Bioinformatics D: Protein Structure and Function

DAVID L. TABB, PH.D.

FEBRUARY 28, 2019

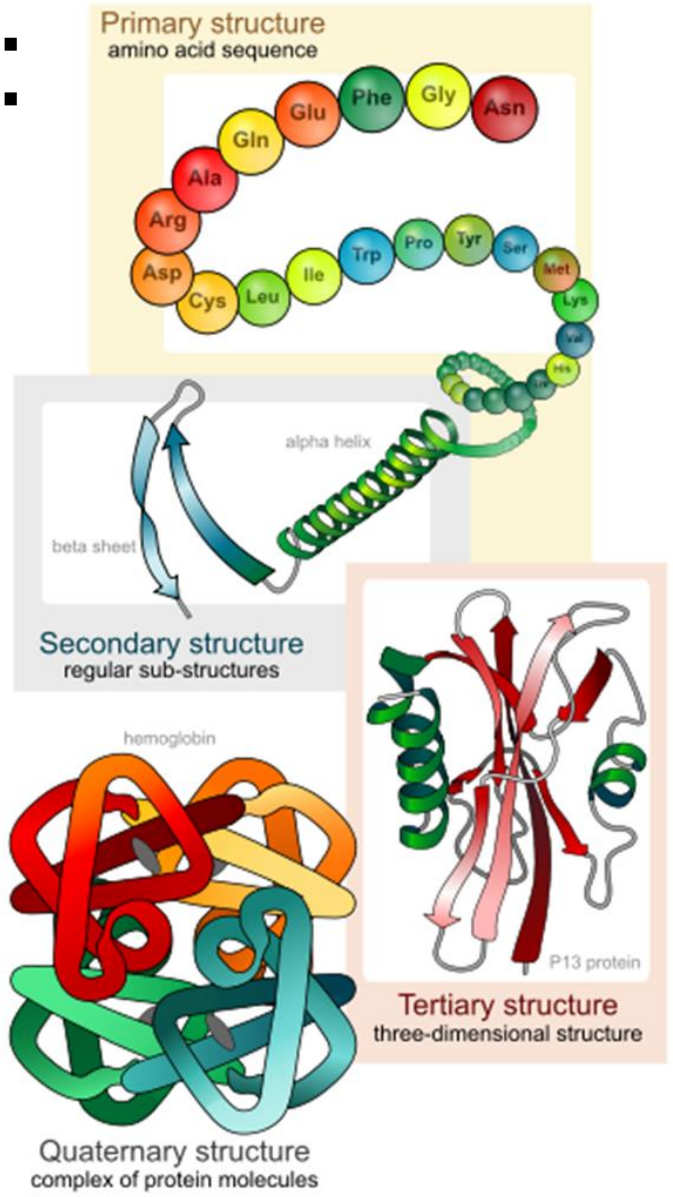
Overview

- Structural basis of function
  - Tanford transition: beta-lactoglobulin has a pocket
  - Catalytic triad: serine proteases chop peptide bonds
- Protein structure
  - Experimental determination of structures
  - Phi and Psi angles; Levinthal's Paradox
  - *De novo* prediction of protein structure
  - PDB and CASP

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Four levels of protein structure:

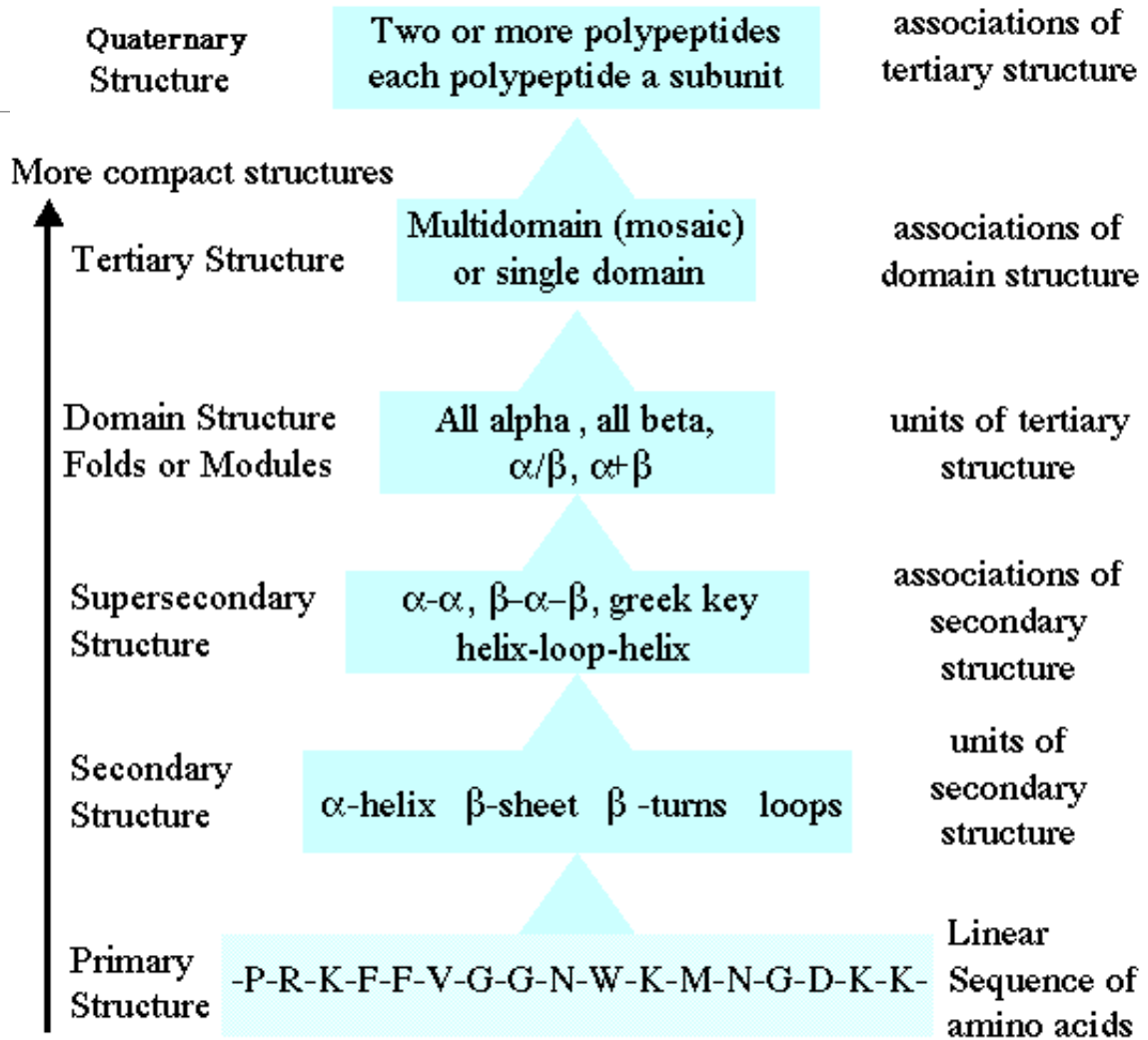
- Primary Structure
  - Amino acid sequence
- Secondary Structure
  - Alpha helices
  - Beta Sheets
- Tertiary Structure
  - Fully folded polypeptide chain
- Quaternary Structure
  - Multichain proteins
  - Protein-Protein complexes



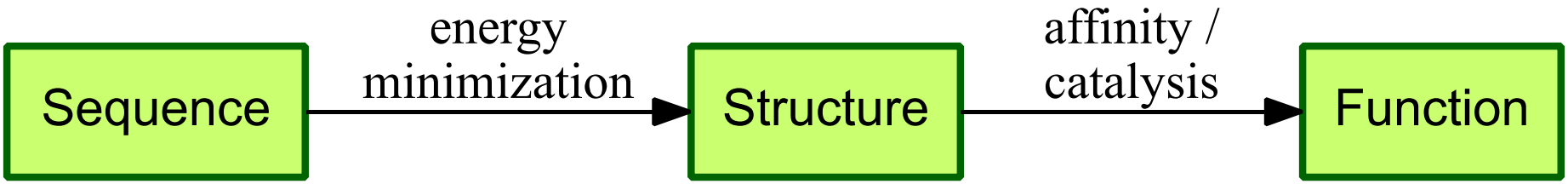
D. Eisenberg. *PNAS* (2003) 100: 11207-11210.

UNIVERSITEIT STELLENBOSCH UNIVERSITY

An update to structure hierarchy



https://biomedapps.curtin.edu.au/biochem/tutorials/prottute/hierarchy.htm



What value are structures?

- If you know the structure, you have strong arguments for function.
- Detailed structures define protein-protein interfaces, clarifying interactions among proteins.
- Structures help us understand the impacts of sequence variability.
- Ligand interfaces support drug discovery and molecular modelling.

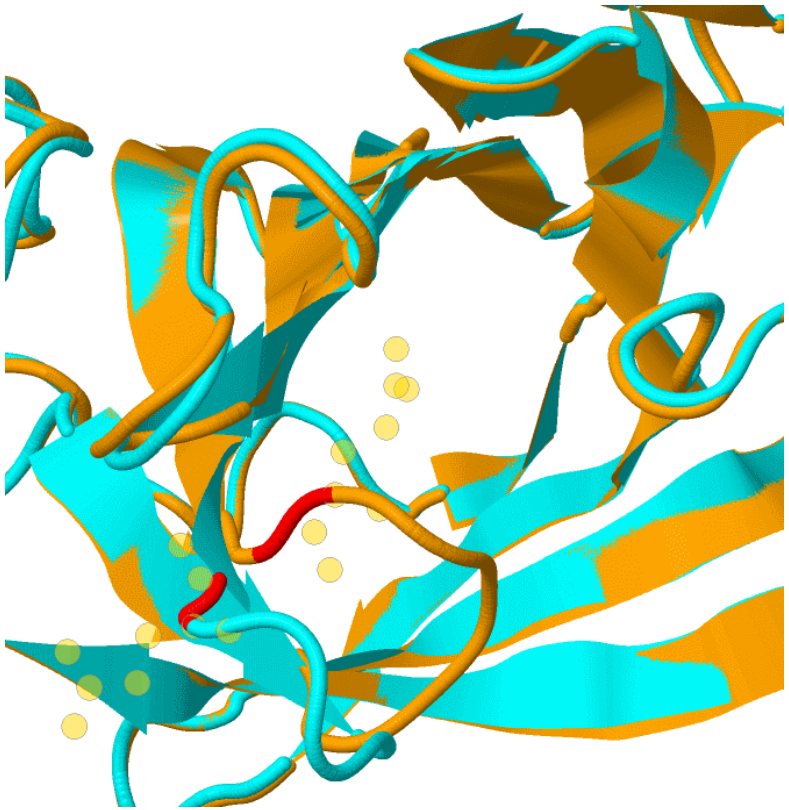
UNIVERSITEIT STELLENBOSCH UNIVERSITY

Beta-lactoglobulin structure confers lipocalin activity

- *Homodimer*: two polypeptides of the same sequence form a duplex.
- *Beta sheet*: *antiparallel* strands form binding pocket for small hydrophobics.

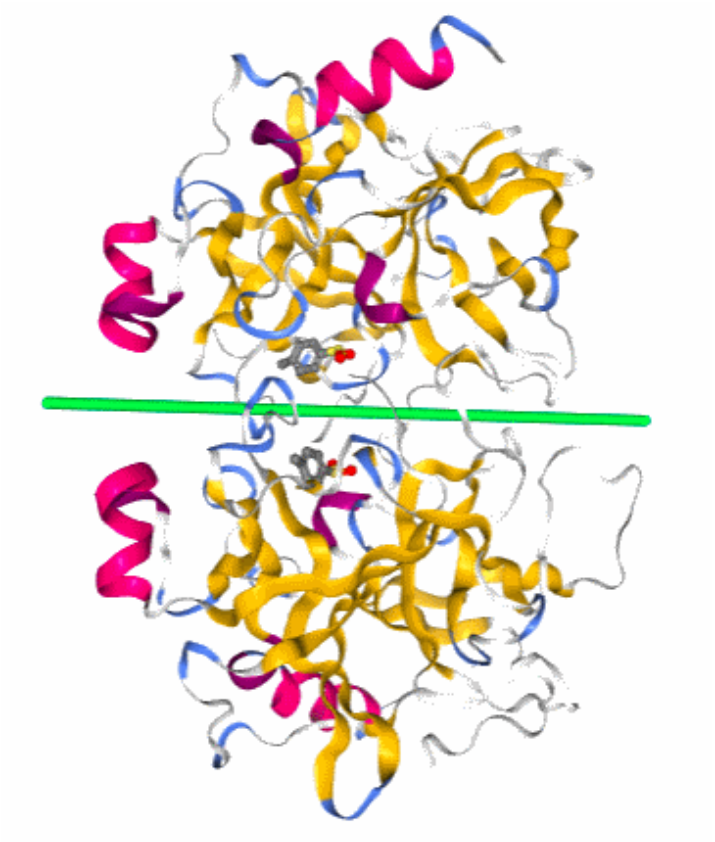
PDB 1BSY; BY Qin et al. *Biochemistry* (1998) 37: 14014-14023.

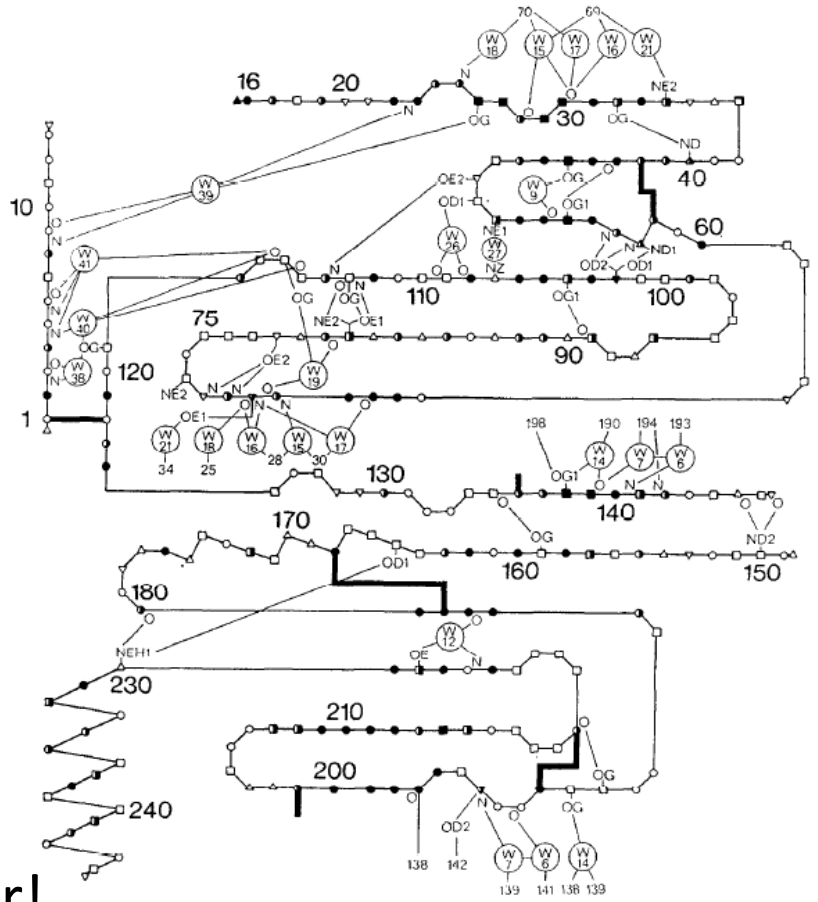
Tanford Transition



- Image compares pH 6.2 structure (orange) to pH 8.2 (blue). "EF-Loop" swings in pH response.
- Glu89 (red plus dots) is buried at low pH but external at high.
- EF-Loop is pocket "door."

Chymotryptin, the model serine protease





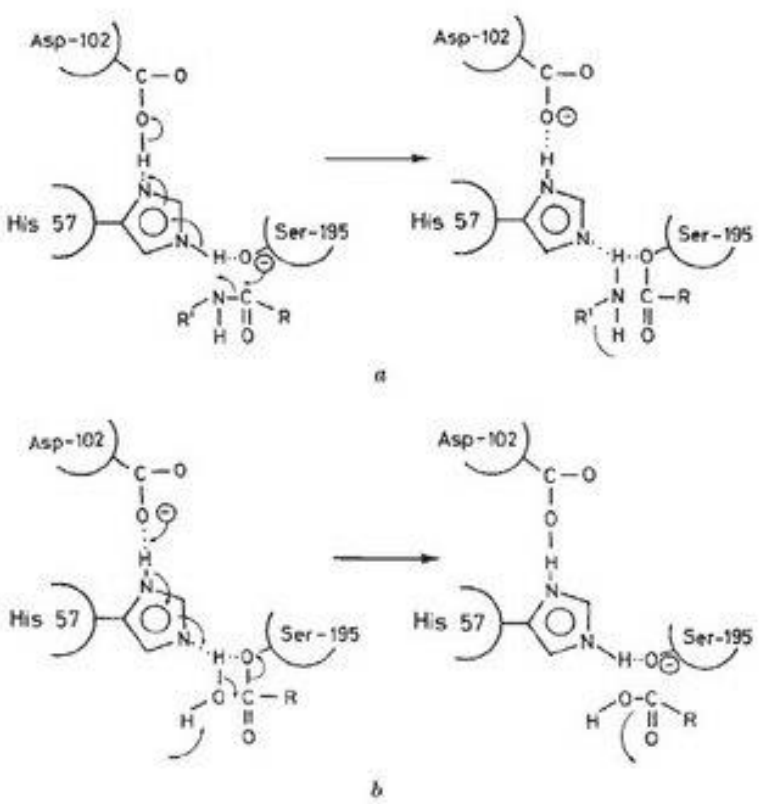
Visualization has gotten easier!

JJ Birktoft and DM Blow. *J. Mol. Biol*. (1972) 68: 187-240.

The catalytic triad

Acylation

Deacylation



1. Ser195 attacks peptide carbonyl, lysing peptide bond.
2. Water's oxygen displaces Ser oxygen to release digestion products.
3. His57 both accepts and donates protons.

Blow et al. *Nature* (1969) 221: 337-340 Neitzel, J. J. *Nature Education (2010)* 3:21

Phi and psi angles define protein backbone shape

- Phi: Rotation of bond between nitrogen and alpha carbon
- Psi: Rotation of bond between alpha carbon and carbonyl.

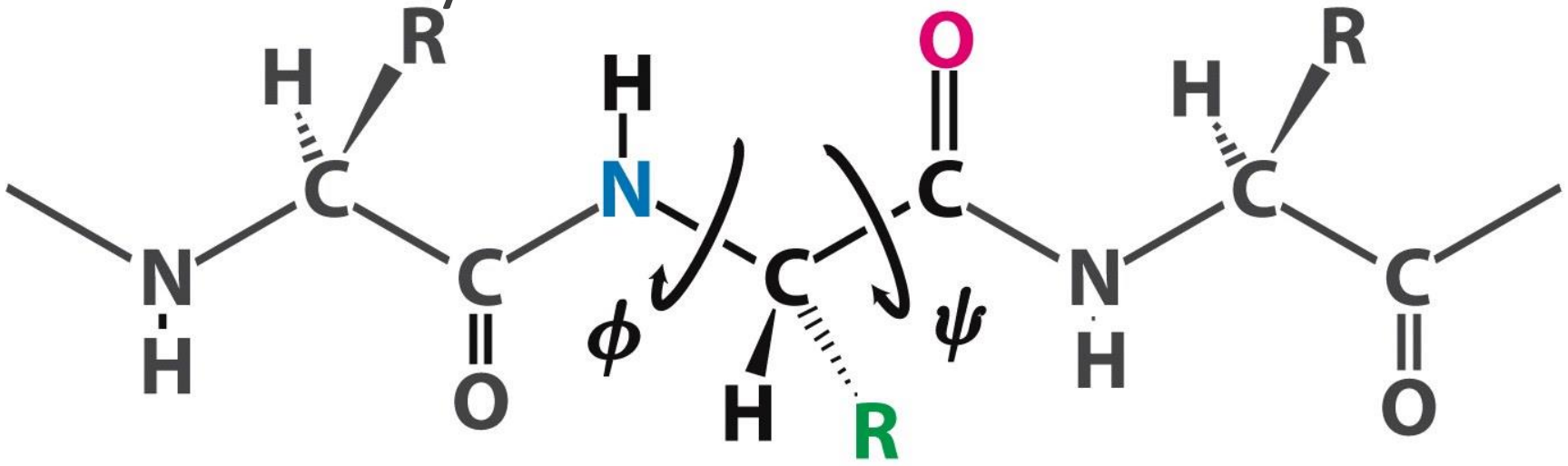


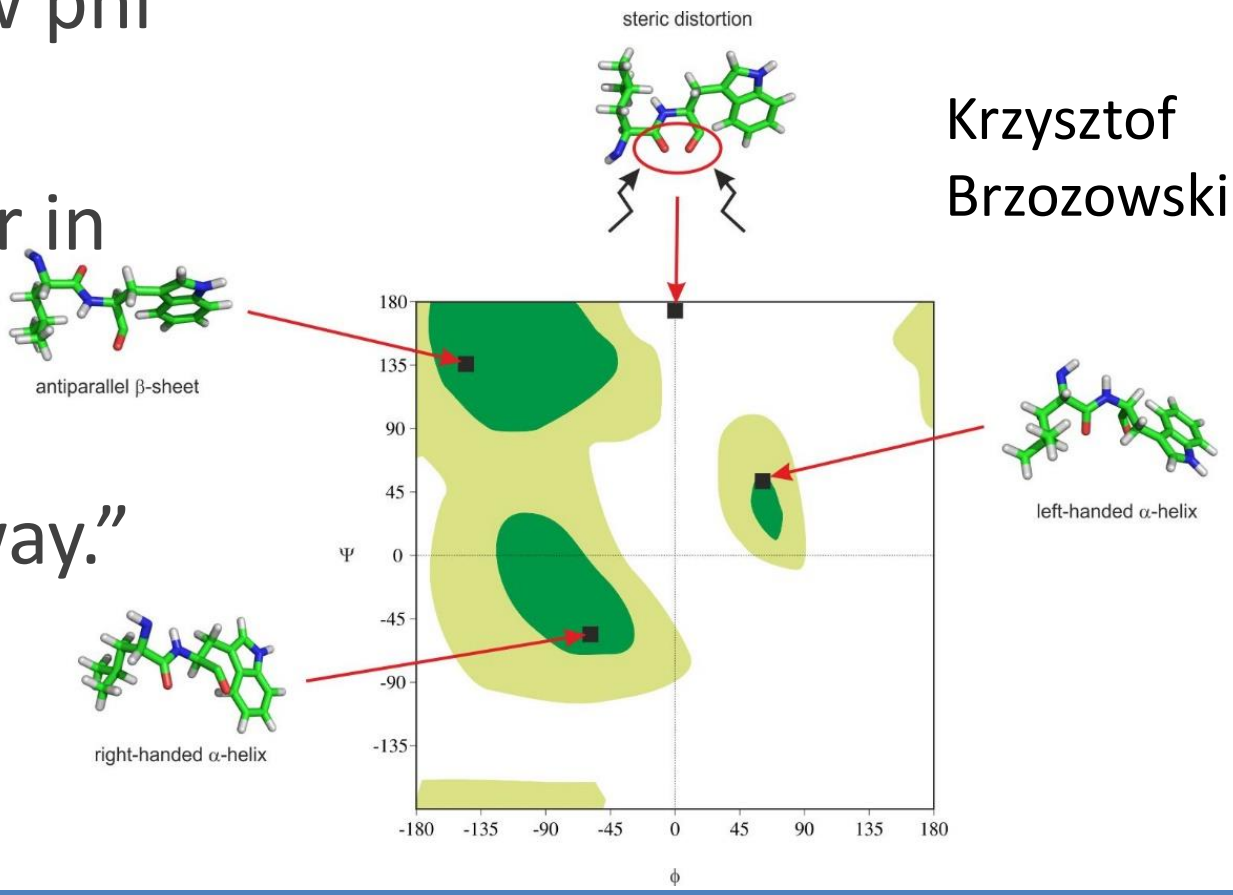
Figure 2.22a
Biochemistry, Seventh Edition

Ramachandran et al. *J. Mol. Biol*. (1963) 7:95-99.

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Ramachandran related phi and psi to secondary structure

- High psi and low phi → β sheet.
- α helices appear in two regions.
- "Steric" means "getting in the way."

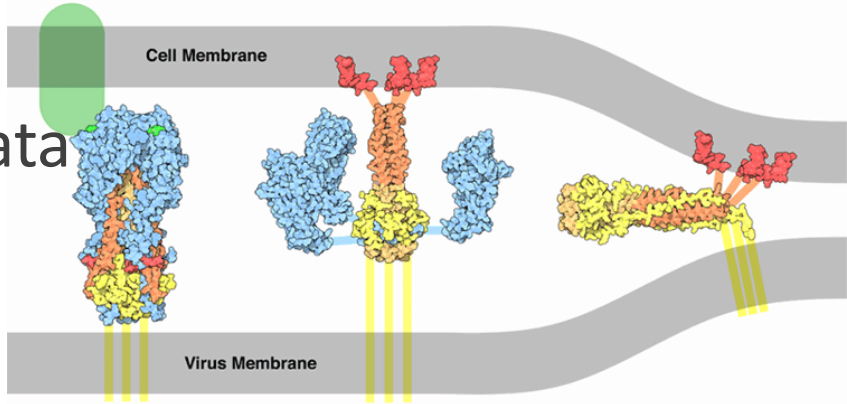


Krzysztof Brzozowski

Ho et al. *Protein Science* (2003) 12: 2508-2522

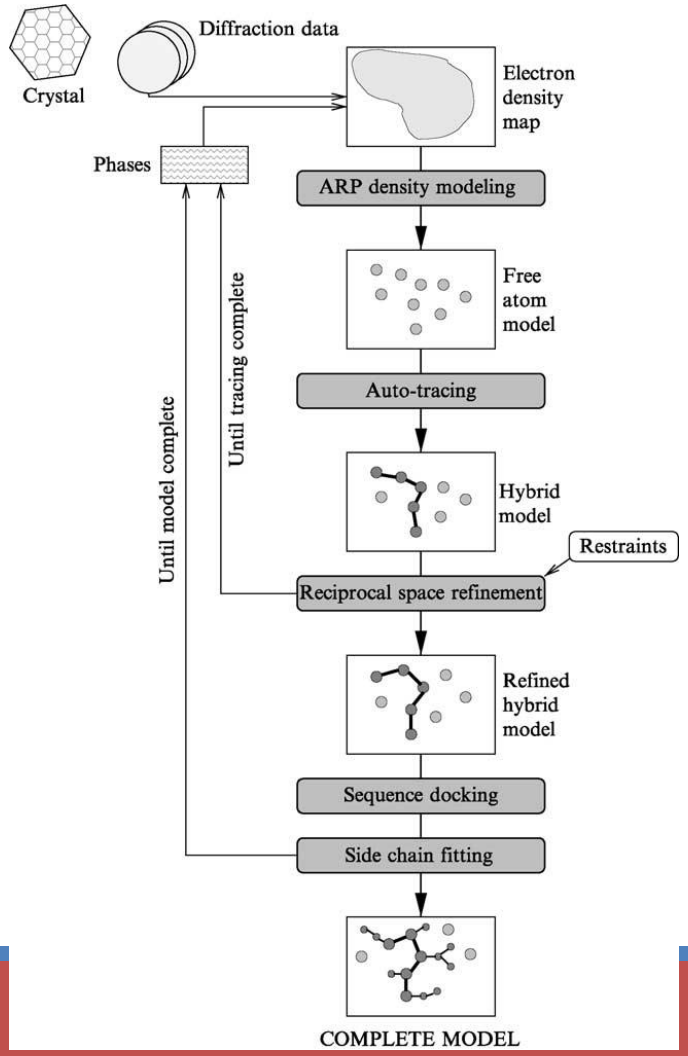
Methods to measure structure empirically

- X-Ray Crystallography
  - Highest resolution (1-3 Å) data
  - Good for soluble proteins
- NMR Spectroscopy
  - 4-6 Å data…protein fold (backbone)
  - Good for small, flexible proteins
- Electron Microscopy
  - ~10 Å or lower resolution (overall shape)
  - Good for macromolecular complexes

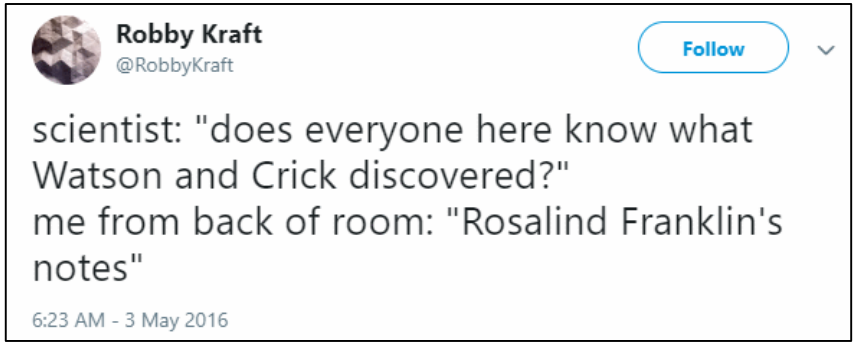


Hemagglutinin image from https://pdb101.rcsb.org/motm/76

Algorithms infer structures from electron density



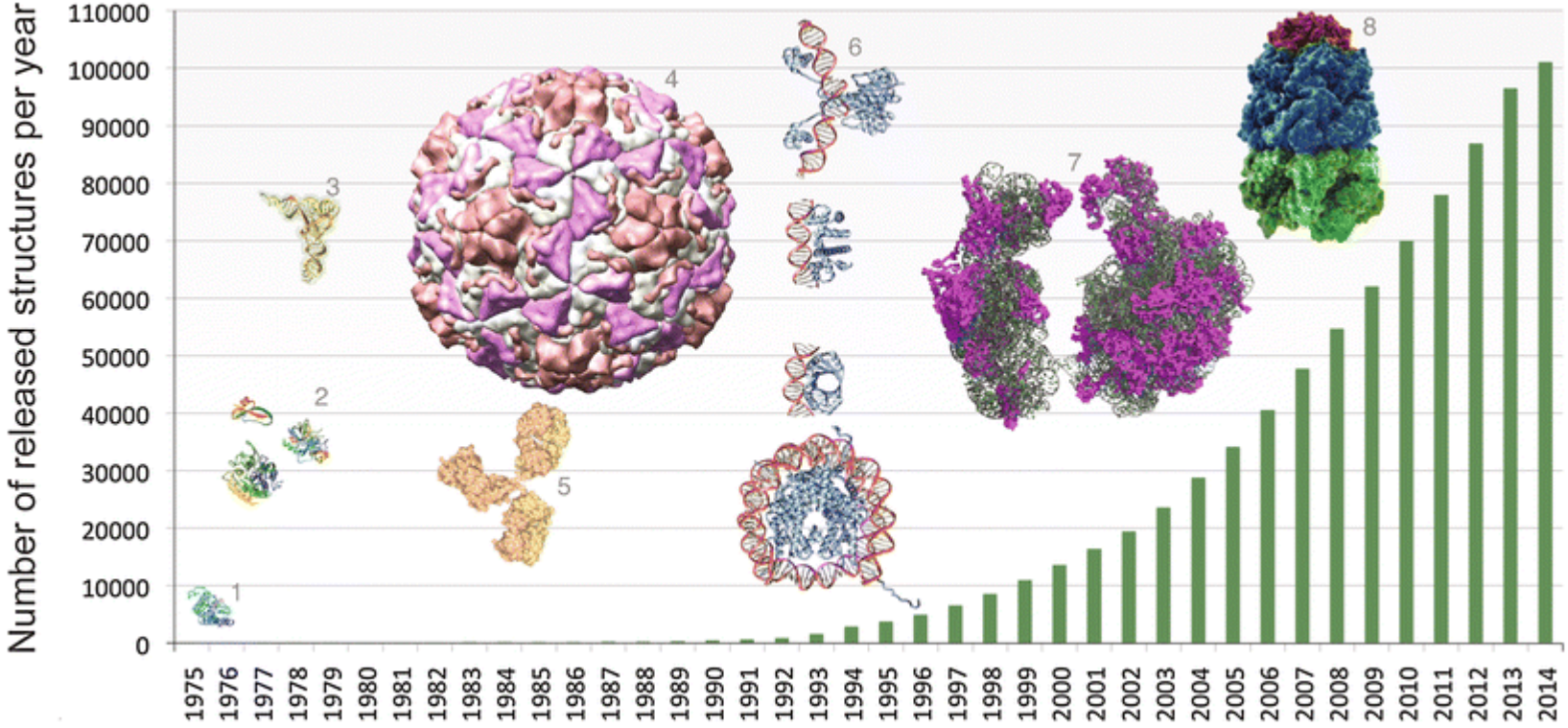
- Recognizing relationships between atoms enables backbone tracing.
- Side chains are only attempted when backbone is in place.



RJ Morris et al. *Methods Enzymol*. (2003) 374: 229-244.

Structures since 1971

The Protein Data Bank is an international resource for structures.



http://www.wwpdb.org/

Berman et al. *J. Comp. Aid. Molec. Design* (2014) 28: 1009-1014

Levinthal's paradox

- "It would not be possible in a physically meaningful time for a protein to reach the native (functional) conformation by a random search of the enormously large number of possible structures."

So how do proteins fold correctly, the first time?

- "The introduction of a small energetic bias toward the native state leads to decreases in folding rates such that they can be compared with experiment."

Martinez. *J. Chem. Educ*. (2014) 91: 1918-1923

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Threading: perturbing a known structure with AA changes

- When a structure has already been determined for a closely related sequence, one can estimate the structure for a query sequence by *comparative modelling* or *threading:*

1. Identify structures for related sequences.
2. Align the sequence to the template structure.
3. Build a model reflecting the altered side chains.
4. Assess solvent-accessible surface area for model.

Pieper et al. *Nucl. Acids Res*. (2011) 39: D465-D474 Kroncke et al. *Biochemistry* (2015) 54: 2551

UNIVERSITEIT STELLENBOSCH UNIVERSITY

Concepts in *ab initio* structural inference

- Anfinsen's thermodynamic hypothesis (1973): native structure is a unique, stable and kinetically accessible minimum of the free energy.
- Monte Carlo sampling: thousands or millions of structures may simultaneously be generated and "scored" for free energy.
- Hill climbing: the best starting points for round *n* were the "best scorers" in round *n-1*.
- Rotamer library: a collection of low-energy side chain conformations from *k*-mers of sequence

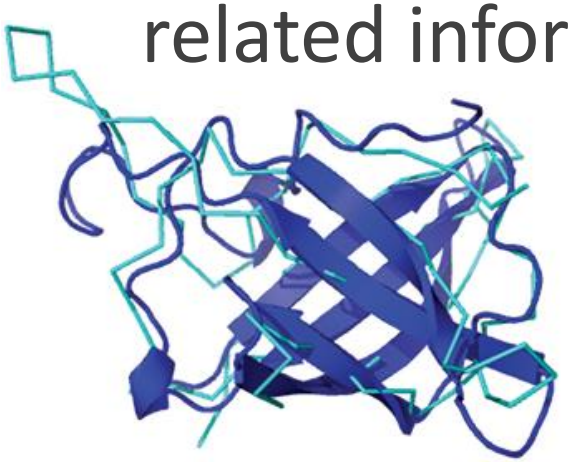
D Shortle et al. *PNAS* (1998) 95: 11158-11162 JC Calvo et al. *J Supercomput*. (2011) 58: 253-260.

CASP 12

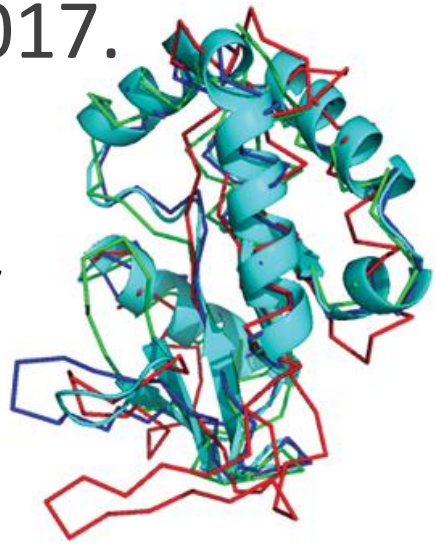
UNIVERSITEIT STELLENBOSCH UNIVERSITY

Critical Assessment of Structural Prediction

- "Participants are provided sequence information and in turn provide protein structure models and related information."
- CASP takes place every two years, starting in 1994. The most recent was reported in 2017.



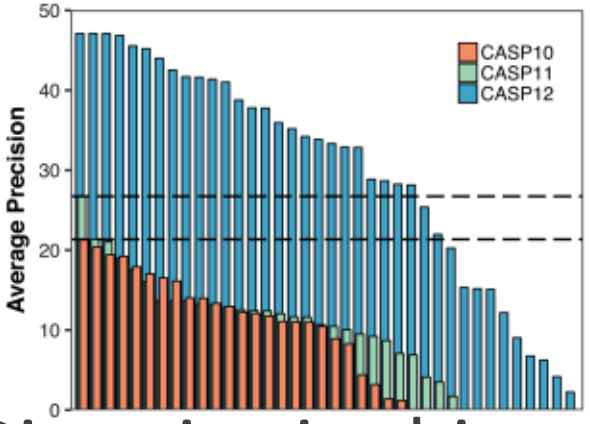
http://predictioncenter.org/



Moult et al. *Proteins* (2017) 86: 7-15.

Are structural prediction methods improving?

- "in this round of CASP the large majority of methods are using machine learning approaches including coevolution data."
- 38 groups registered, and 36 gave results. 32 distinct methods were evaluated.



- Big gains in this round were seen in finding contacts between AAs that are separated by 24 residues, with no available template.

J. Schaarschmidt et al. *Proteins* (2017) 86: 51-66.

Takeaway messages

- The conformations adopted by proteins are essential to their activity!
- Protein structures can be measured through X-ray crystallography, NMR, or electron microscopy.
- Computational methods in protein structure continue to evolve, and structures for small proteins can generally be inferred to a good approximation from sequence alone.