



Homology modelling

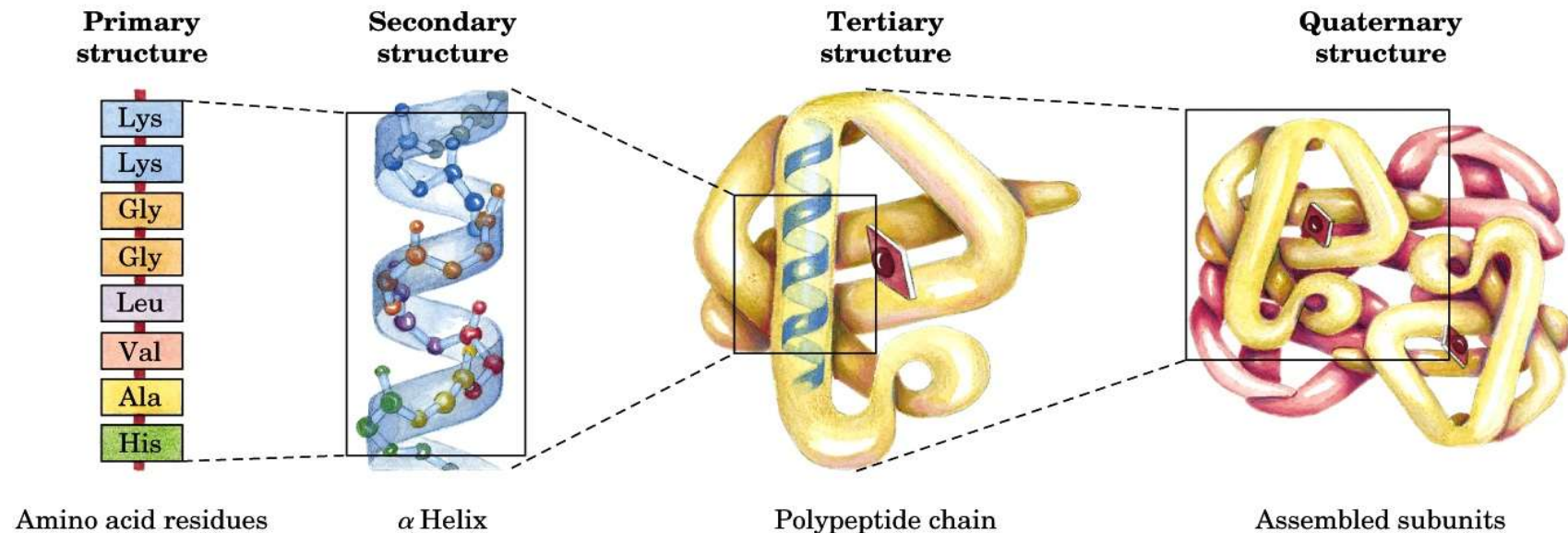
David Abia
dabia@cbm.csic.es

<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

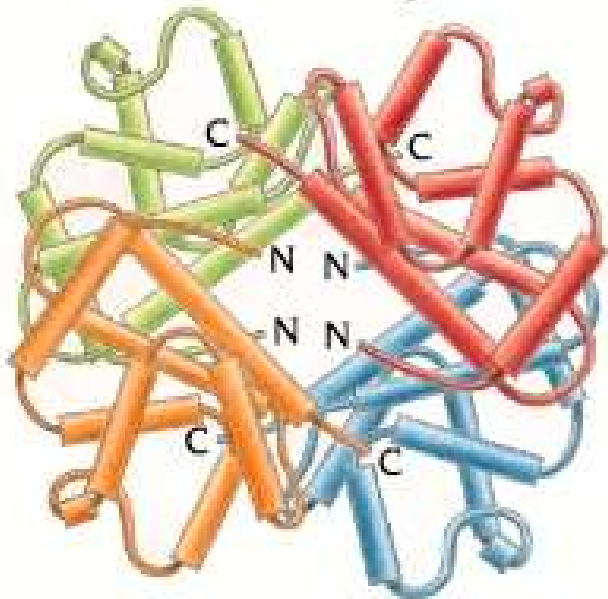


Many Factors Influence Protein Folding

Proteins Assume the Lowest Energy Structure

Protein

Quaternary



© 1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

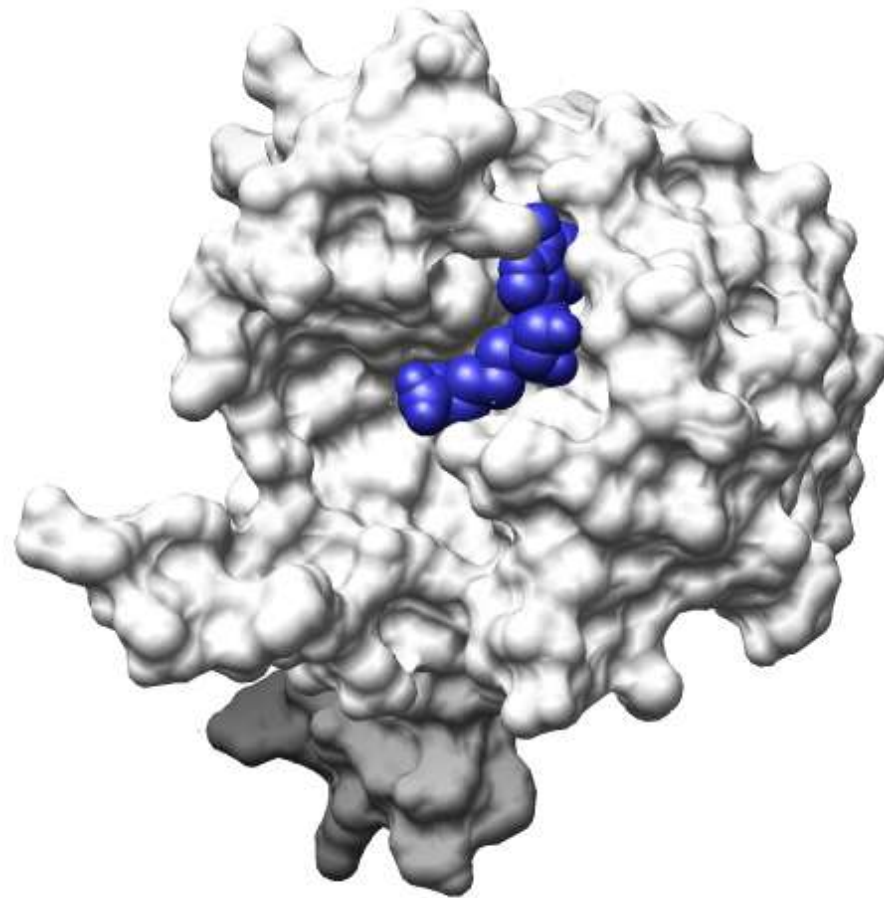
Factors that influence folding include:

- 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

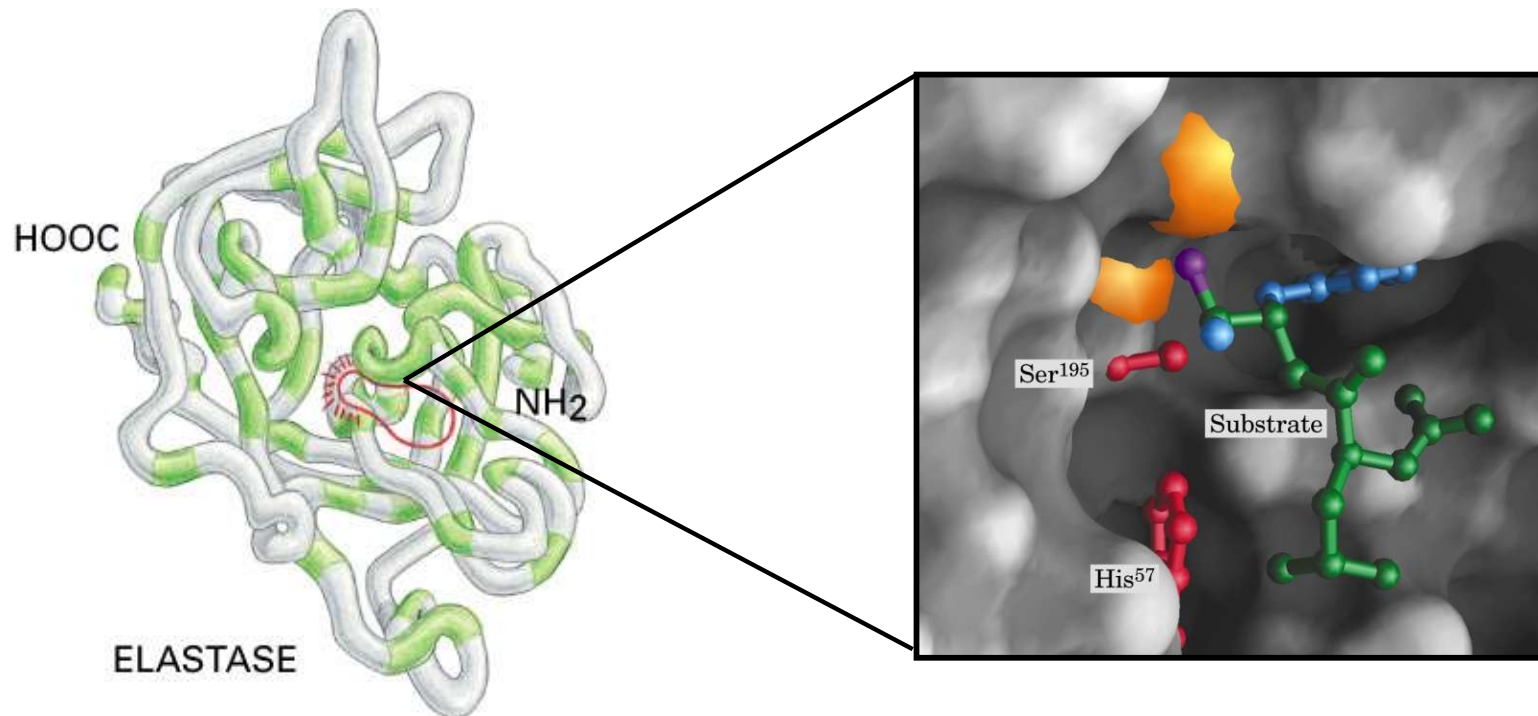
A 3D surface representation of a protein structure, shown in white and grey. The protein is a complex, multi-domain structure with a prominent binding pocket. Inside this pocket, a small molecule ligand is bound, represented by a cluster of blue spheres. The protein's surface is highly textured, showing various protrusions and indentations. The ligand is positioned in a way that suggests it is interacting with the protein's binding site.

PDB ID 1nw7

Protein 3D Structures

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

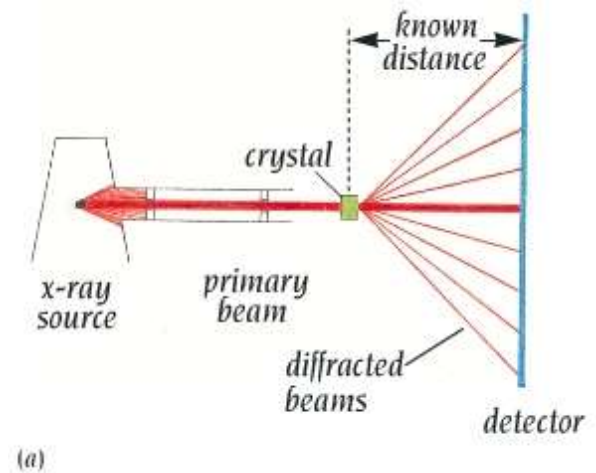
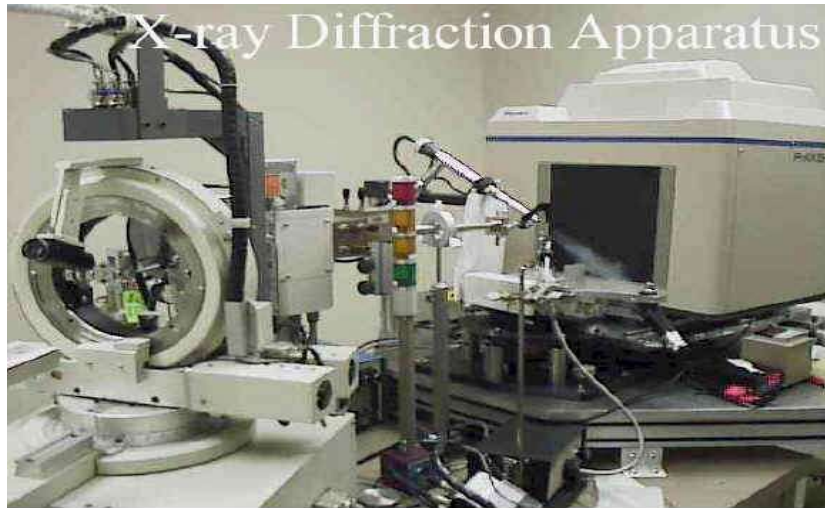
2. Areas of specific chemical\electrical properties



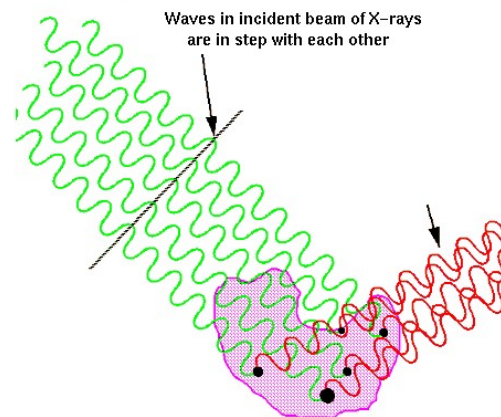
Experimental Methods of Structure Determination

X-ray crystallography

High resolution structure determination

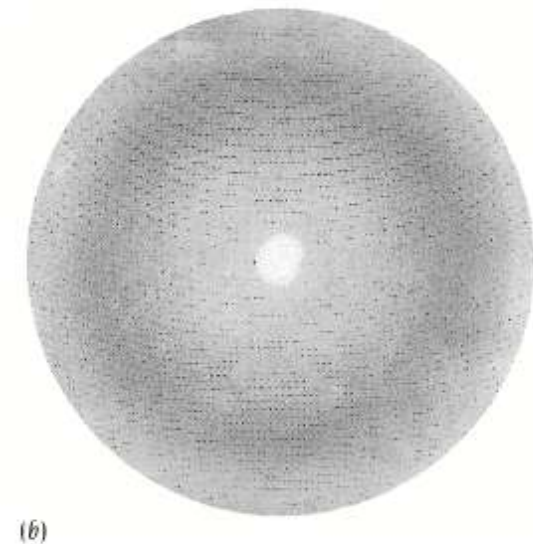


Scattering from the atoms in a molecule



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

RCT North 1997



Database of Protein Structures

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

The screenshot shows the RCSB PDB website homepage. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, About, Careers, and COVID-19. A search bar is prominently displayed with the text "Enter search term(s), Entry ID(s), Ligand ID or sequence". Below the search bar, there are statistics: "231,356 Structures from the PDB" and "1,068,577 Computed Structure Models (CSM)". A banner for "PDB-101" and "PDB-IHM" is visible. The main content area features a "Welcome" message, a list of navigation options (Deposit, Search, Visualize, Analyze, Download, Learn), and a central section titled "RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:". This section lists "Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive" and "Computed Structure Models (CSM) from AlphaFold DB and ModelArchive". A banner for "Explore NEW Features" and "PDB-101 Training Resources" is also present. On the right, there is a "February Molecule of the Month" section featuring a 3D structure of H5 Hemagglutinin.

Latest Entries

As of *Tue Feb 11 2025*

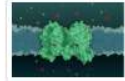

9168

Features & Highlights

-  **rcsb-api**
Register Now for Webinar on API Access with Python
Join us for *Streamlining Access to RCSB PDB APIs with Python* on Monday March 24
- pdb_00001abc**
Register for the March 13 Virtual Office Hour on Supporting Extended PDB IDs
Interested in learning more about plans to extend PDB IDs to 12 characters? Bring you questions to a virtual Office Hour.

News

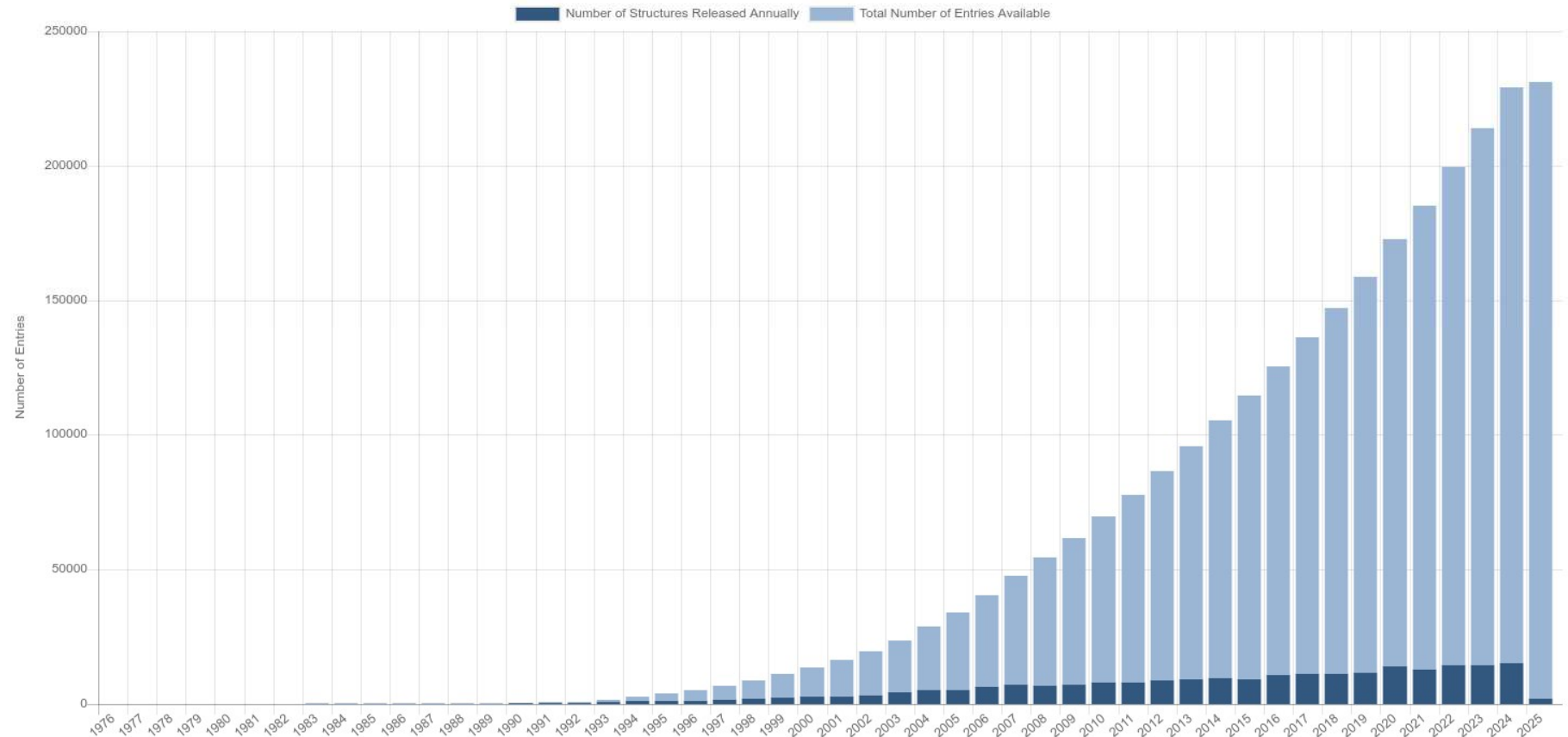
Publications

-  **PDB-101 Focus: Peak Performance**
PDB-101 materials explore the structural biology of nutrition.
» 02/17/2025
-  **Meet RCSB PDB at The Biophysical Society Meeting**
Visit posters B65 *Improving Data Representation of Metalloproteins in the PDB* (Tues Feb 18) and B518 *Enhancing Exploration and Visualization of Biomolecular 3D Information at RCSB.org* (Wed Feb 19)

Database of Protein Structures

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

PDB Statistics: Overall Growth of Released Structures Per Year

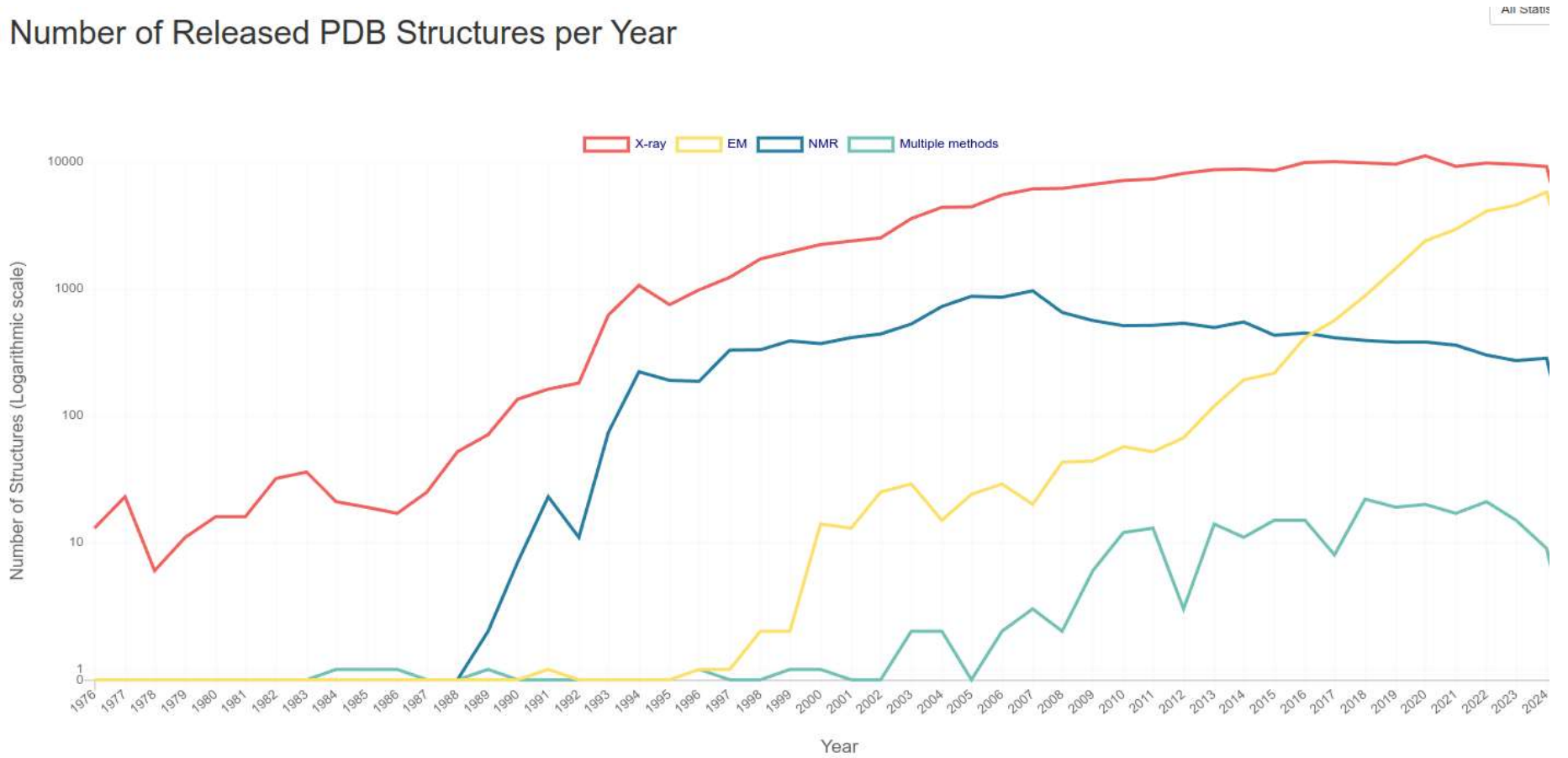


Even so, the number of solved structures greatly lags behind the rate of new genes being sequenced ... Solution: Computational Structural Methods

Obtained structures per technique

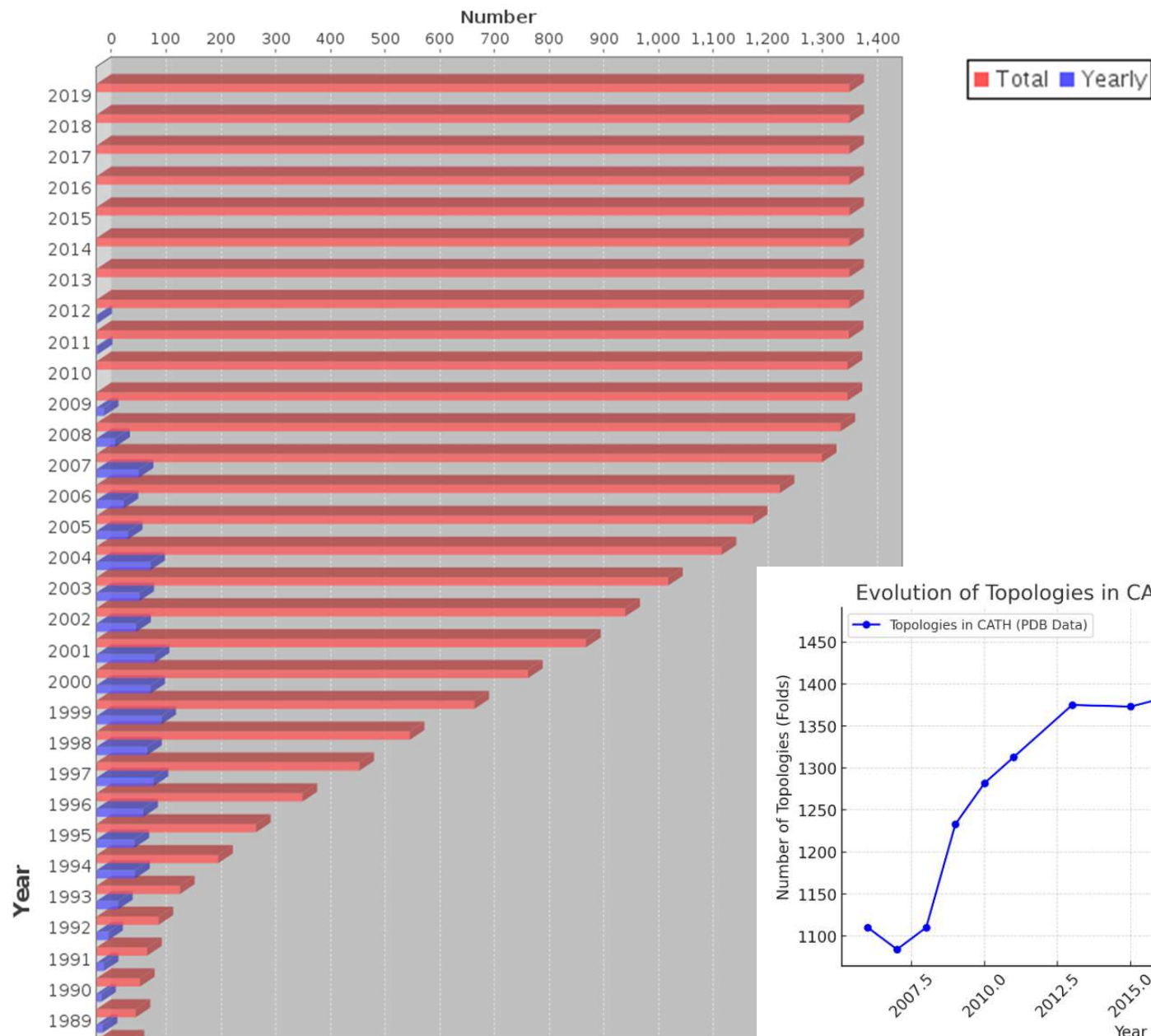
<https://www.rcsb.org/stats/all-released-structures>

Number of Released PDB Structures per Year



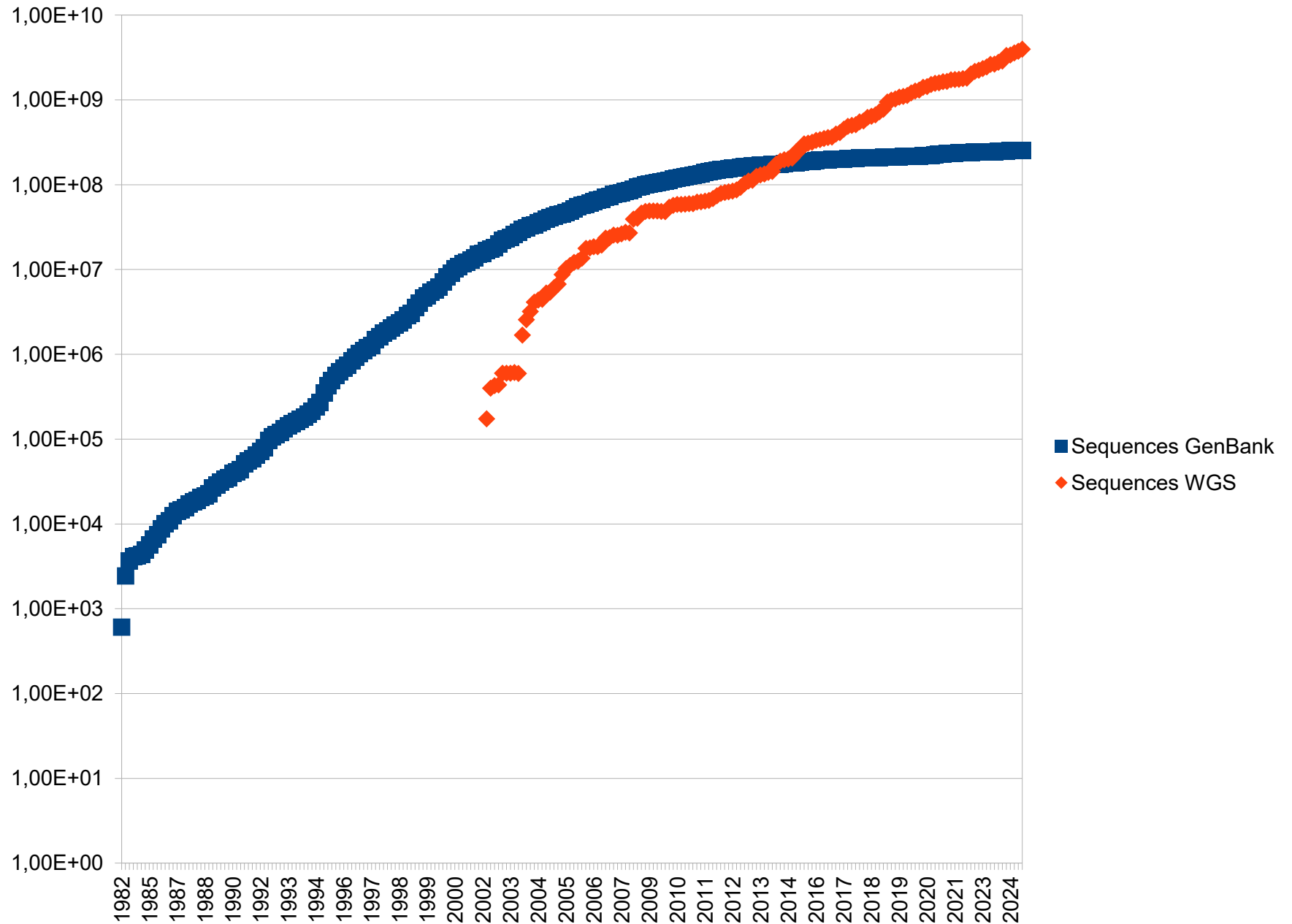
Growth Of Unique Folds (Topologies) Per Year As Defined By CATH (v4.0.0)

number of folds can be viewed by hovering mouse over the bar



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

GenBank Sequences



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

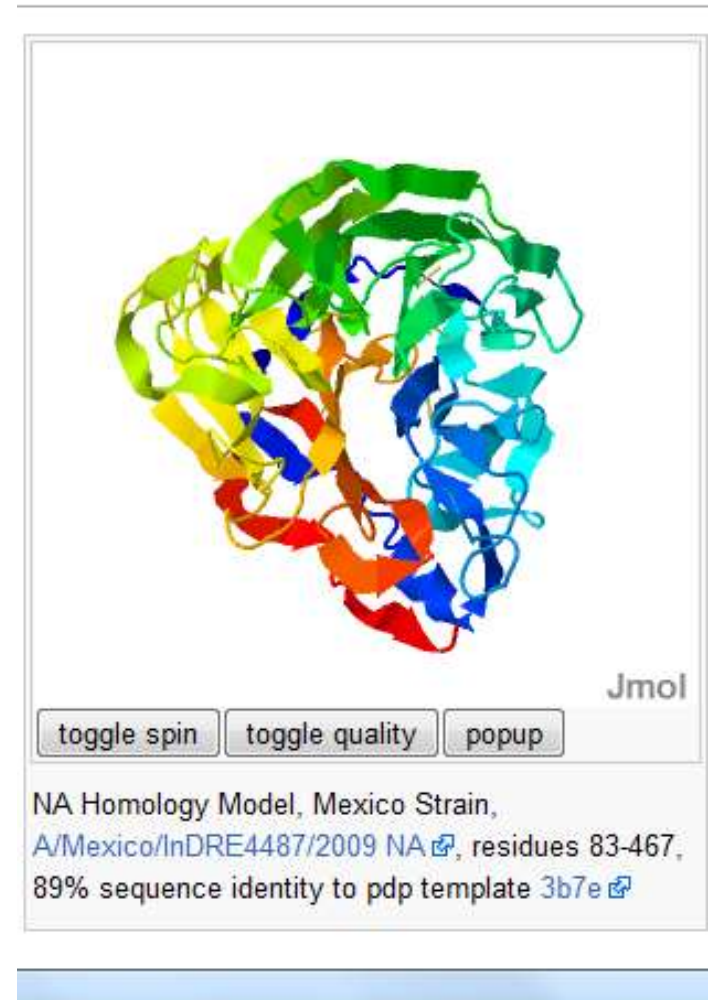
From Sequence to Structure

H1N1 NA

MNPNQKIITIGSVCMTIGMANLIL
QIGNIISIWISHSIQLGNQN
QIETCNQSVITYENNTWVNQTY
VNISNTNFAAGQSVVSVKLAGNSS
LCPVSGWAIYSK
DNSVRIGSKGDVFIREFPISCSPLE
CRTFFLTQGALLNDKHSNGTIKD
RSPYRTLMS
CPIGEVPSPYNSRFESVAWSASACH
DGINWLTIGISGPDNGAVAVLKYN
GIITDTIKS
WRNNILRTQESECACVNGSCFTV
MTDGPSNGQASYKIFRIEKGKIVKS
VEMNAPNYHY
EECSCYPDSEITCVCRDNWHGS
NRPWVSFNQNLEYQIGYICSGIFG
DNPRPNDKTGS
CGPVSSNGANGVKGFSEFKYGNGV
WIGRTKSISSRNGFEMIWDPNGW
TGTDNNFSIKQD
IVGINEWVSGYSGSFVQHPELTGLD
CIRPCFWVELIRGRP KENTIWTSGS
SISFCGVNS
DTVGWSWPDGAELPFTIDK"



Computational
Approach



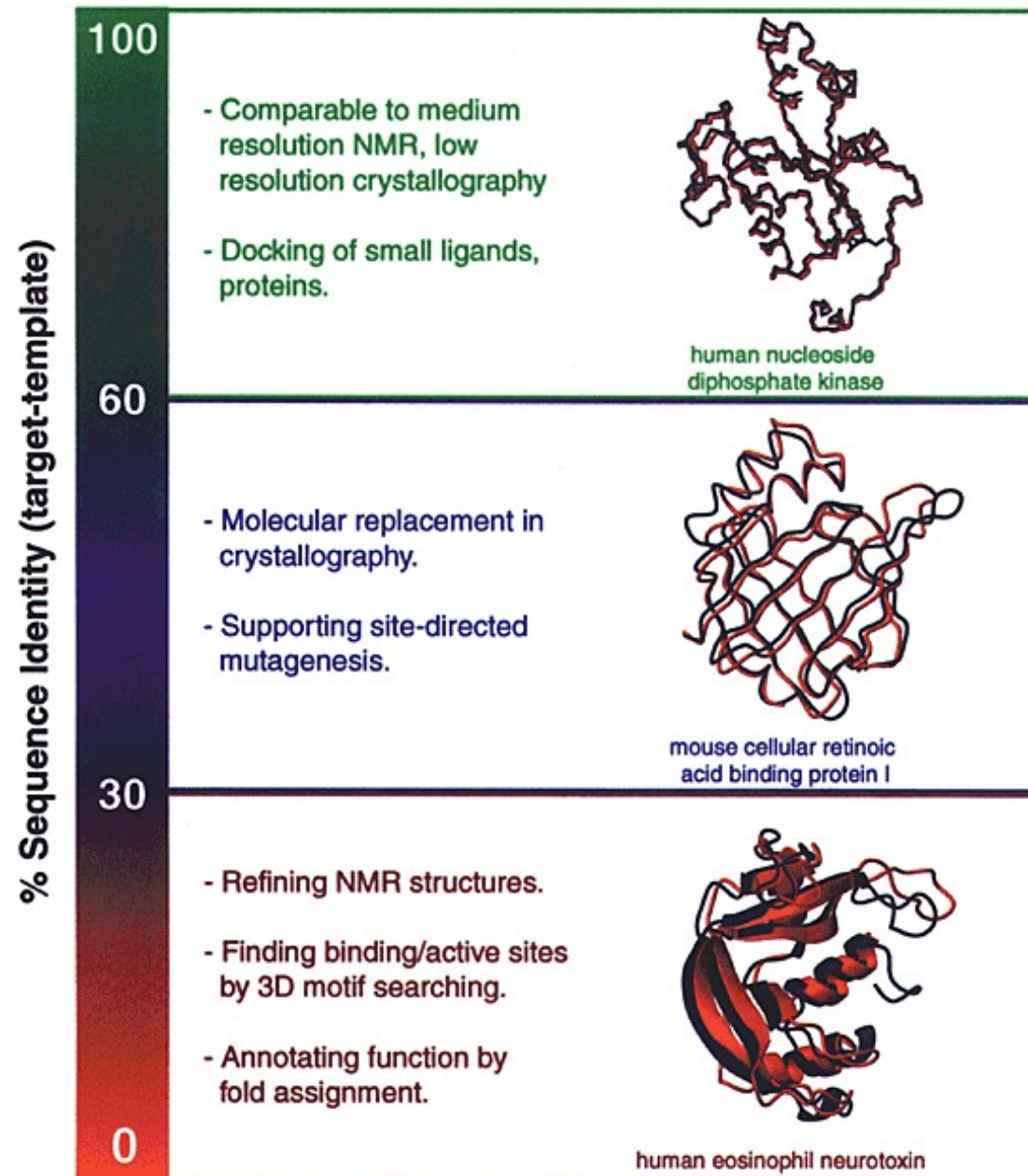
Comparative Modeling “Homology Modeling”

Steps include:

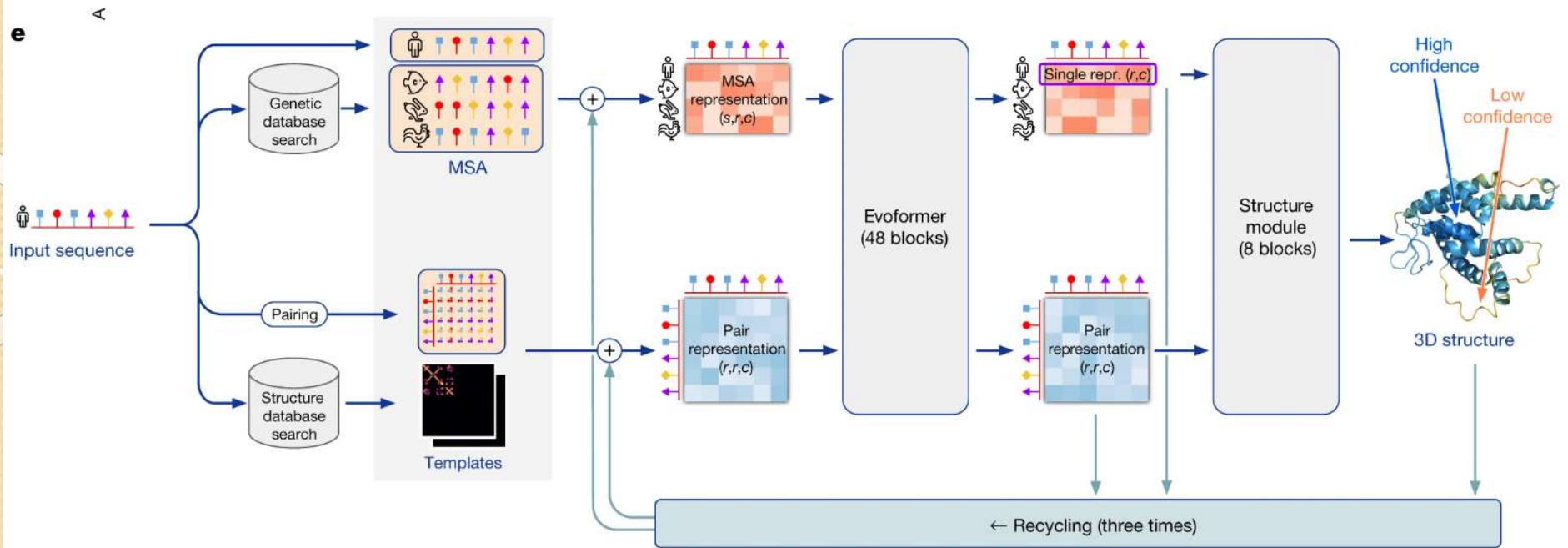
- 1. Template recognition and initial alignment**
- 2. Alignment Correction** (Multiple Sequence Alignment can Help)
- 3. Backbone Generation** (transfer coordinates from template)
- 4. Loop Modeling** (loops hard to predict with insertions)
- 5. Side Chain Modeling** (usually similar torsion angles at high sequence 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)

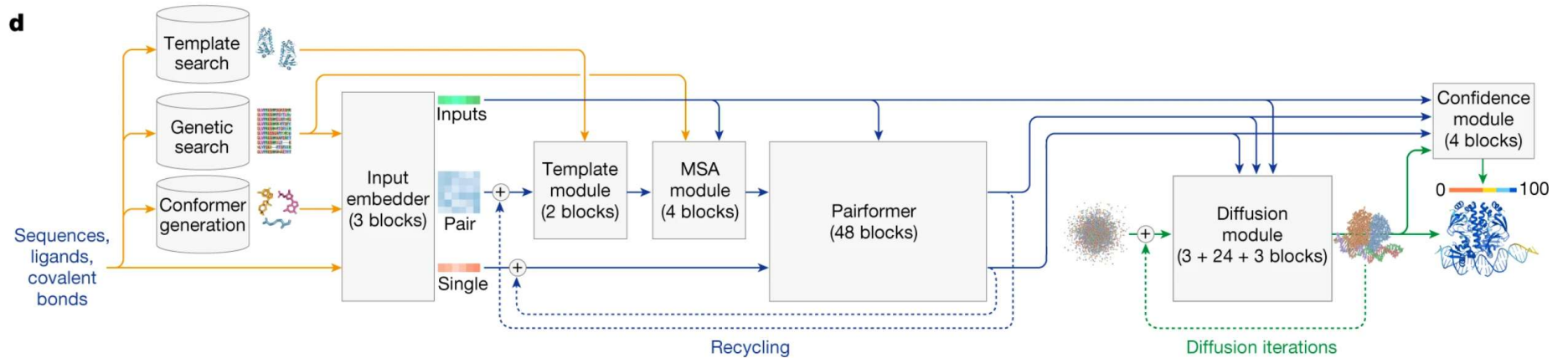


AlphaFold 2



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>