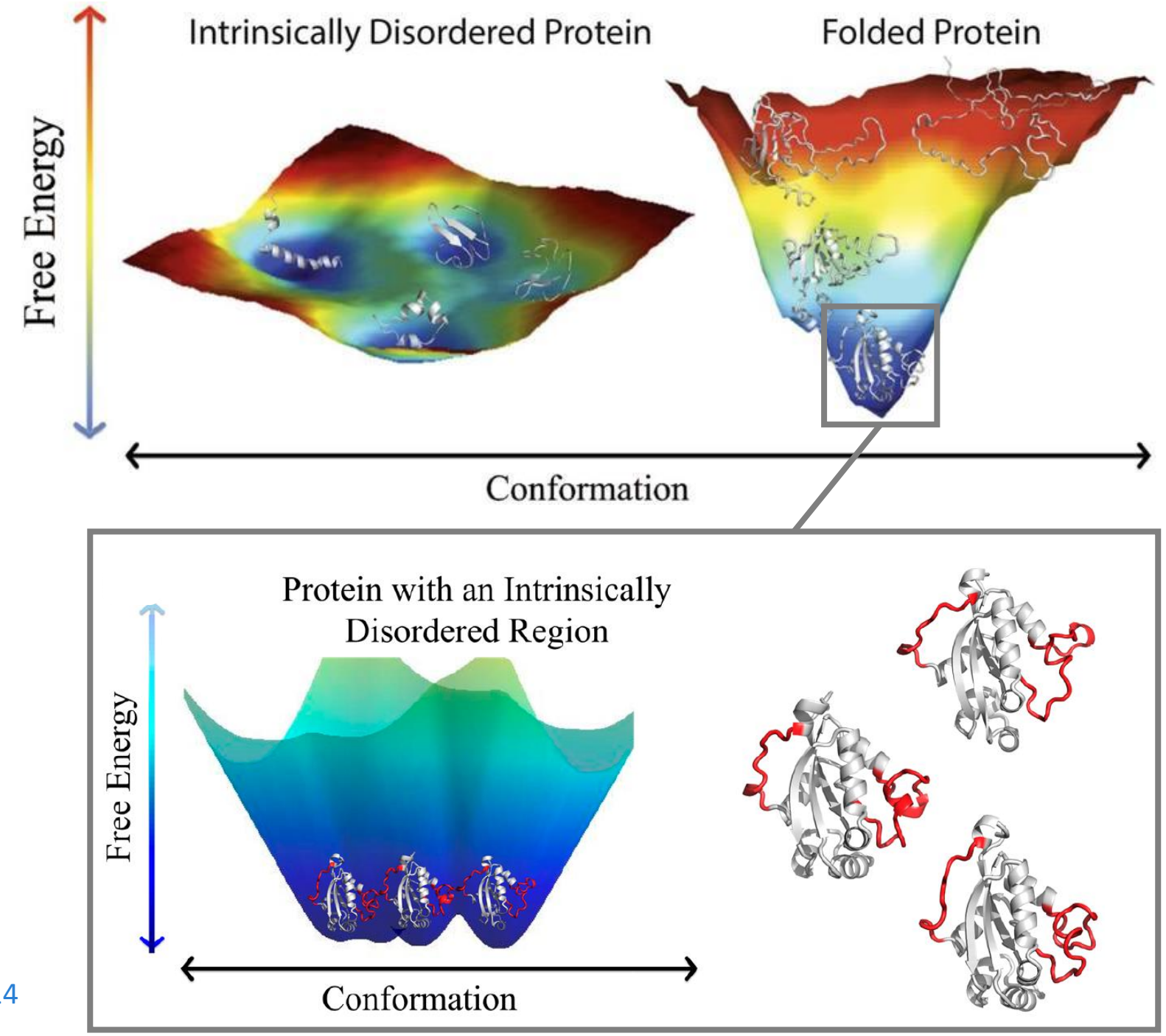


Not all proteins fold

Intrinsically Disordered Proteins (IDP):

- have a shallow energy landscape
- Do not have a well-defined structure

Proteins may contain Intrinsically Disordered Regions (IDR)



V.M. Burger et al., *Intrinsically Disordered Proteins: Where Computation Meets Experiment*, Polymers, 2014

Take home messages

- *The structure determines the function*: proteins function determined by interacting with specific binding partners
- CASP: community effort to improve the capacity of predicting protein 3D structure from their amino acid sequence
- Sequence + environment = conformational space
- Dynamic proteins are difficult to observe experimentally. *Our view of the proteome is biased by the techniques we use to observe it*