Homology modelling

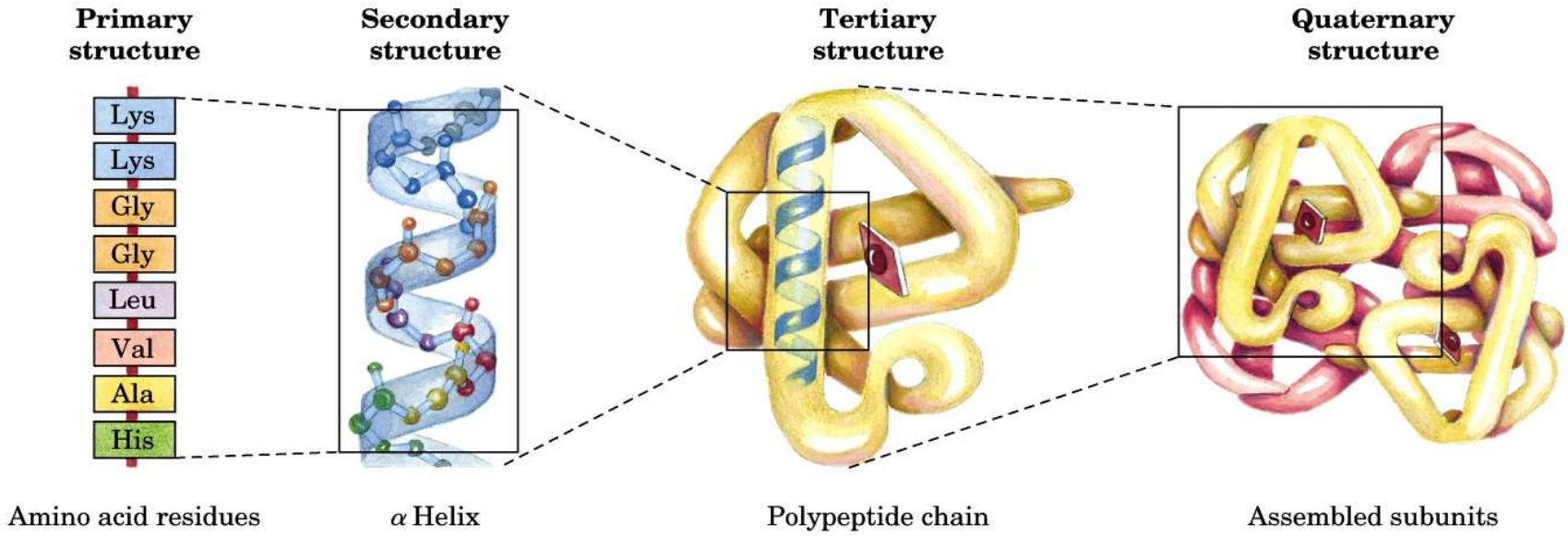
David Abia
dabia@cbm.csic.es

https://tinyurl.com/HMTH26

From Sequence to Structure

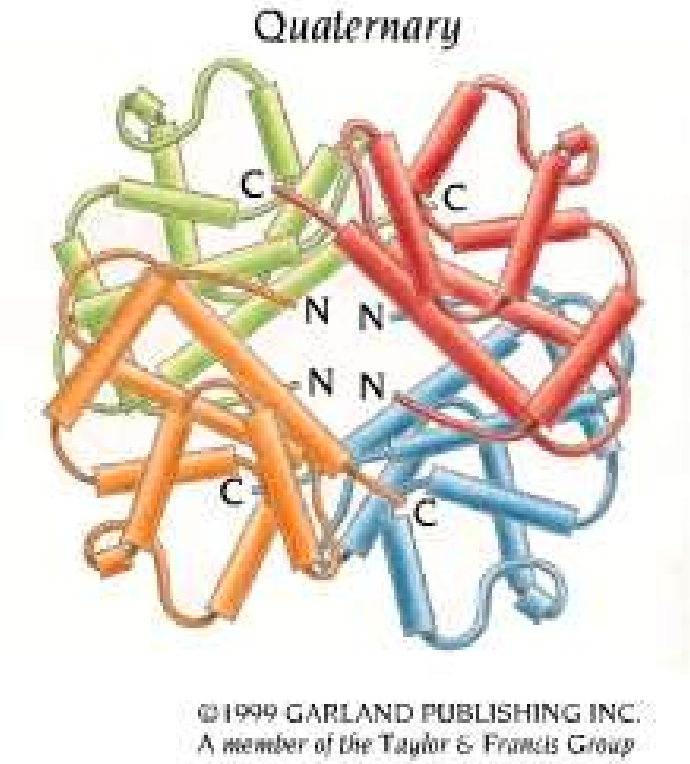
Protein structure is hierarchic:

- Primary – sequence of covalently attached amino acid
- Secondary – local 3D patterns (helices, sheets, loops)
- Tertiary – overall 3D fold
- Quaternary – two or more protein chains



Many Factors Influence Protein Folding

Protein



Proteins Assume the Lowest Energy Structure

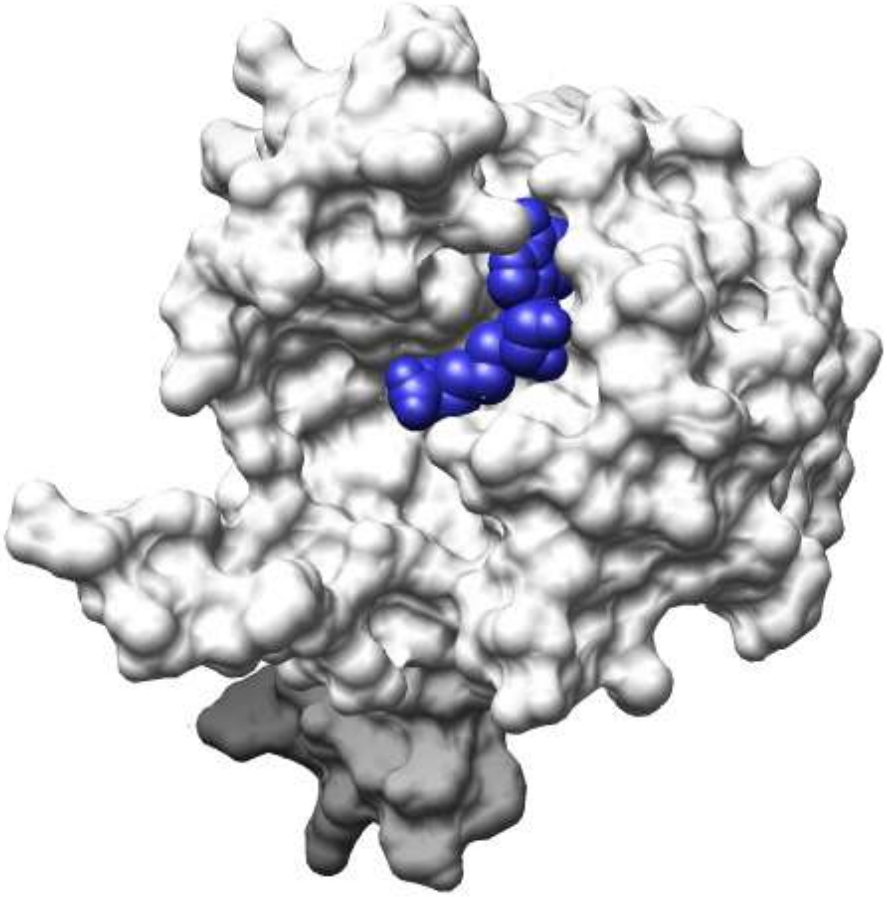
Factors that influence folding include:

1. **Hydrophobic Interactions** / collapse (particularly within the core)
2. **Hydrogen bonds** – lead to secondary structures
3. **Disulfide Bonds** (Cysteine residues)
4. **Salt Bridges** / Ionic Interactions (among charged residues)
5. **Multimeric interactions** with same type or other proteins

Protein 3D Structures

A protein's structure has a critical effect on its function:

1. Binding pockets

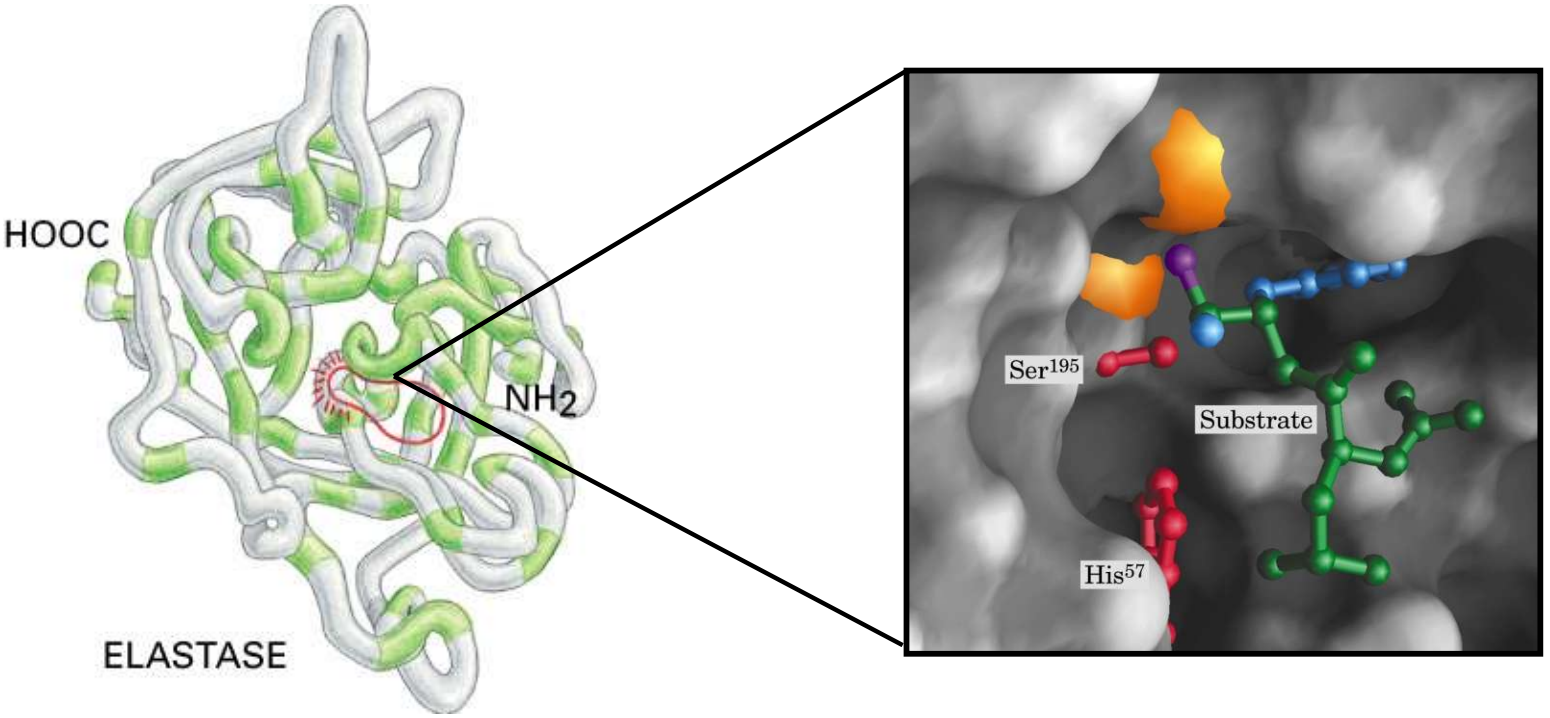


PDB ID 1nw7

Protein 3D Structures

A protein's structure has a critical effect on its function:

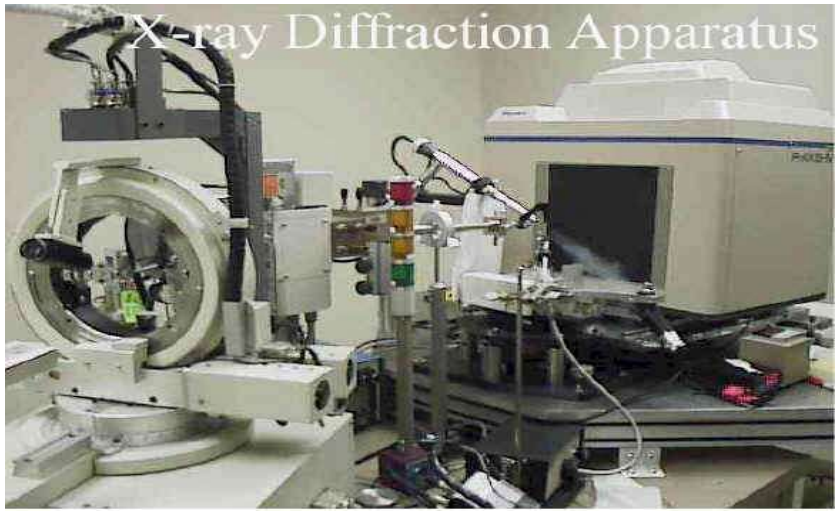
2. Areas of specific chemical\electrical properties

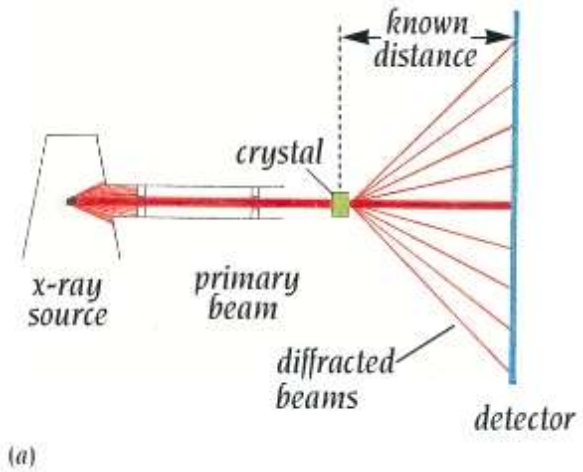


Experimental Methods of Structure Determination

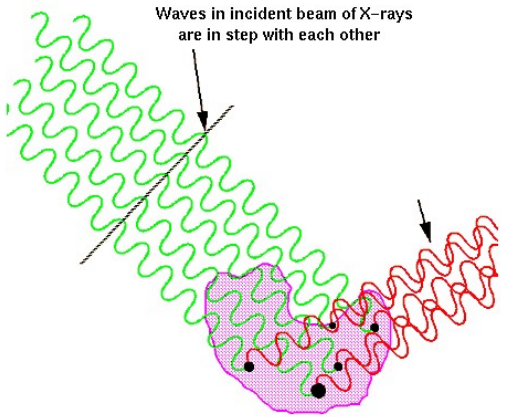
X-ray crystallography

High resolution structure determination



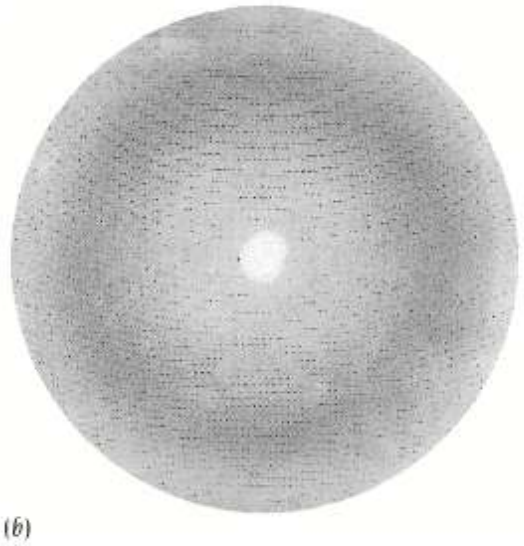


Scattering from the atoms in a molecule



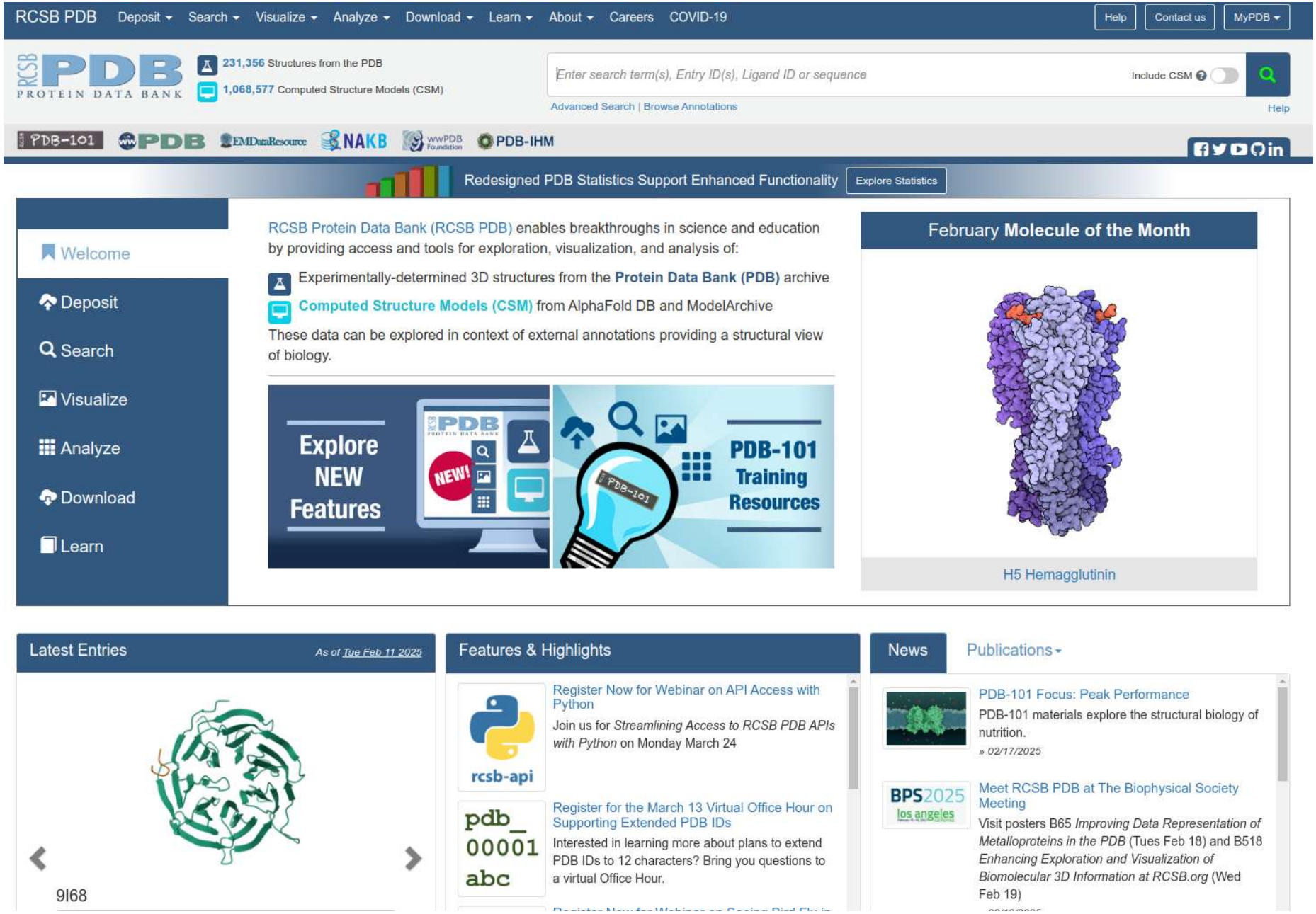
The scattered waves are all shown as having the same amplitude – in practice, the amplitude would vary in proportion to the number of electrons associated with the atom

ACT North 1997



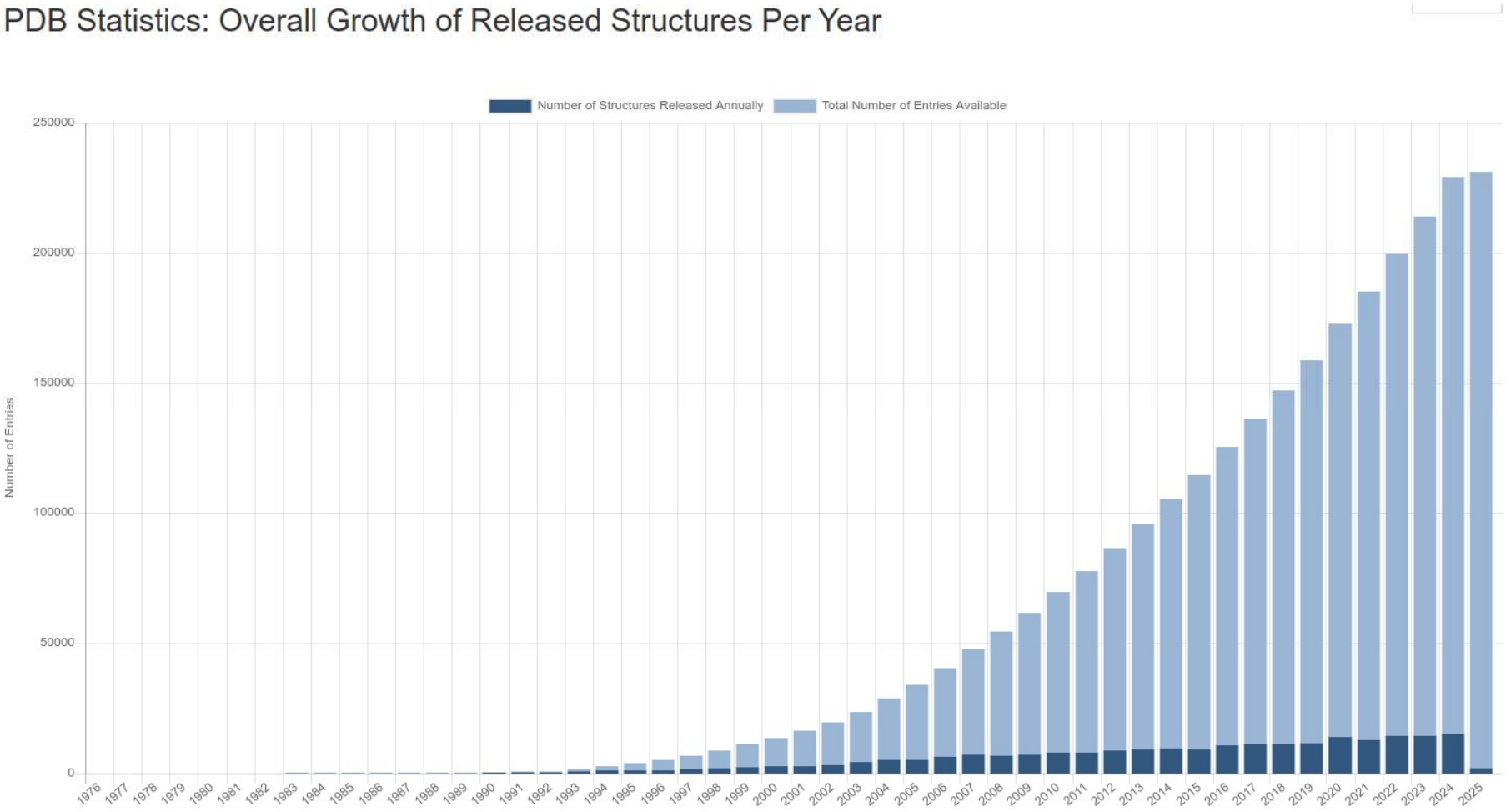
Database of Protein Structures

PDB – Protein Data Bank: https://www.rcsb.org/



Database of Protein Structures

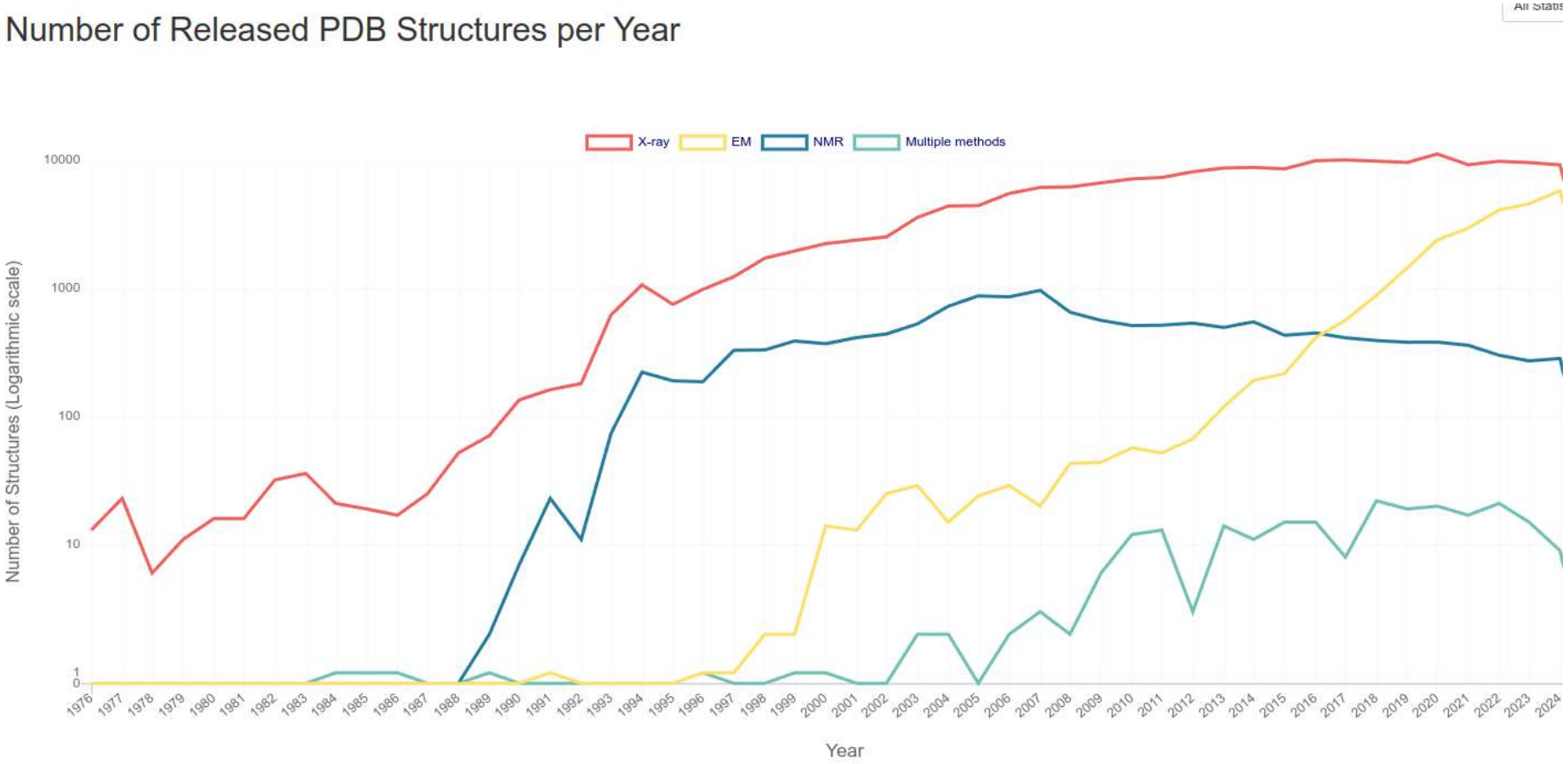
https://www.rcsb.org/stats/growth/growth-released-structures

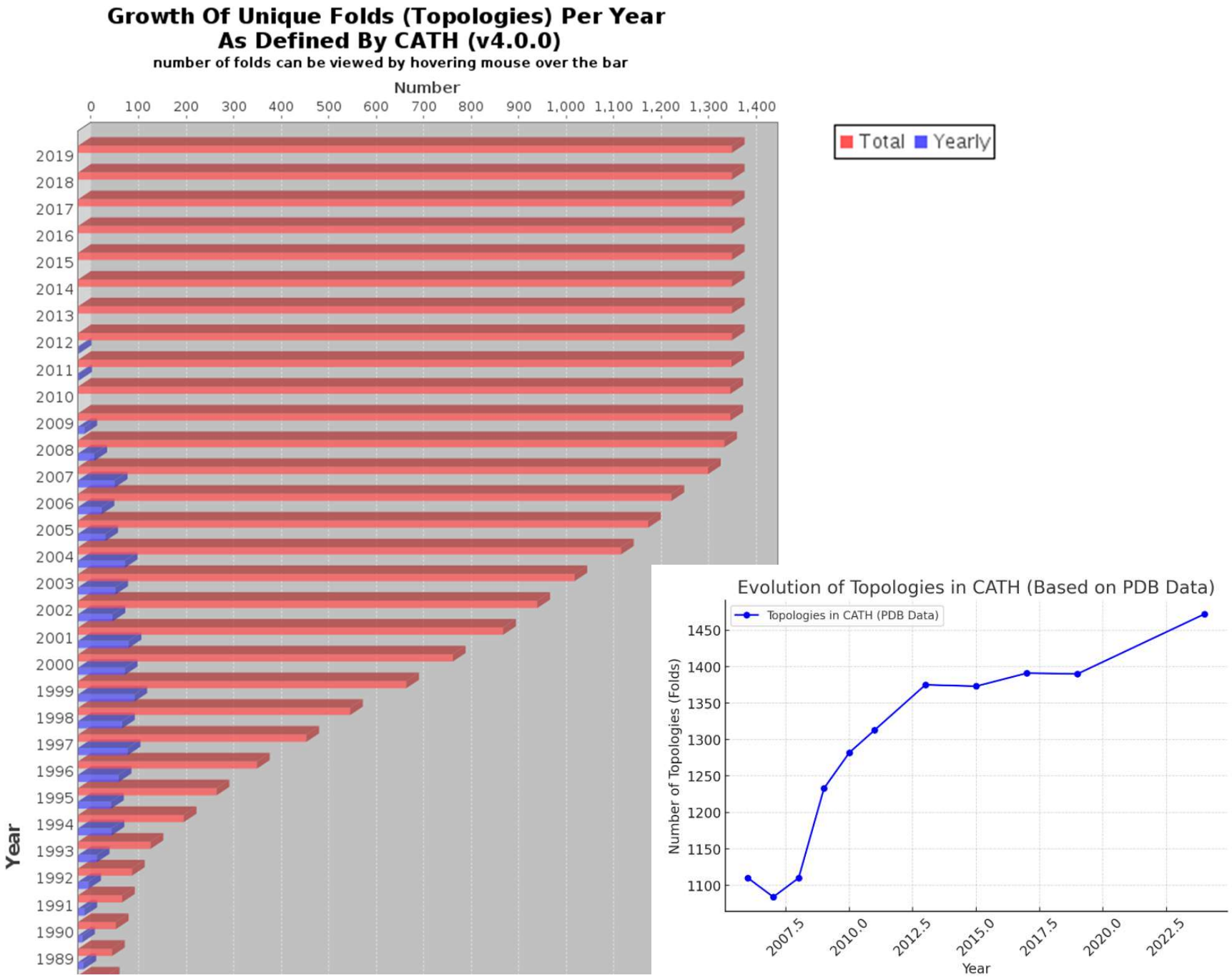


Even so, the number of solved structures greatly lags behind the rate of new genes being sequenced … Solution: Computational Structural Methods

Obtained structures per technique

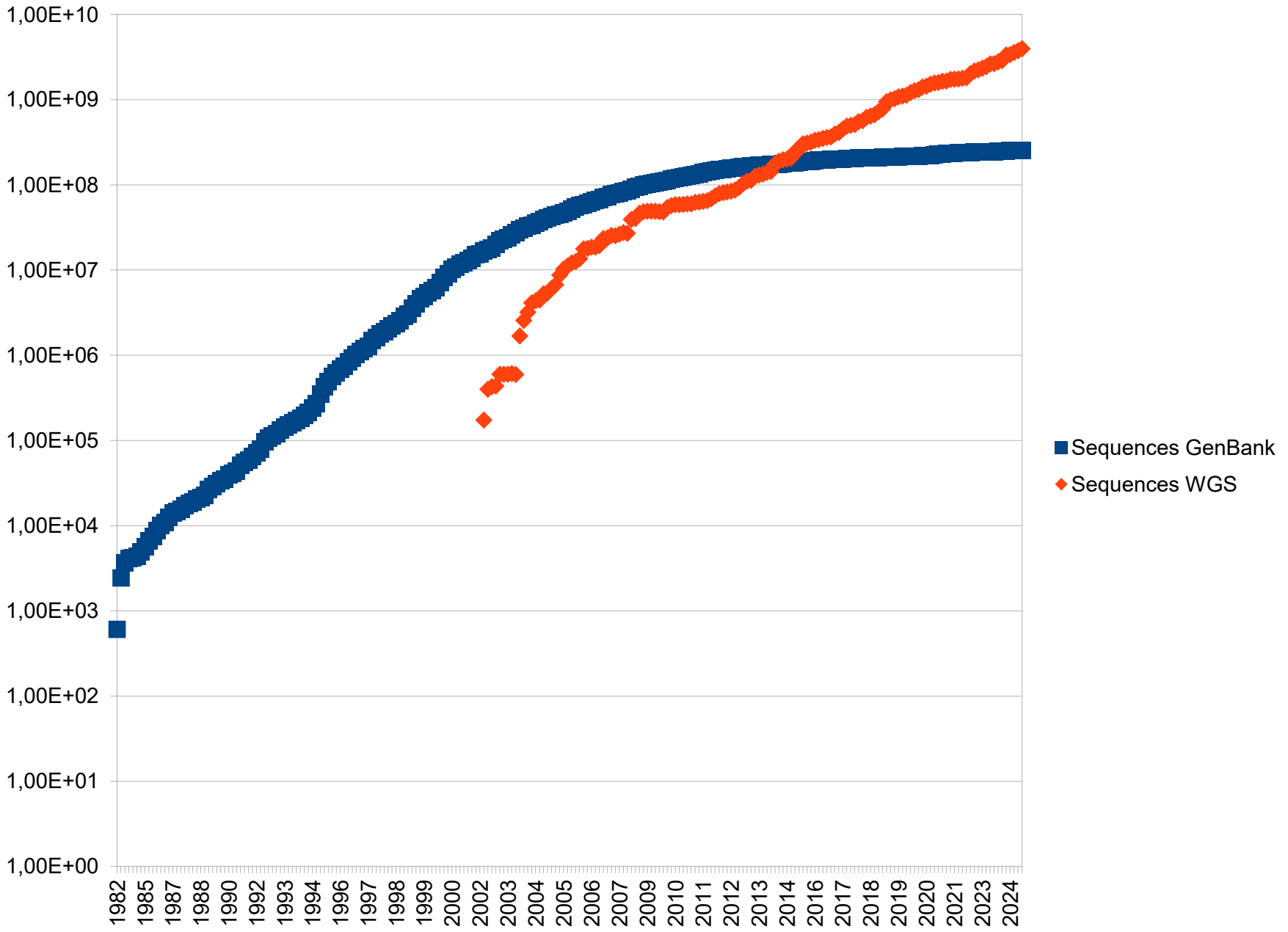
https://www.rcsb.org/stats/all-released-structures





http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=fold-cath

GenBank Sequences



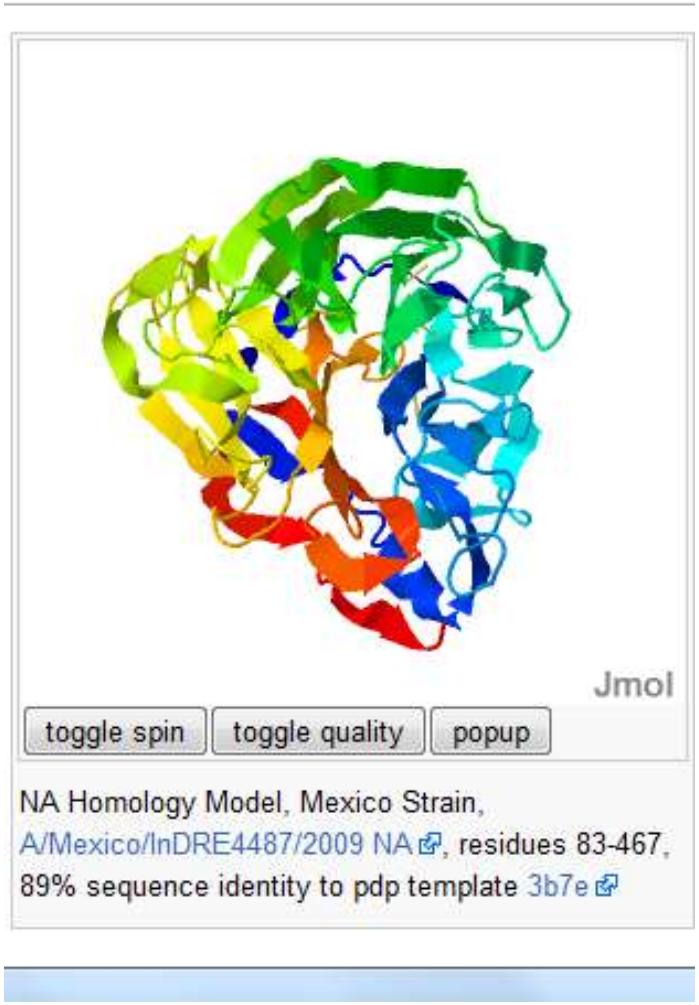
https://www.ncbi.nlm.nih.gov/genbank/statistics/

From Sequence to Structure

H1N1 NA

MNPNQKIITIGSVCMTIGMANLIL
QIGNIISIWISHSIQLGNQN
QIETCNQSVITYENNTWVNQTY
VNISNTNFAAGQSVVSVKLAGNSS
LCPVSGWAIYSK
DNSVRIGSKGDVFVIREPFISCSPLE
CRTFFLTQGALLNDKHSNGTIKD
RSPYRTLMS
CPIGEVPSPYNSRFESVAWSASACH
DGINWLTIGISGPDNGAVAVLKYN
GIITDTIKS
WRNNILRTQESECACVNGSCFTV
MTDGPSNGQASYKIFRIEKGKIVKS
VEMNAPNYHY
EECSCYPDSSEITCVCRDNWHGS
NRPWVSFNQNLEYQIGYICSGIFG
DNPRPNDKTGS
CGPVSSNGANGVKGFSFKYGNGV
WIGRTKSISSRNGFEMIWDPNGW
TGTDNNFSIKQD
IVGINEWSGYSGSFVQHPELTGLD
CIRPCFWVELIRGRPKENTIWTSGS
SISFCGVNS
DTVGWSWPDGAELPFTIDK"

Computational Approach

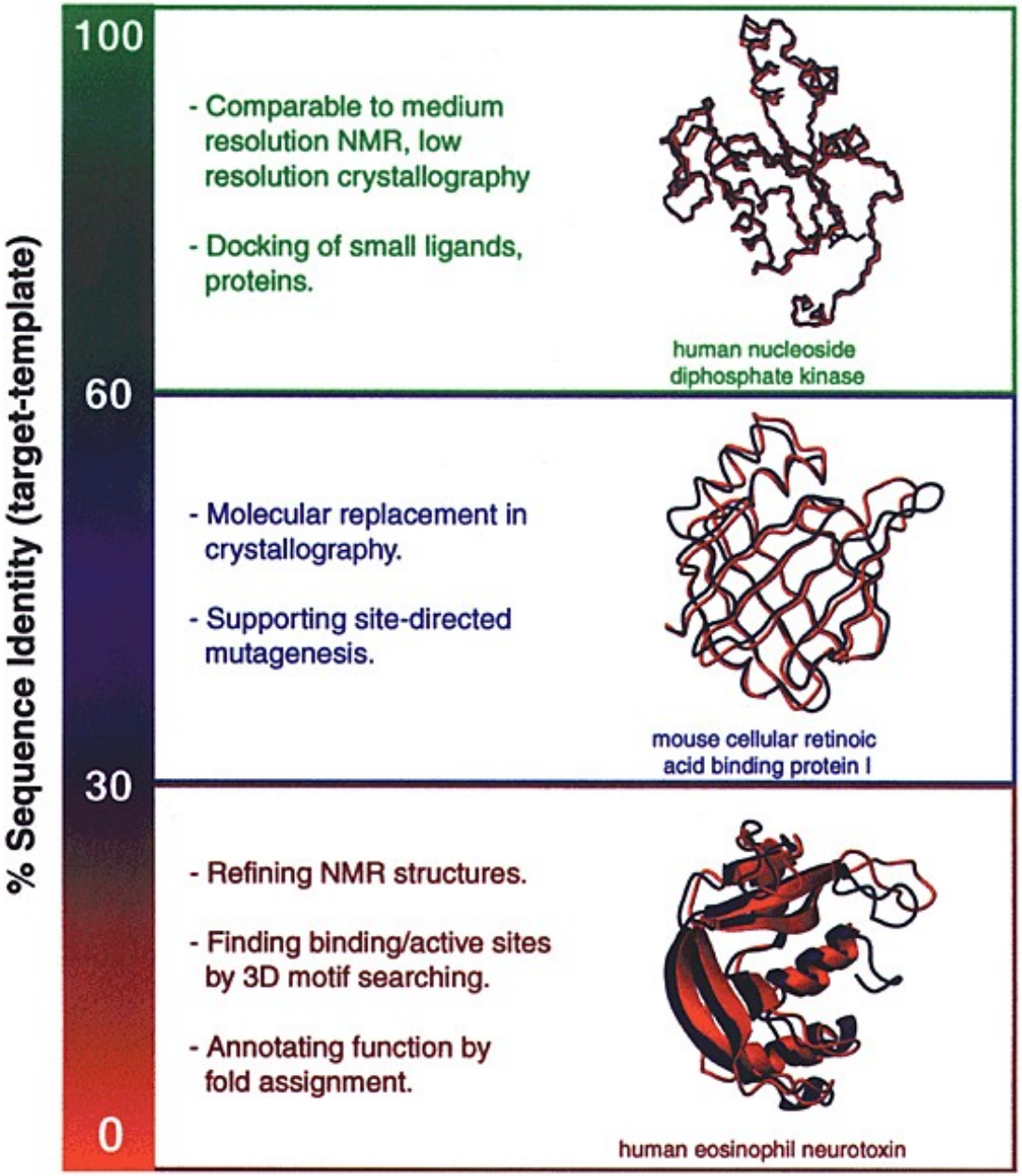


Comparative Modeling "Homology Modeling"

Steps include:

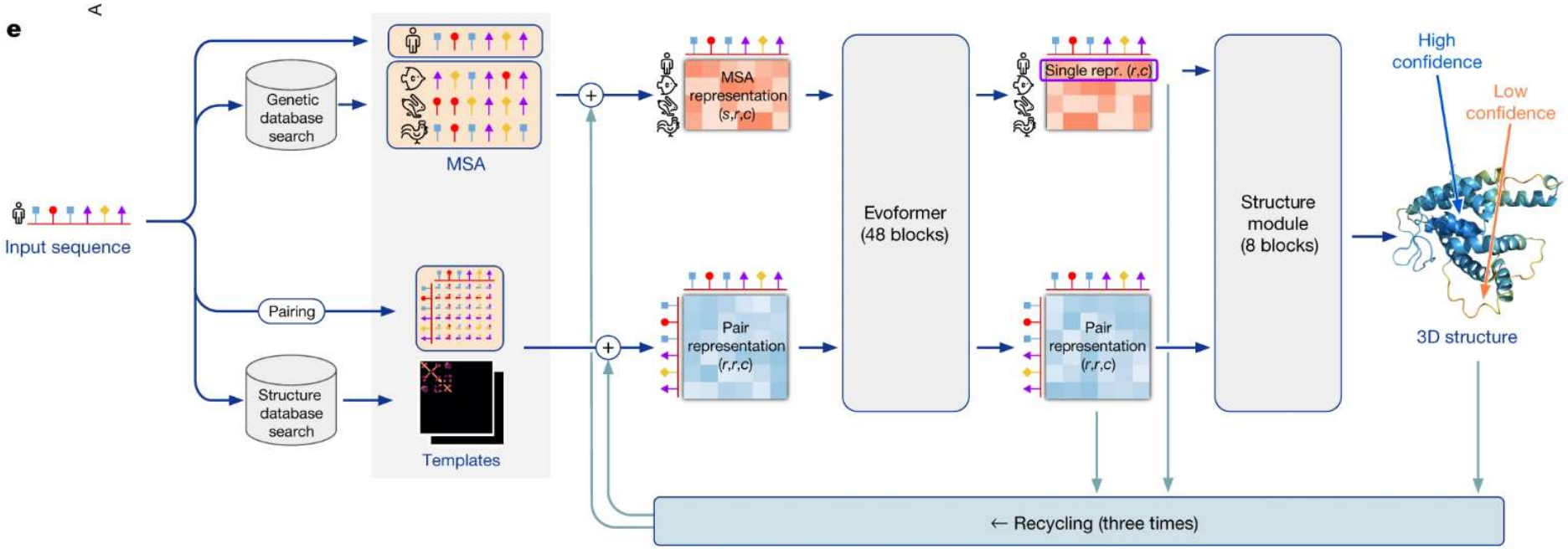
1. **Template recognition and initial alignment**
2. **Alignment Correction (Multiple Sequence Alignment can Help)**
3. **Backbone Generation (transfer coordinates from template)**
4. **Loop Modeling (loops hard to predict with insertions)**
5. **Side Chain Modeling (usually similar torsion angles at high sequenc ID)**
6. **Model Optimization (minor energy minimization steps or restrain some atom positions)**
7. **Model Validation (Higher ID more accurate usually, Calculate energy, or normality index (bond length, torsion angles))**
8. **Iteration (to refine)**

How Well Can We Do It?



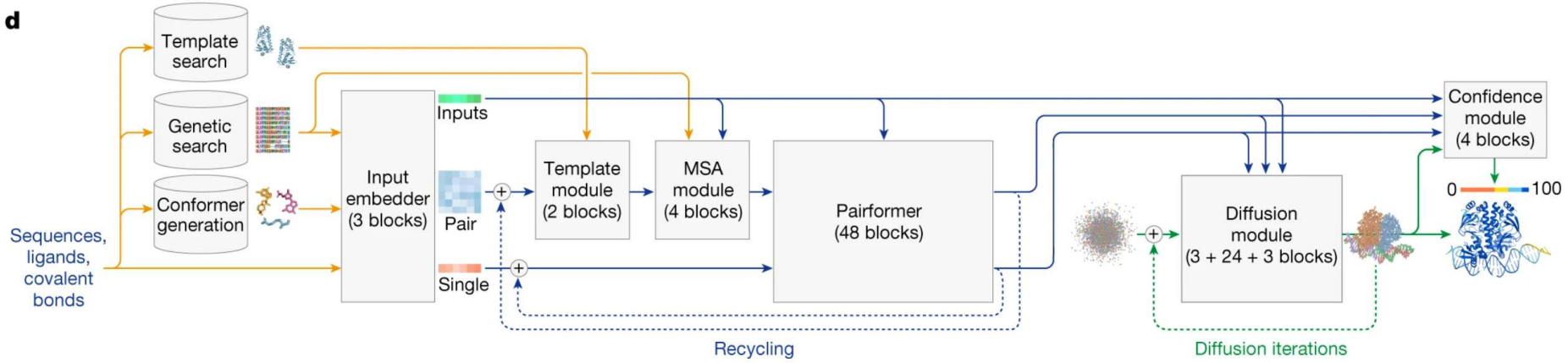
Sali, A. & Kuriyan, J. *Trends Biochem. Sci.* **22**, M20–M24 (1999)

Alphafold 2



Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

Alphafold 3



Abramson, J., Adler, J., Dunger, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w