3000788 Intro to Comp Molec Biol

Lecture 20: RNA and protein structural models

Fall 2025

Sira Sriswasdi, PhD

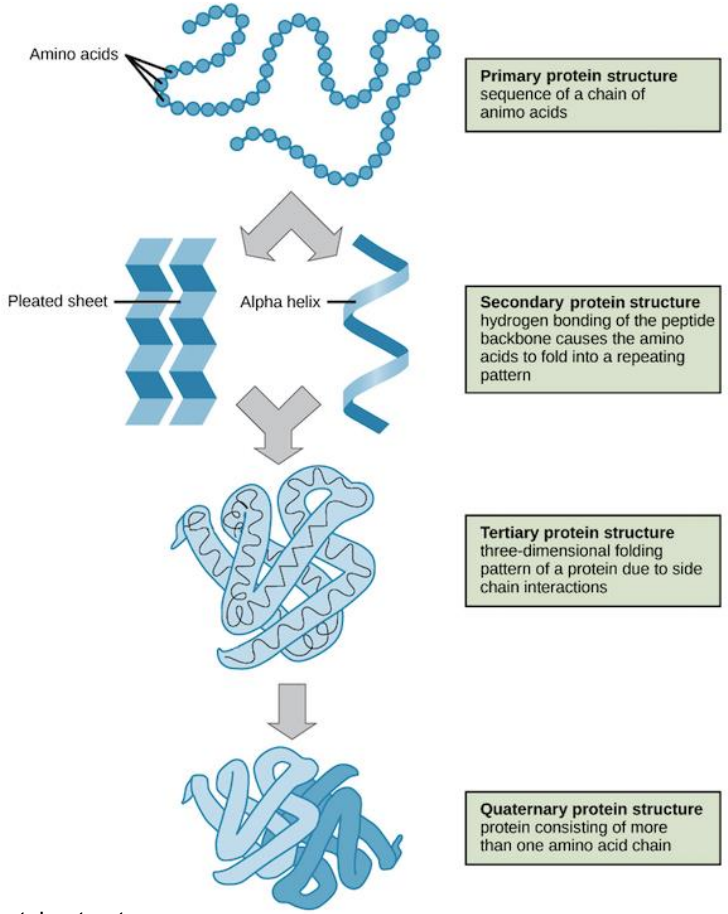
- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

Today's agenda

- Importance of RNA and protein structures
- Assay for studying RNA and protein structures
- Computational modeling of RNA and protein structure

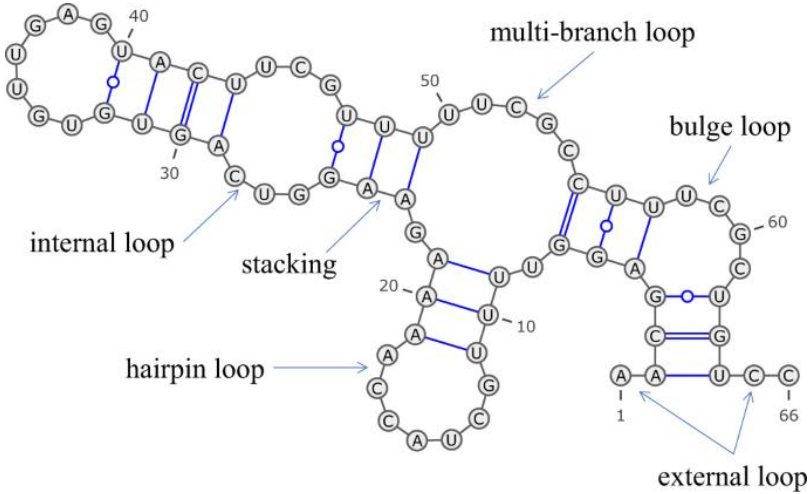
RNA and protein structures

- Primary structure = sequence
- Secondary structure*
- Tertiary structure = 3D structure
- Quaternary structure = multi-molecule complex

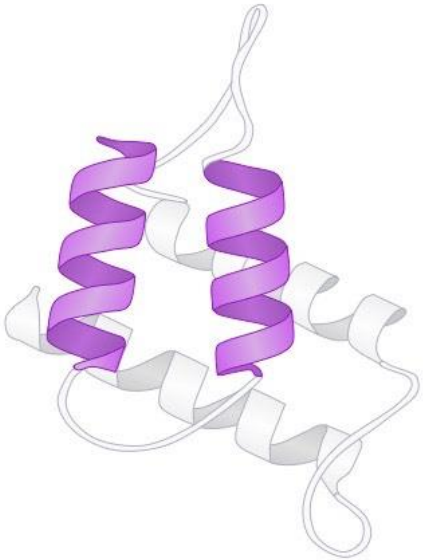


https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/a/orders-of-protein-structure

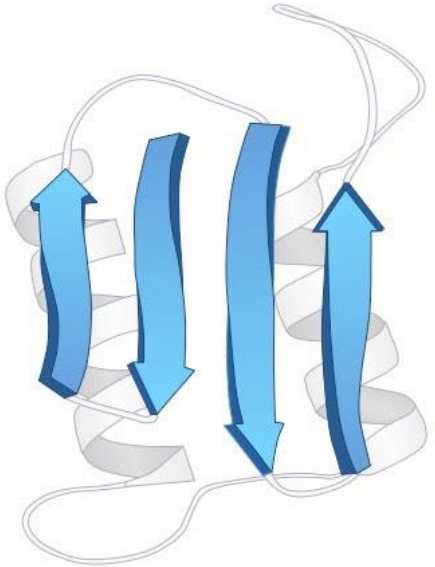
Secondary structures



Sato, K. et al. Nature Communications 12:941 (2021)



α - helices

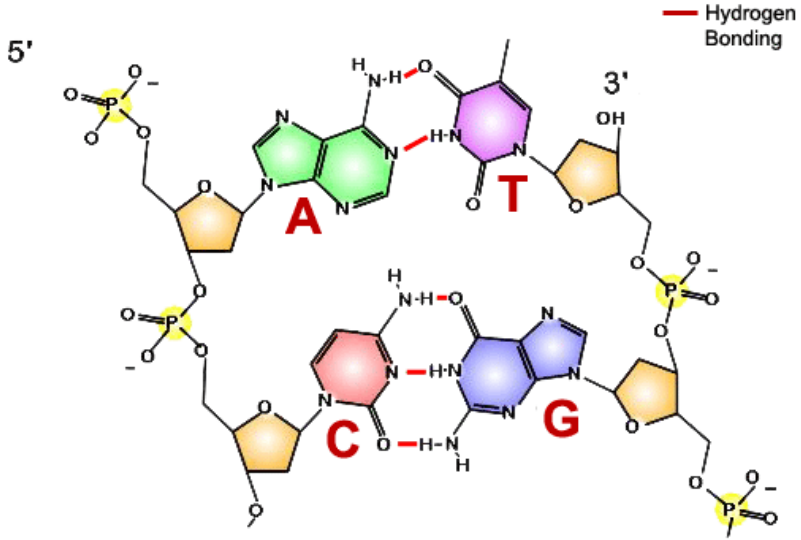


β - pleated sheets

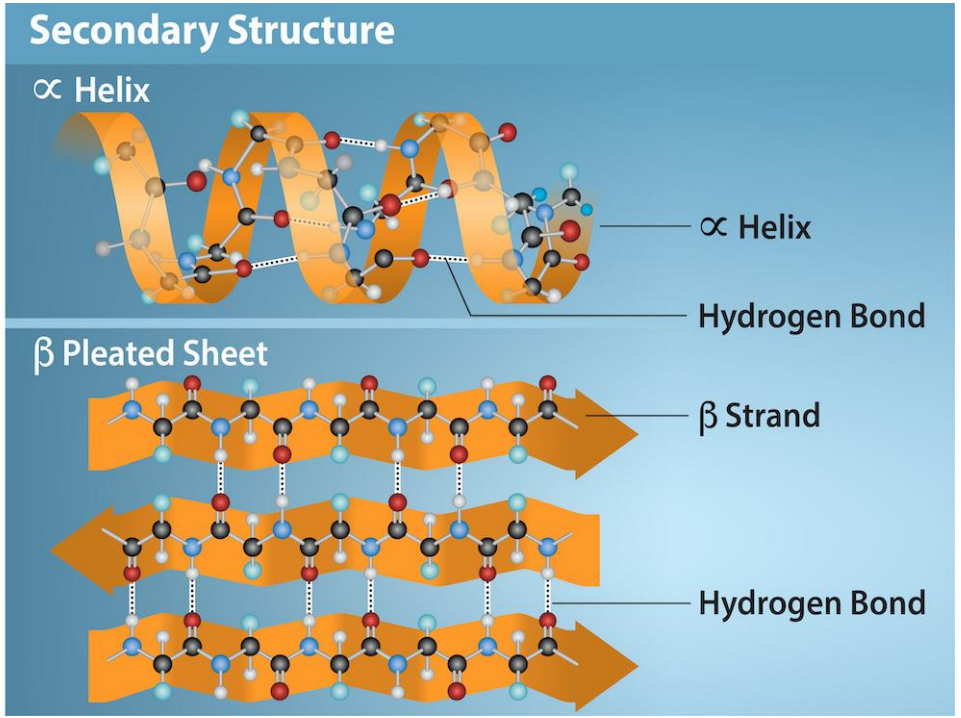
https://old-ib.bioninja.com.au/standard-level/topic-2-molecular-biology/24-proteins/protein-structure.html

- Defined by hydrogen bonds between complementary nucleotides or the C=O and NH part of nearby peptide bonds

Hydrogen bonds

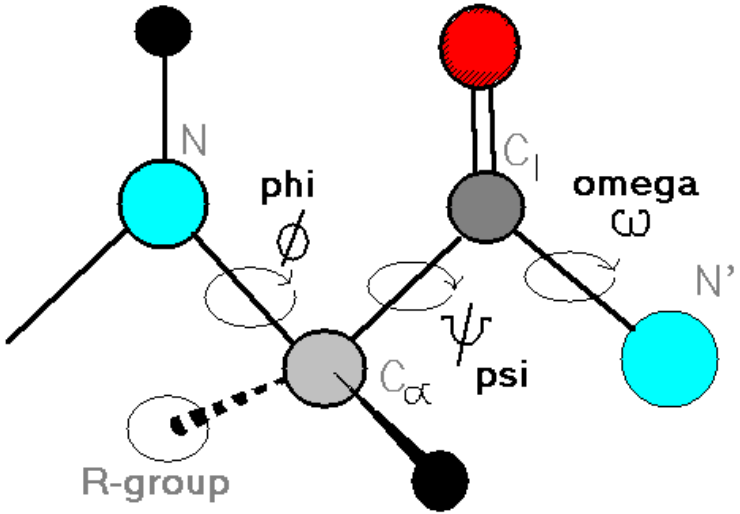


https://courses.lumenlearning.com/hccs-waymakerbiology1/chapter/reading-major-enzymes/

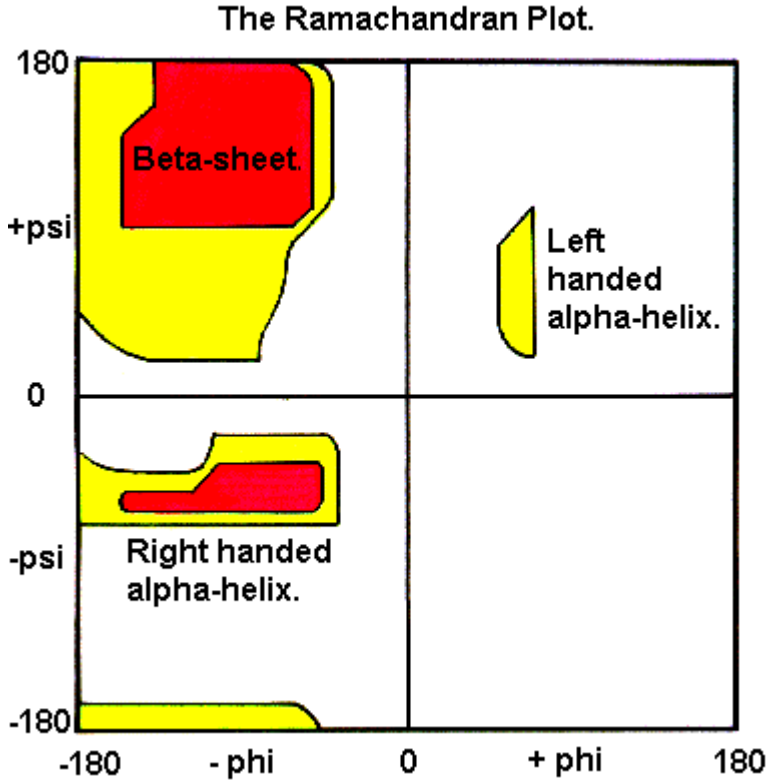


https://lmu.pressbooks.pub/conceptsinbiology/chapter/protein-structure/

Torsion angle in peptide bond



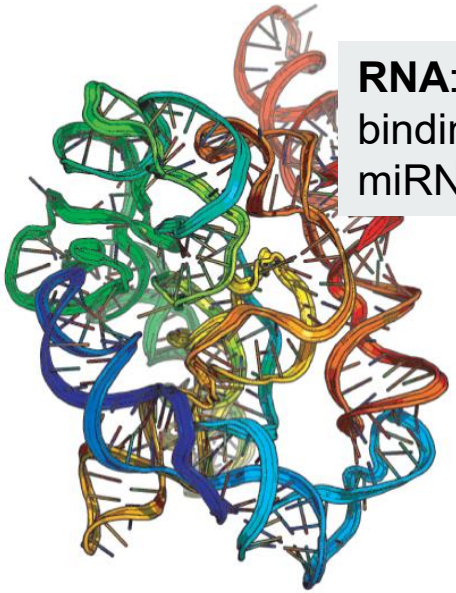
http://www.bioinf.org.uk/teaching/bioc0008/page03.html



http://www.bioinf.org.uk/teaching/bioc0008/page05.html

- Signature of protein secondary structures, use for QC

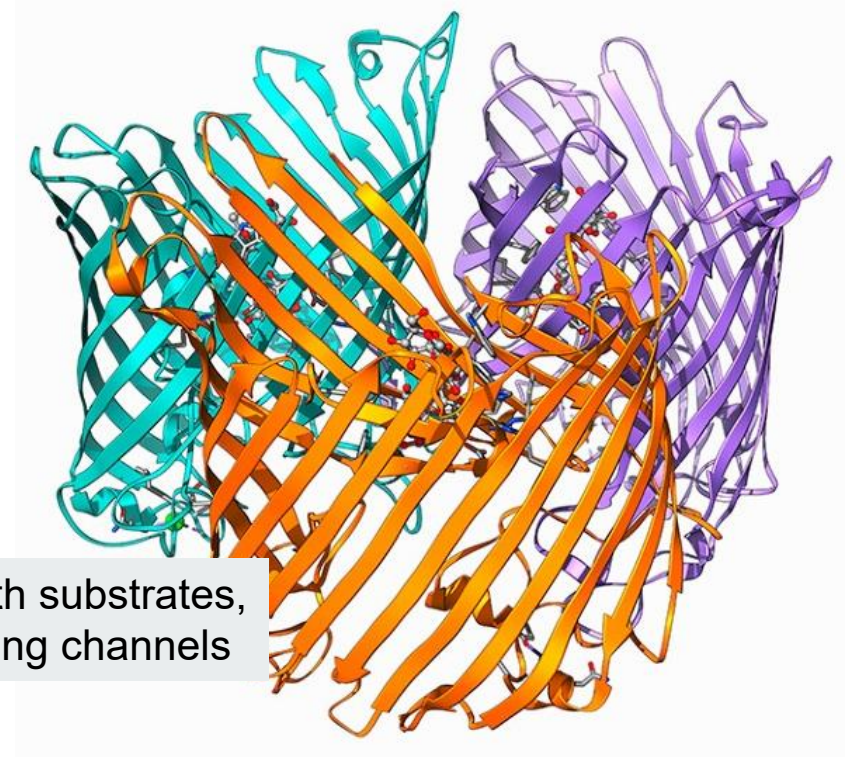
Tertiary and quarternary structures



RNA: scaffold for protein binding, protective against miRNA

Das, R. Nature Methods 18:439 (2021)

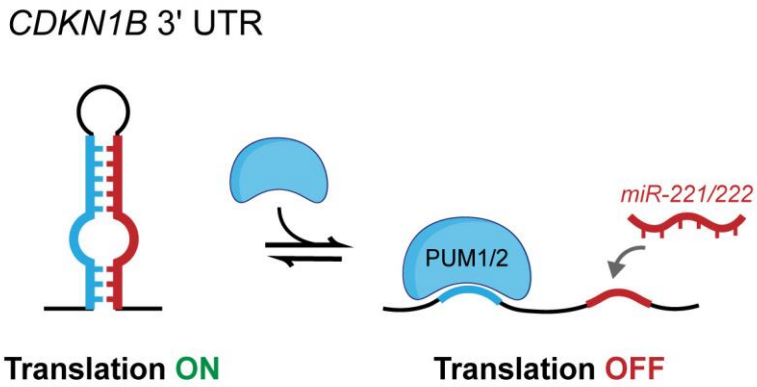
***Tetrahymena* ribozyme**
Discovered: 1980
Structure solved: 2020



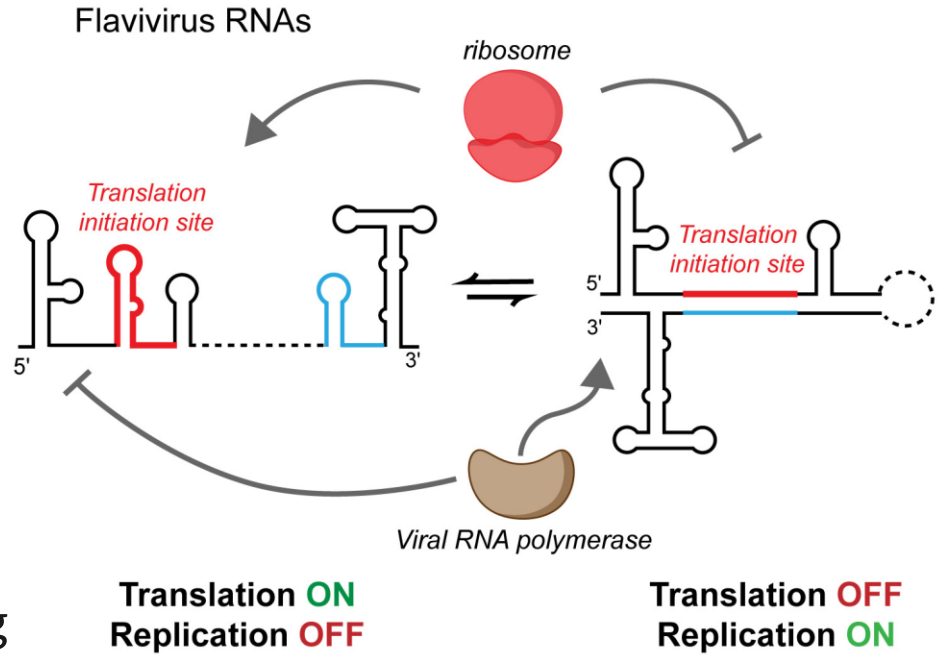
Protein: Binding with substrates, and cofactors, forming channels

Callaway, E. Nature 621:455 (2023)

Regulatory roles of RNA secondary structure



Bose, R. et al. Cell Chemical Biology 31:17-35 (2024)



- Secondary structure protects against hybridization and binding by other RNA and proteins

Functional roles of RNA tertiary structure

Scaffold for recruiting proteins

Guide proteins to target RNA/DNA

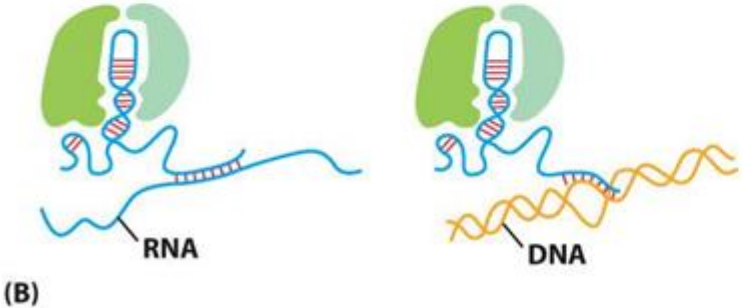
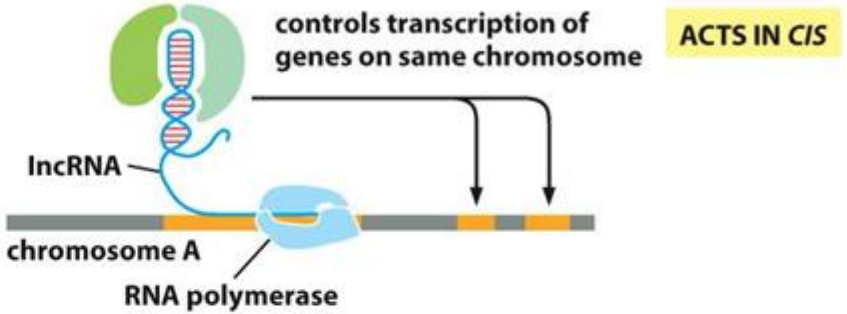


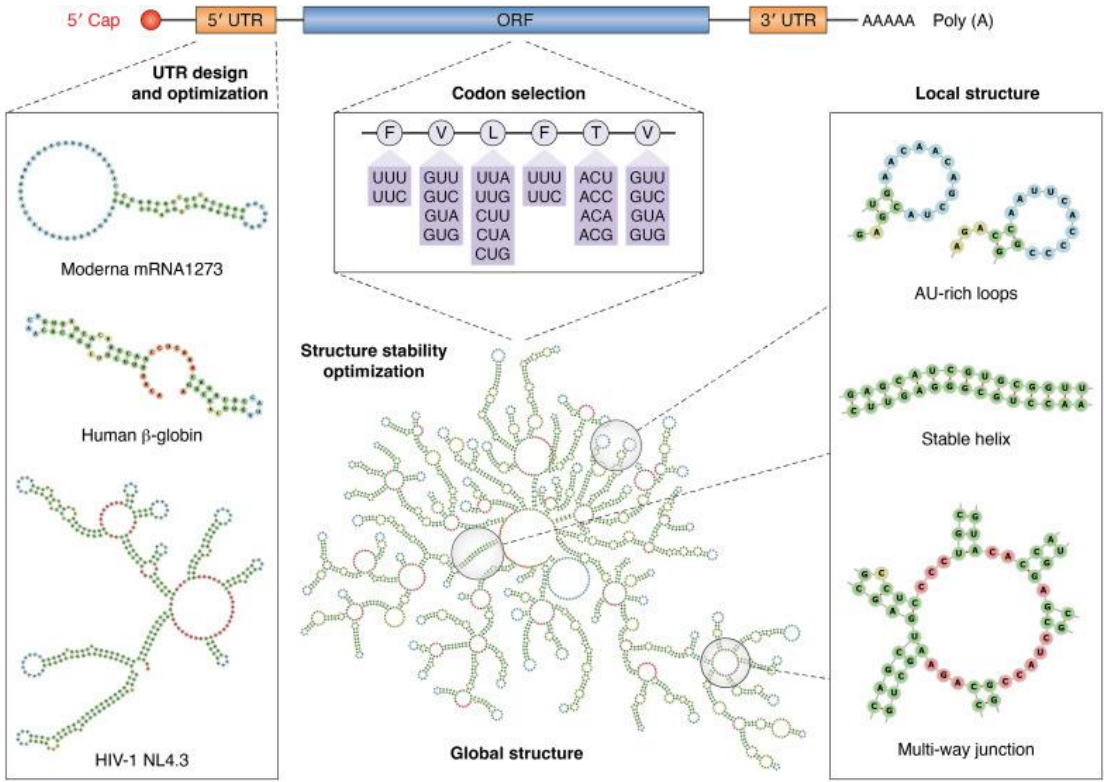
Figure 7-79 Molecular Biology of the Cell 6e (© Garland Science 2015)

Self-regulation



- **Tertiary structure**: binding
- **Primary structure**: base-pairing
- **Self-regulatory**: structures in UTR

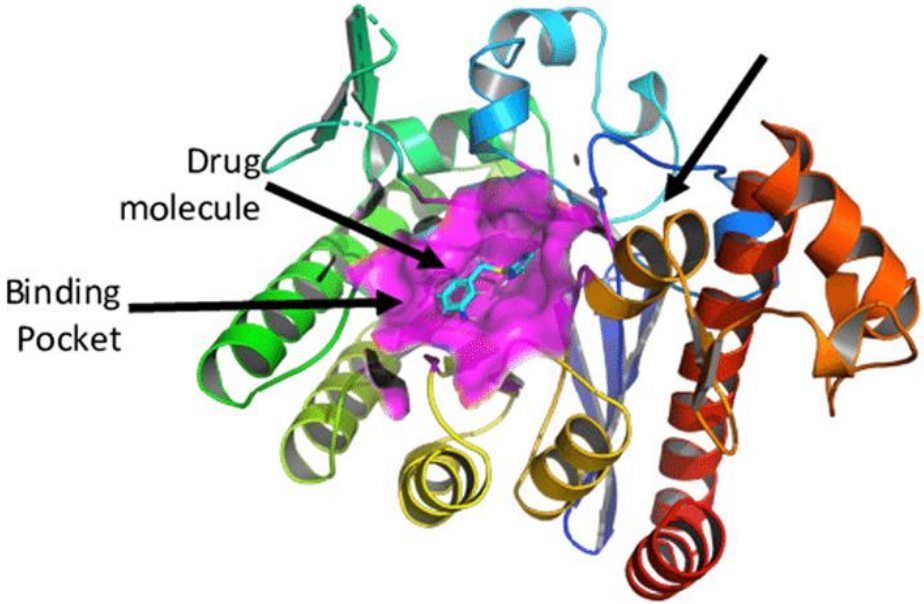
UTR optimization for mRNA vaccine development



- **Global**: structural stability
- **5'UTR**: ribosome loading efficiency
- Learn from highly expressed genes and viral UTRs

Jin, L. et al. Journal of Biological Chemistry 301:108015 (2025)

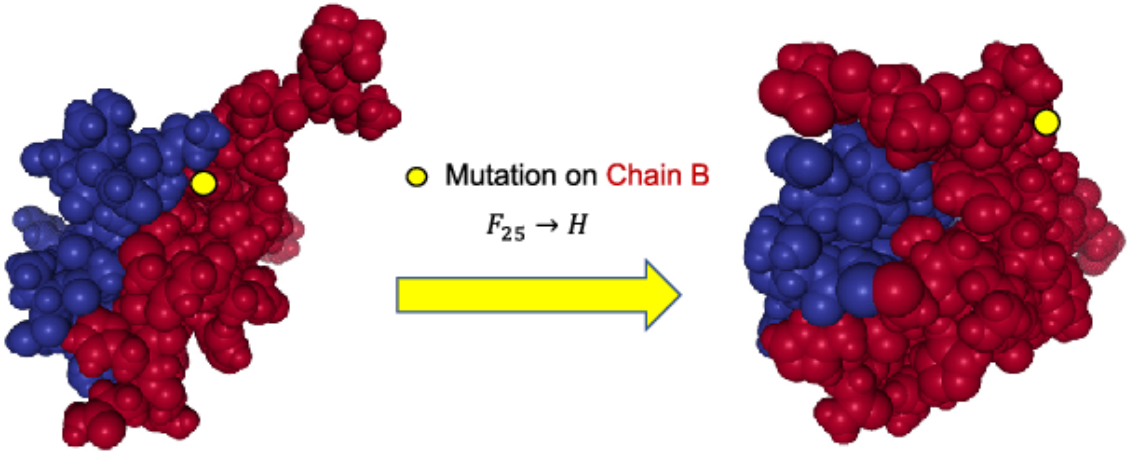
What can we do with knowledge of structure?



- Visualize
- Predict binding/interaction with other molecules
- Hypothesize effect of mutations on the structure and interaction

Jones, D. et al. Journal of Chemical Information and Modeling 61:1583-1592 (2021)

Correlating structural and functional changes

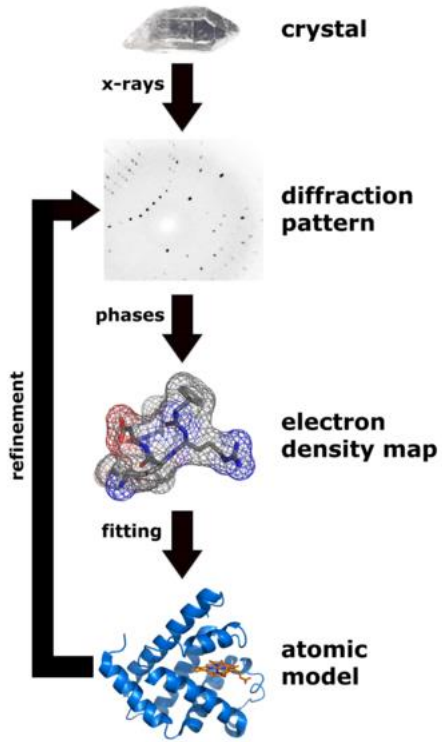


Zhou, G. NAR Genomics and Bioinformatics 2:lqaa015 (2020)

- Structural change → change in binding orientation / dynamics → change in response and downstream functional activity

Assays for RNA and protein structure

X-ray crystallography



- Account for >80% of known protein structures
- Solve for possible electron density from diffraction pattern of x-ray
- Fit most likely amino acids on to electron density
- Median resolution of around 2.0 Å
- Not all proteins can crystalize (not native state)

https://proteopedia.org/wiki/index.php/X-ray_crystallography

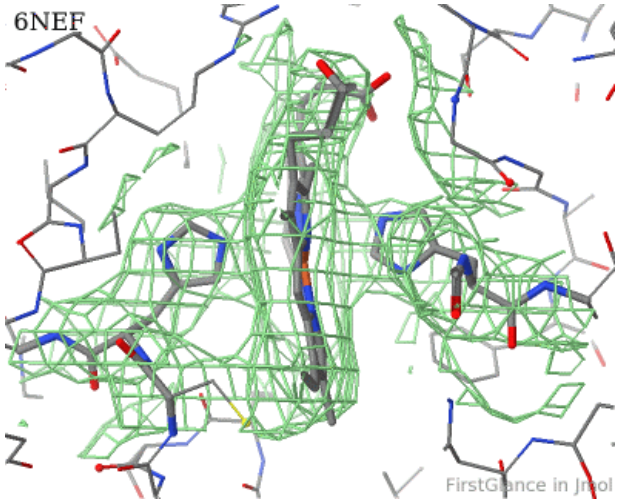
Other biophysical techniques

- **Nuclear magnetic resonance (NMR) spectroscopy**
  - Proteins in solution / mixture of possible structures
  - Work for small molecules
  - ~6% of known protein structures

- **Cryo-EM (electron microscopy)**
  - Protein frozen at 4-77 Kelvin
  - Median resolution of 3.5 Å
  - ~10% of known protein structures

- **Small-angle X-ray scattering (SAXS)**

https://proteopedia.org/wiki/index.php/Cryo-EM



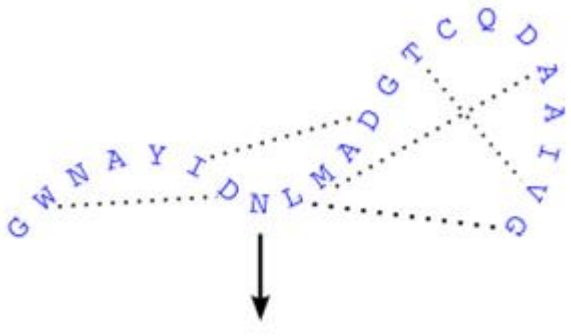
Density map for 3.4 Å cryo-EM of a cytochrome

Distance and structural restraints

- Partial structural information help choosing the correct protein structures among possibilities

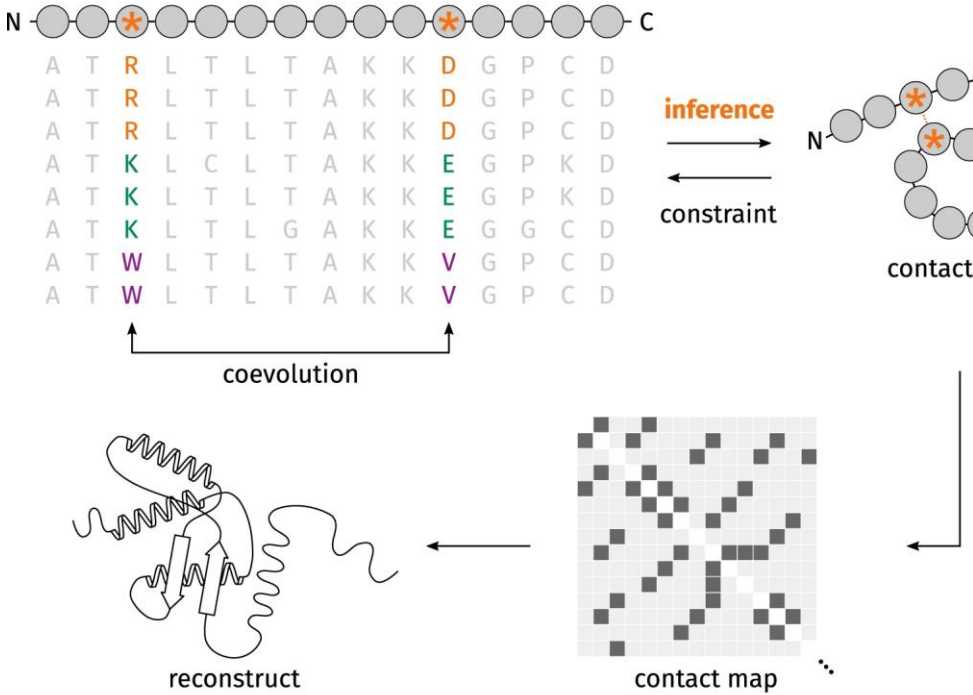
- **Structural mass spectrometry**
  - Chemical cross-linking
  - Hydrogen-deuterium exchange

- **Computational analysis**
  - Sequence alignment and evolution
  - Mutational scanning experiment



https://salilab.org/modeller/manual/node11.html

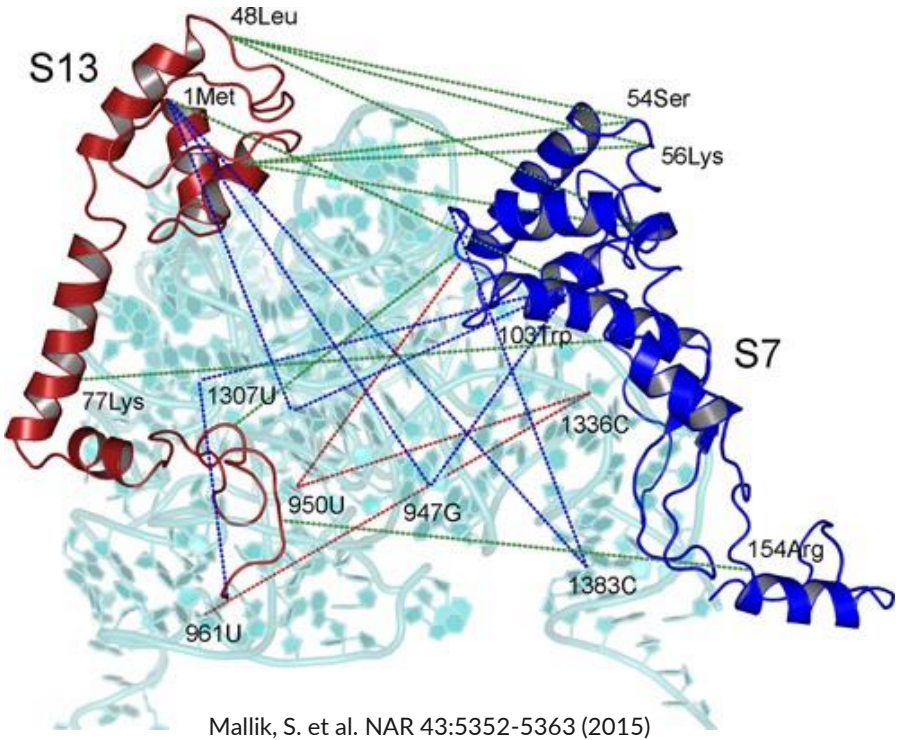
Co-evolution



Bittrich *et al*. Sci Rep 2019

- Structurally linked residues tend to evolve together to maintain the interactions
- Discover by calculating **mutual information**
  - Deviation of statistical independence
  - Compare **P(x, y)** to **P(x)P(y)**

Co-evolution across proteins

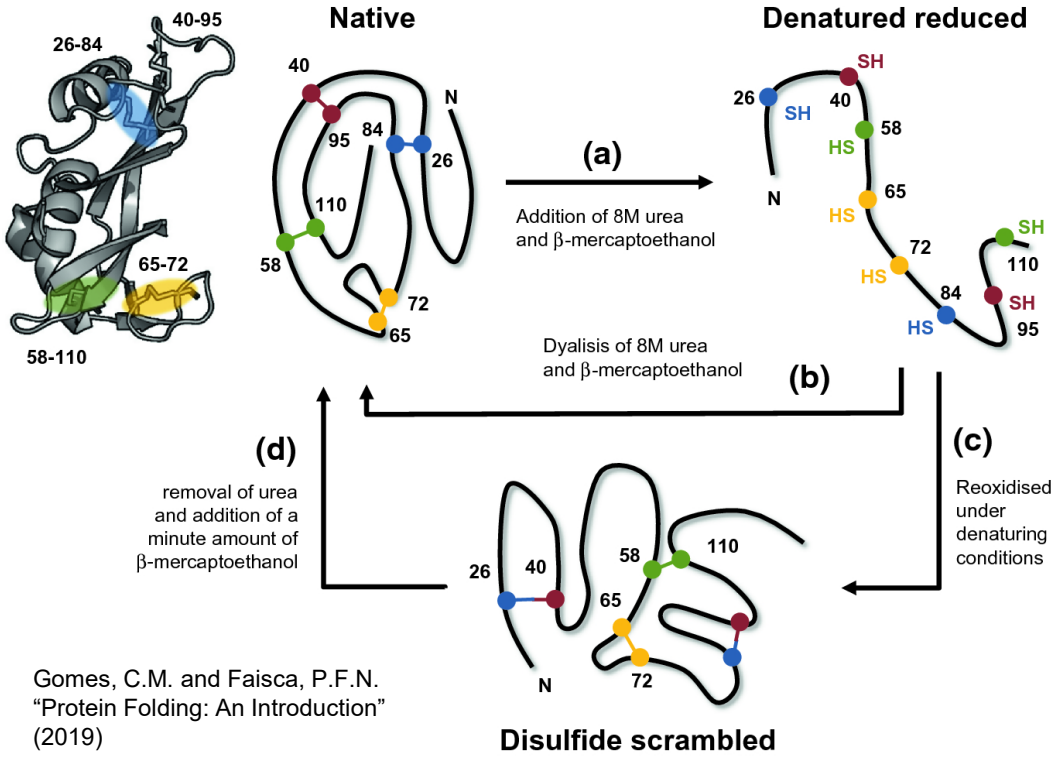


Mallik, S. et al. NAR 43:5352-5363 (2015)

- Between direct interaction partners
  - Compensatory mutation to maintain binding
- Between proteins sharing a common target
  - Maintain structural compatibility with the common target

Modeling of protein structure

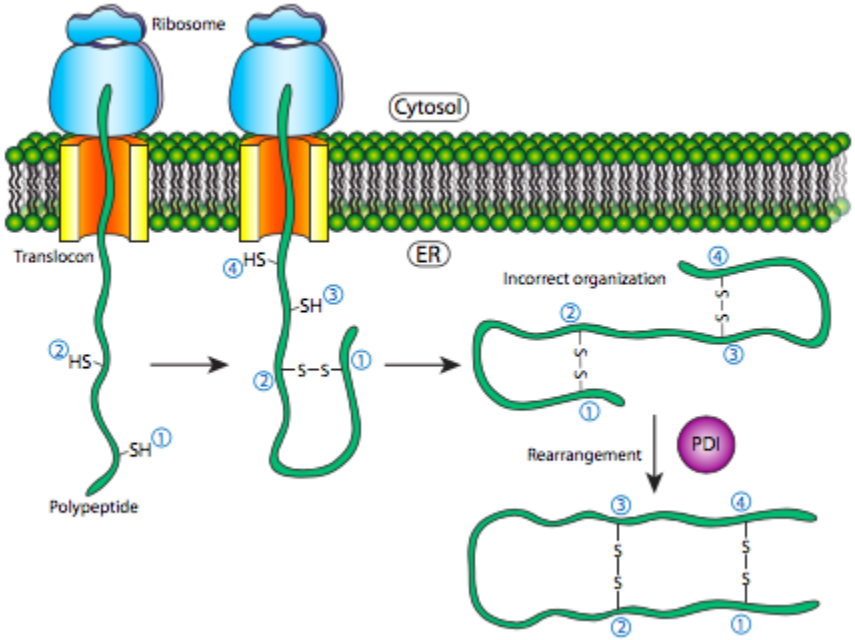
Anfinsen's Dogma

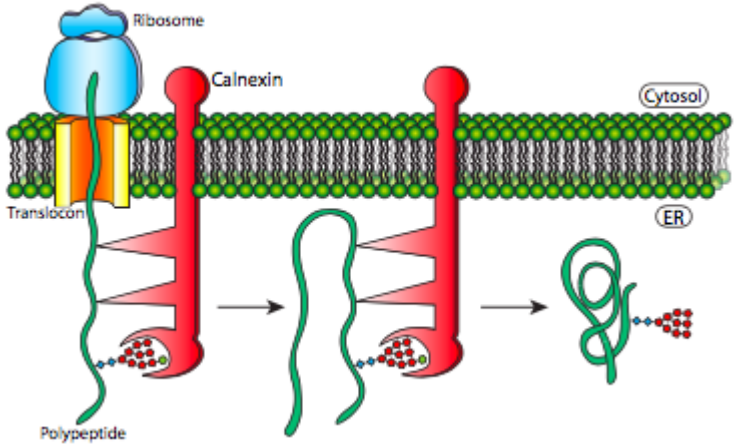


Gomes, C.M. and Faisca, P.F.N. "Protein Folding: An Introduction" (2019)

- Christian B. Anfinsen's experiment on protein folding and re-folding
- **Protein structure is deterministic** – subject to environmental condition
- **Caveat**: Cellular conditions can change

Protein folding inside the cell

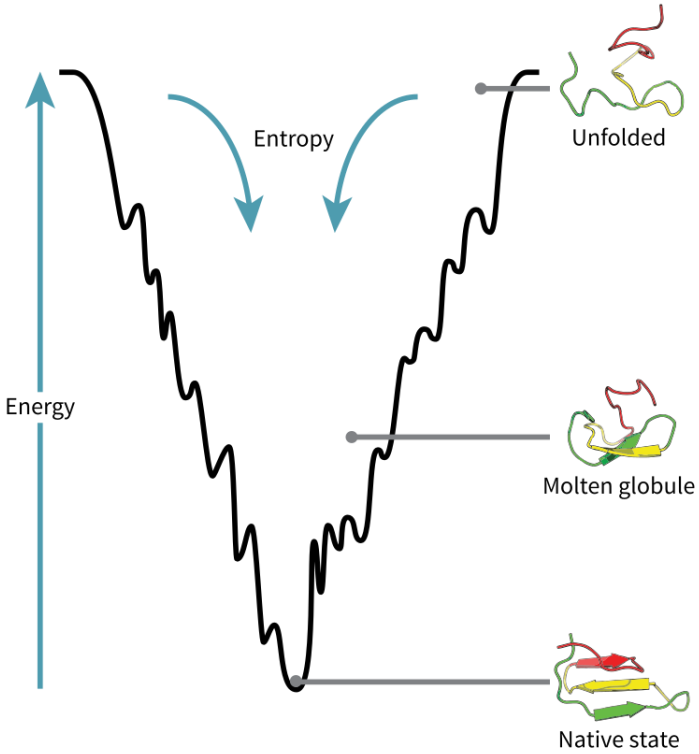




- Fold in parallel with translation
- Transported to organelle
- Aided by chaperone

https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book

Primary forces driving protein folding

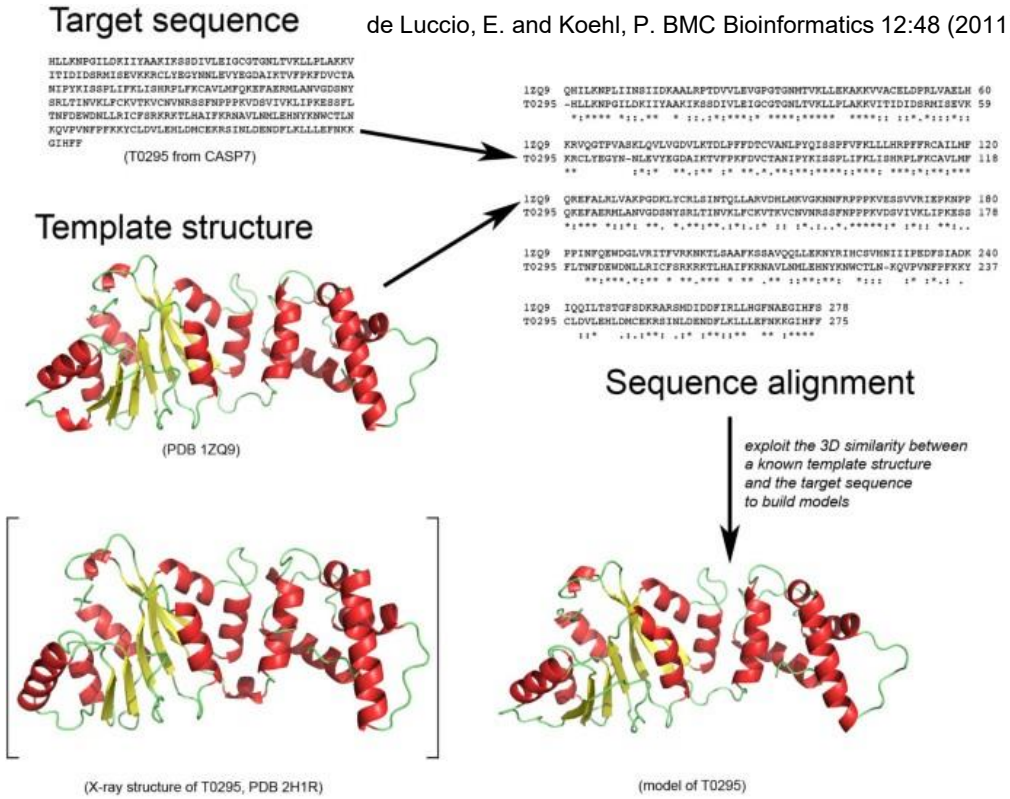


https://en.wikipedia.org/wiki/Protein_folding

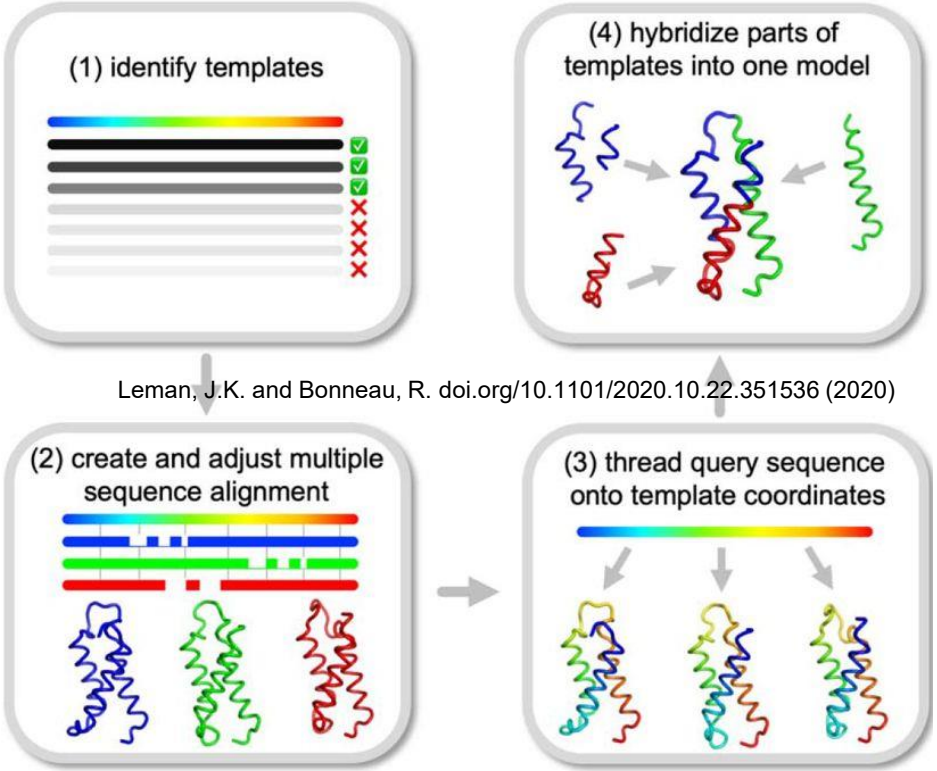
- **Global**: Hydrophobic force guides early stage of folding
- **Local**: Hydrogen, ionic, disulfide, and Van der Waals are responsible for maintaining local structures
- Affected by pH and presence of other solvents inside the cell

Homology modeling

- Identify proteins with similar amino acid sequences and known structure
- Optimize template structure to fit the new amino acids
- Can utilize multiple templates to generate possibilities



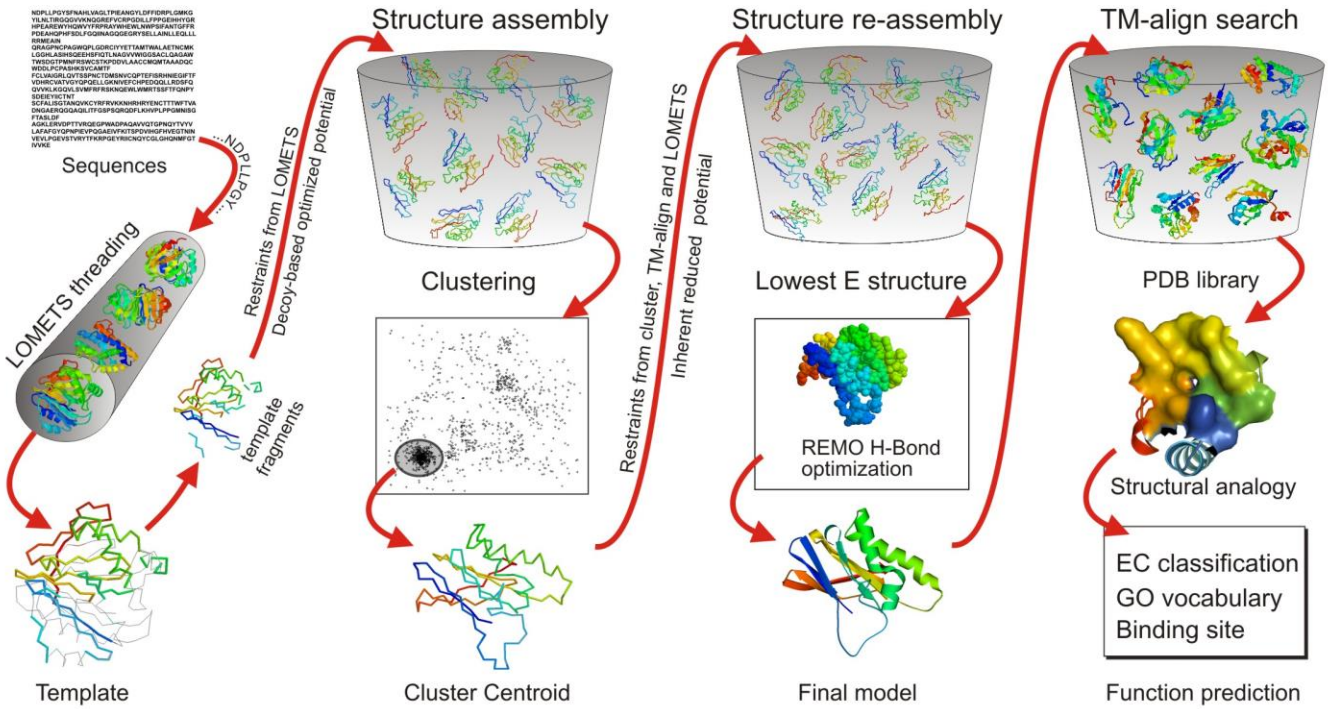
Protein threading (sequence-to-structure)



Leman, J.K. and Bonneau, R. doi.org/10.1101/2020.10.22.351536 (2020)

- Allow different parts of the protein to align with different protein families
  - New protein may contain common domains
- Stitch matches into a new contiguous template structure
- Optimize the structure

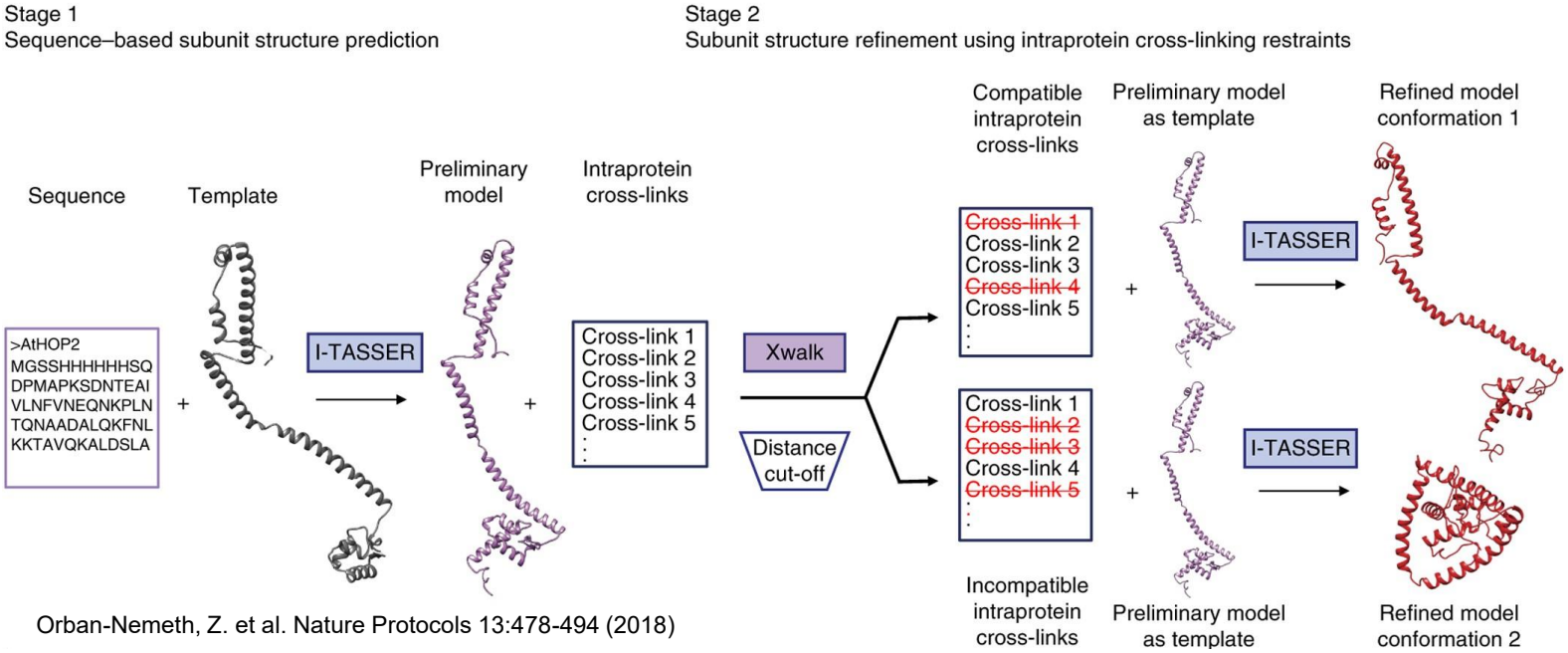
A pipeline for *de novo* protein structure prediction



- Start with protein threading
- Generate many structures with low energy
- Cluster and remove outliers
- Repeat the process

https://zhanggroup.org/I-TASSER/

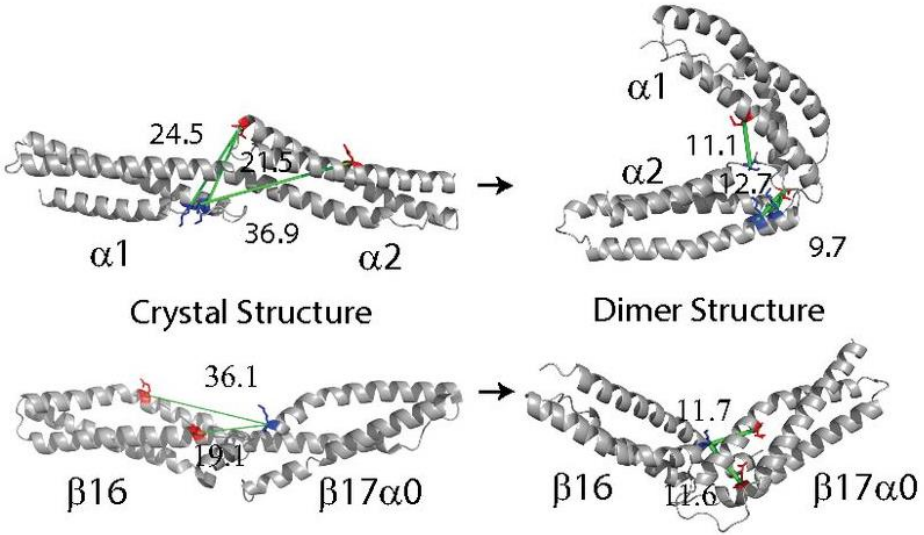
Distance restraints with *de novo* prediction



Orban-Nemeth, Z. et al. Nature Protocols 13:478-494 (2018)

- Use distance restrains to rule-out and refine candidate models

Discrepancy between crystal structure and native form

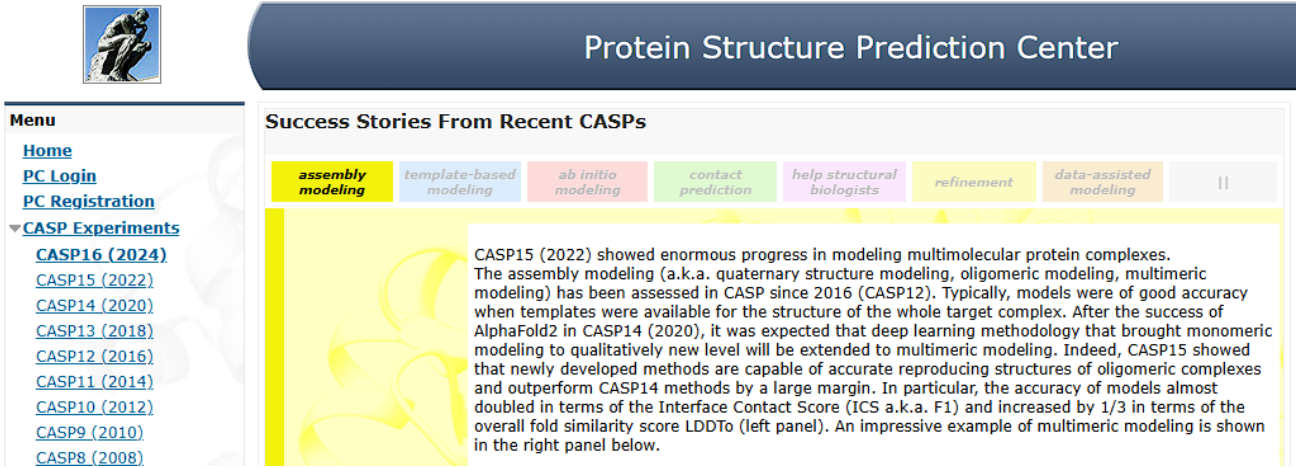


Sriswasdi, S. et al. PNAS 111:1801-1806 (2014)

- Distance restraints determined from chemical cross-linking
  - Expected to occur <12 Å
- Crystal structures show >24 Å
- Update the structure with homology modeling to demonstrate flexibility of native protein conformations

Emergence of AI for protein folding

Critical Assessment of Protein Structure Prediction

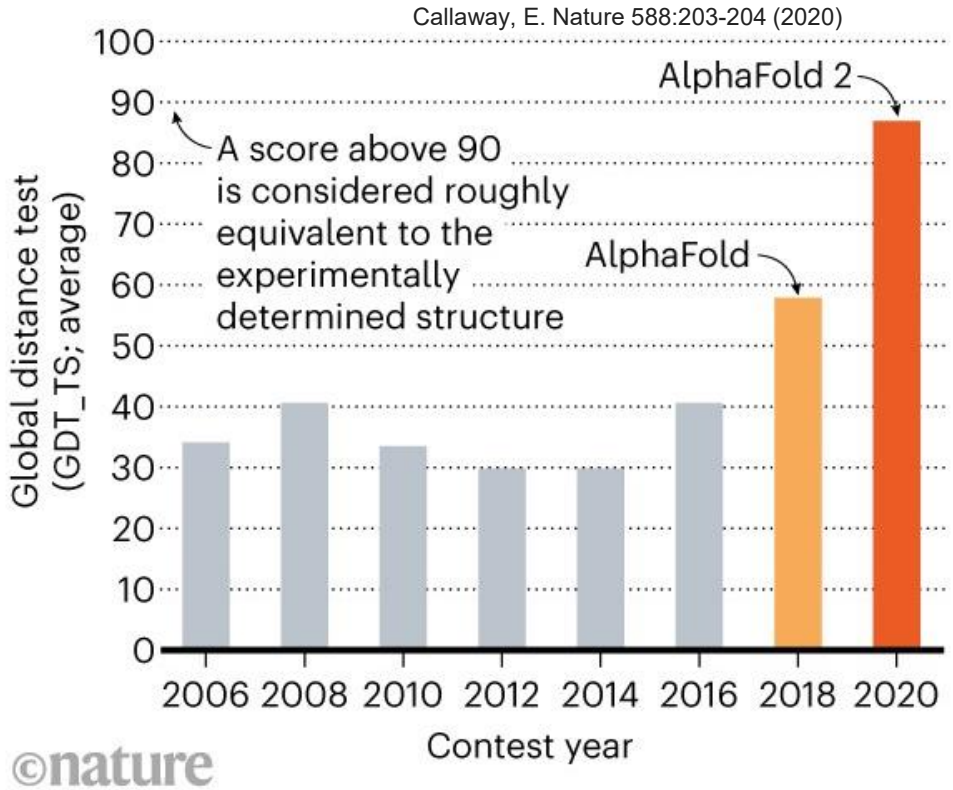


- International community promoting research in protein folding prediction
  - Held every 2 years
  - Many tasks: long-range contact, homology, *de novo*, oligomer
  - Withhold experiment data for evaluating algorithm

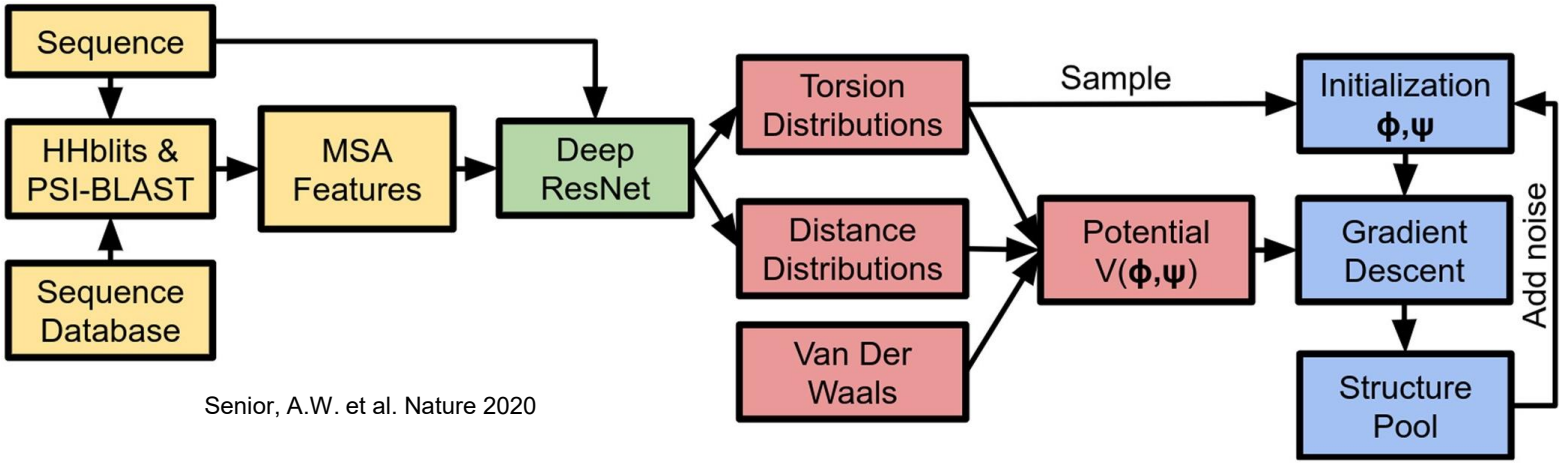
AlphaFold's breakthrough in protein folding prediction

- AF1 **predicts co-evolutionary signal as distance restraints** within the structure and use it to predict 3D structure

- AF2 was able to **directly predict 3D structure** with as low uncertainty as crystal structure (~2 Å)



Design of AlphaFold version 1



Senior, A.W. et al. Nature 2020

- Perform multiple sequence alignment
- Predict co-evolution → distance restraints
- Predict torsion angle → additional constraints
- Optimize 3D structure using physics-based equations

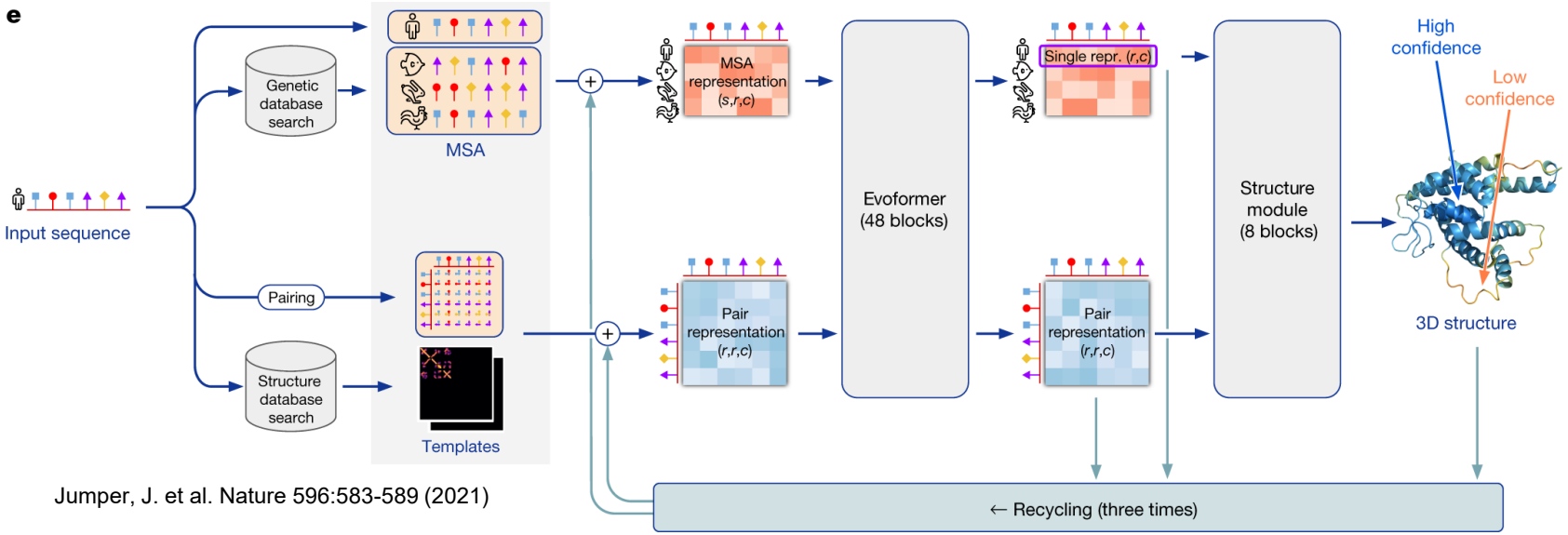
Lesson learned from AlphaFold version 1

Potential	Bins	TM-score	GDT_TS	lDDT	RMSD (Å)	$-\log_{10} P$
Full + relax	51/64	0.649	65.8	54.2	5.94	7.3
Full	51/64	0.642	65.0	53.9	5.91	–
W/o reference	51/64	0.632	64.3	50.0	6.64	4.0
W/o `score2_smooth`	51/64	0.641	64.8	53.7	5.93	1.2
W/o torsions	51/64	0.637	64.3	53.6	6.04	8.2
W/o distogram	51/64	0.266	29.1	19.1	14.88	130

Senior, A.W. et al. Nature 2020

- The ability to learn co-evolution from multiple sequence alignment is the single most important factor for predicting accurate 3D structure

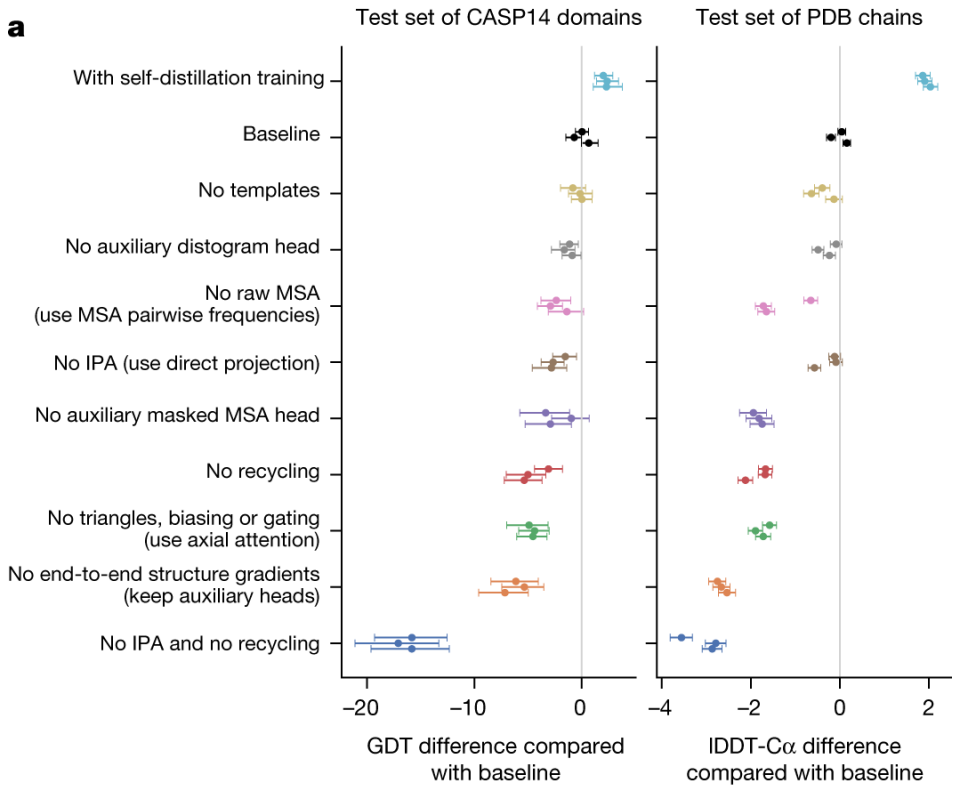
Design of AlphaFold version 2



Jumper, J. et al. Nature 596:583-589 (2021)

- Iterative structure prediction and refinement
- Maintaining consistency across MSA, distance restraint, and structure

Lesson learned from AlphaFold version 2



Important factors

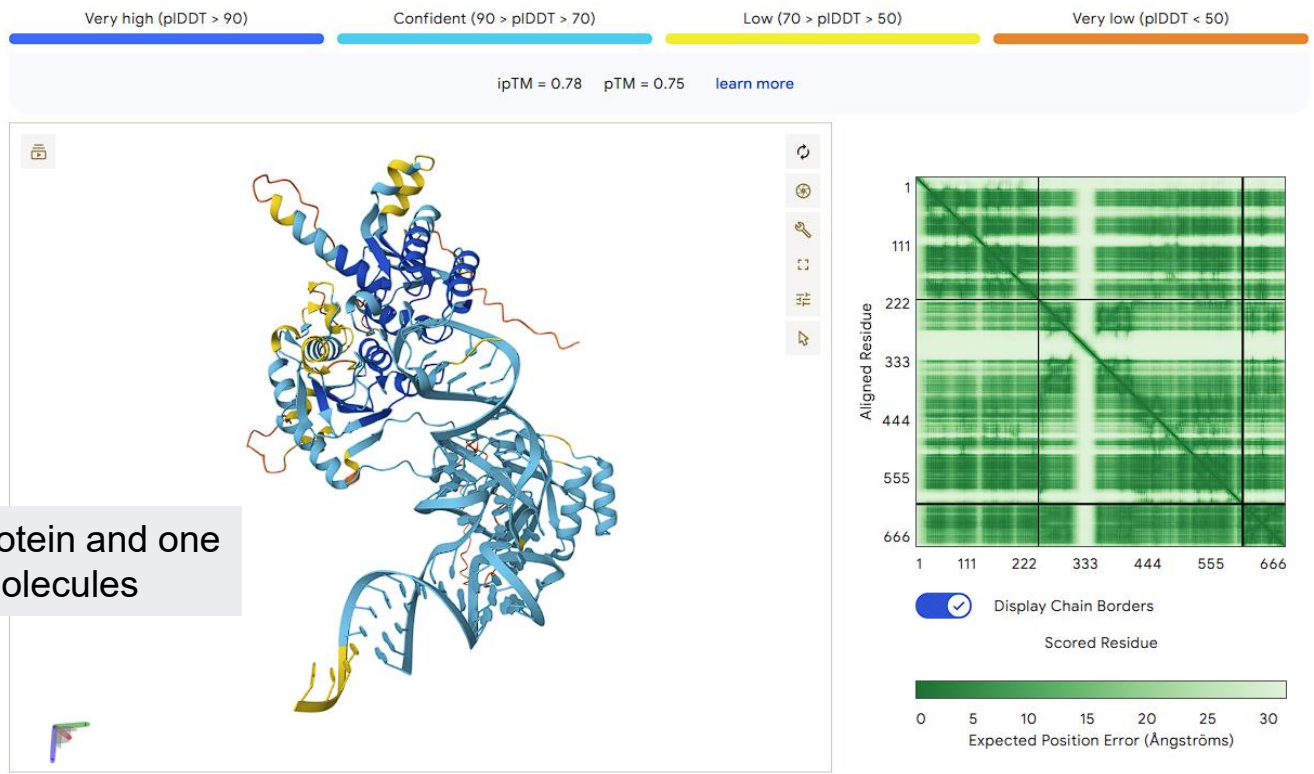
- Iterative structural prediction
- Geometric constraint on structure
- Large protein sequence database

Unimportant factors

- Template structures
- Explicit prediction of distance restraints
- Full MSA data (AA frequency is sufficient)

Jumper, J. et al. Nature 596:583-589 (2021)

Multimer prediction with AlphaFold version 3



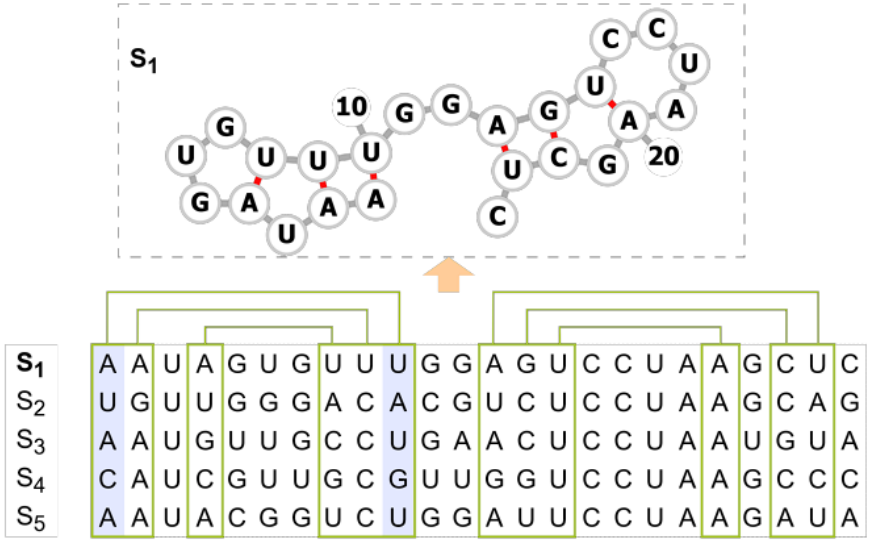
Two protein and one RNA molecules

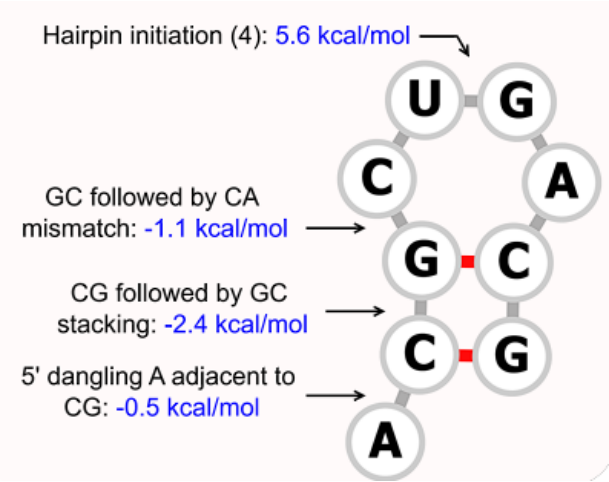
Current limitations of protein folding prediction

- Trained on only crystal structures
  - Not completely representative of native conformations
- One structure per protein / complex
  - Protein structures are dynamics, can change conformation
- No context information
  - Cellular conditions

Computational modeling of RNA structure

RNA secondary structure prediction



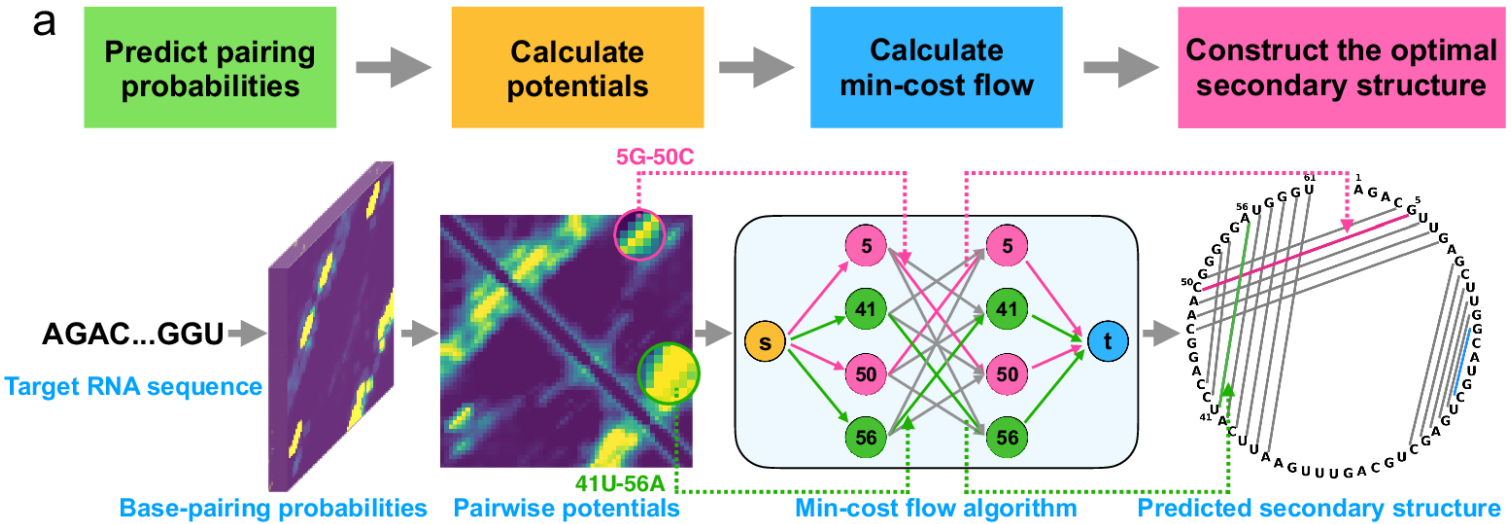


Yang, S. et al. https://arxiv.org/pdf/2501.04056 (2025)

- Co-evolutionary pattern of complementary bases
- Free energy estimate based on neighboring bases ← dynamic programming

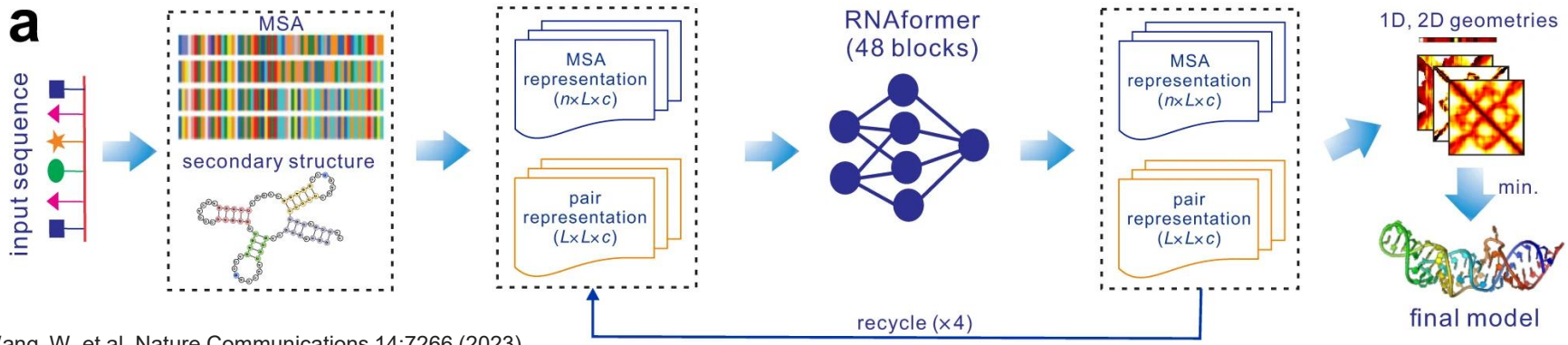
RNA secondary structure prediction

Gong, T. et al. Communications Biology 7:297 (2024)



- Predict base-pairing probabilities for RNA
- Describe RNA secondary structure as a network
  - Edges connecting adjacent nucleotides and base-paired nucleotides

RNA 3d structure prediction

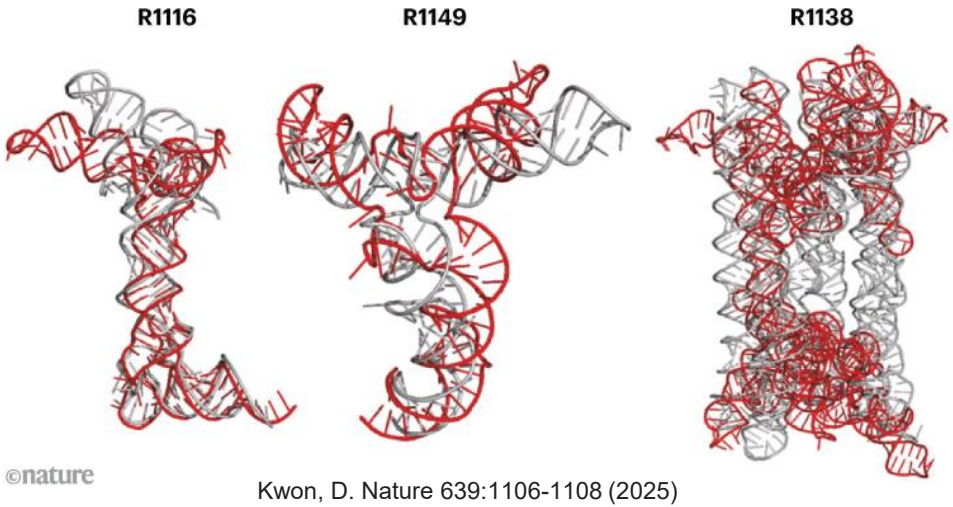


Wang, W. et al. Nature Communications 14:7266 (2023)

- **Pre 2015**: template-based, rely on secondary structure
- **Post 2025**: mimic AI for protein folding prediction

Challenges in prediction of RNA 3D structure

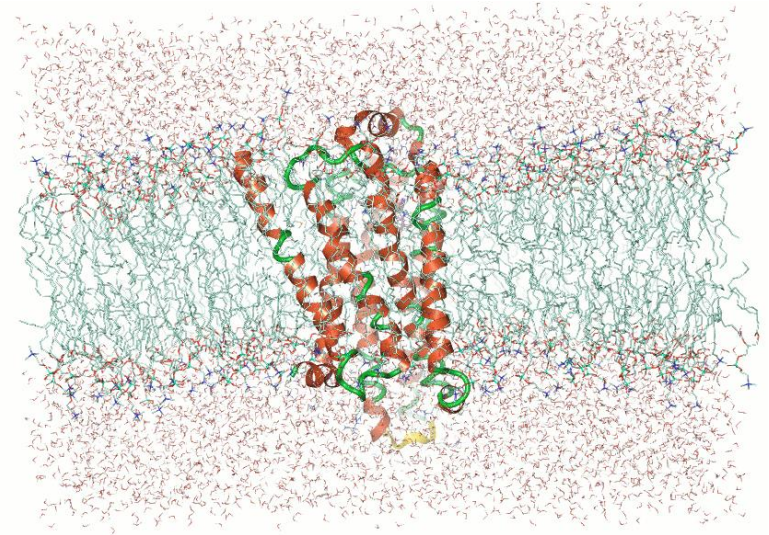
- Few known structures (6,000 compared to >200,000 for proteins)
- Identifying similar RNA via sequence alignment is difficult
- Highly flexibility
- Free energy composition is less well-understood



Kwon, D. Nature 639:1106-1108 (2025)

Molecular dynamics simulation

Extend knowledge from static structures



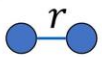
https://cresset-group.com/science/science-resources/membrane-protein-molecular-dynamics/

- Structural flexibility / uncertainty
- Ligand binding stability
- Conformational changes
- Protein folding / unfolding process
- Solvent-accessible surface

Atomistic simulation

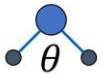
$$U_{total} = \underbrace{U_{bond} + U_{angle} + U_{Pdihe} + U_{Idihe}}_{\text{Bonded interactions}} + \underbrace{U_{LJ} + U_{PC}}_{\text{Non-Bonded interactions}}$$

Bond potential: Spring



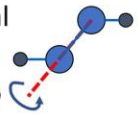
U_{bond}

Angle potential: Hinge



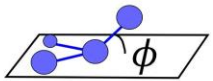
U_{angle}

Proper dihedral potential: Axel



U_{Pdihe}

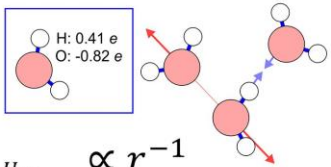
Improper dihedral potential



U_{Idihe}

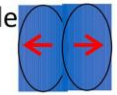
Electrostatic potential

Bonding + Electronegativity: Partial charges on atoms



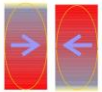
$U_{PC} \propto r^{-1}$

Repulsive Pauli Exclusion principle

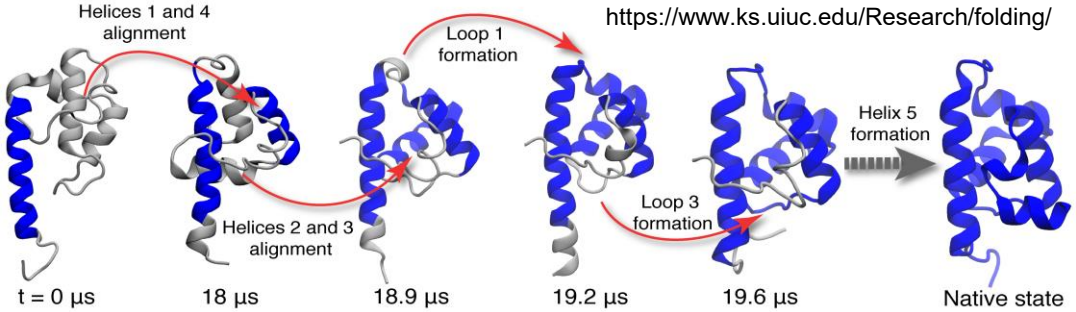


$U_{LJ} \propto r^{-12}$

Attractive Van-der-Waals



$U_{LJ} \propto r^{-6}$



https://www.ks.uiuc.edu/Research/folding/

- Picosecond-nanosecond scale movement of atoms based on physical forces
- Time-series data

Bunker, A. and Rog, T. Frontiers in Molecular Biosciences 7:604770 (2020)

Summary

- Structure determines function of biomolecules
  - Chromatin, RNA, and proteins
- Molecular assays for solving 3D structure
- Emergence of AI for protein folding prediction
- Structure is dynamics and depends on context (cellular states)

Any question?

- See you next time