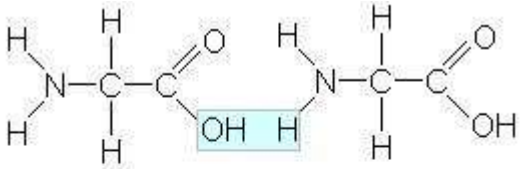
LABORATORY
of THEORY of
BIOPOLYMERS

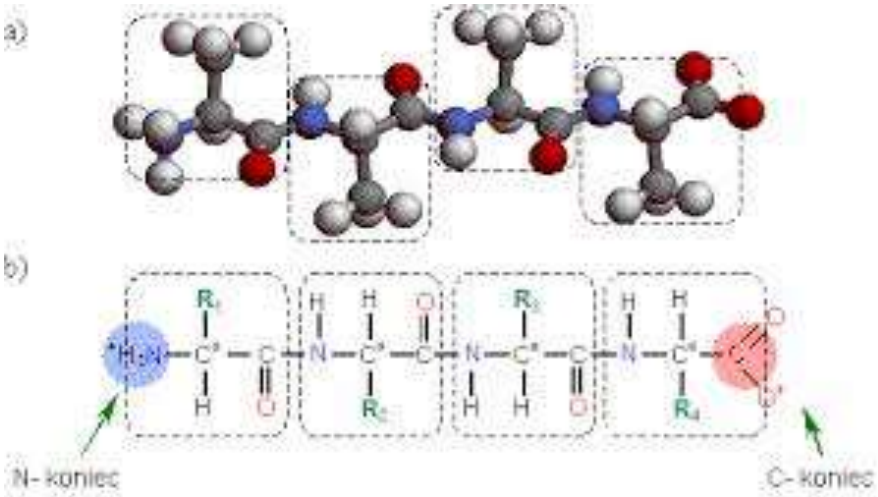
Proteins - structural bioinformatics

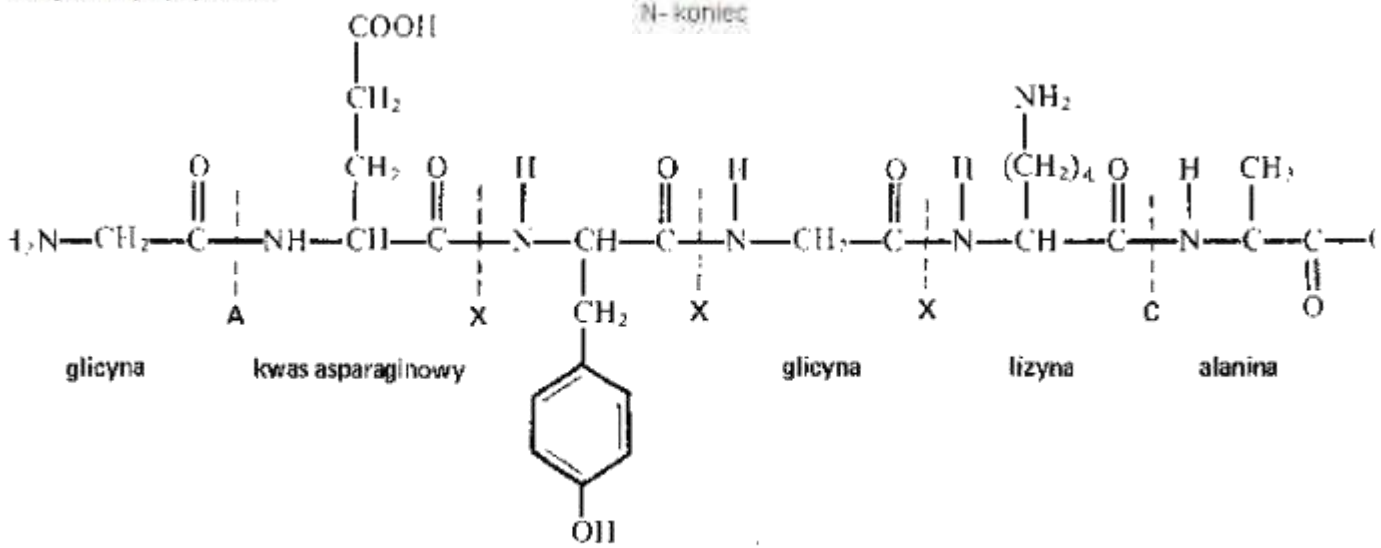
http://biocomp.chem.uw.edu.pl

Polypeptide chain



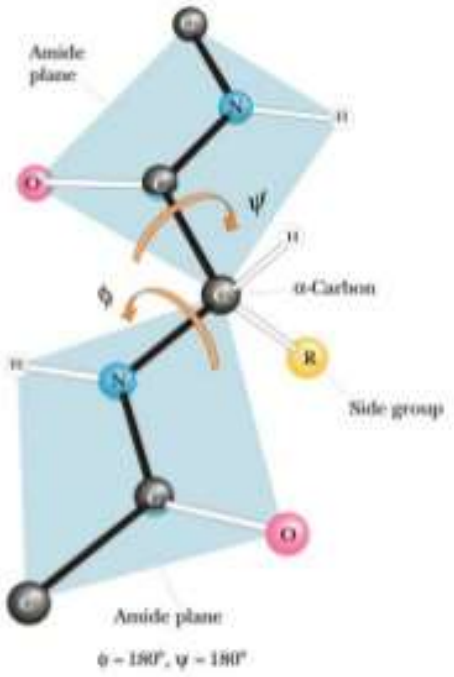
wiązanie peptydowe



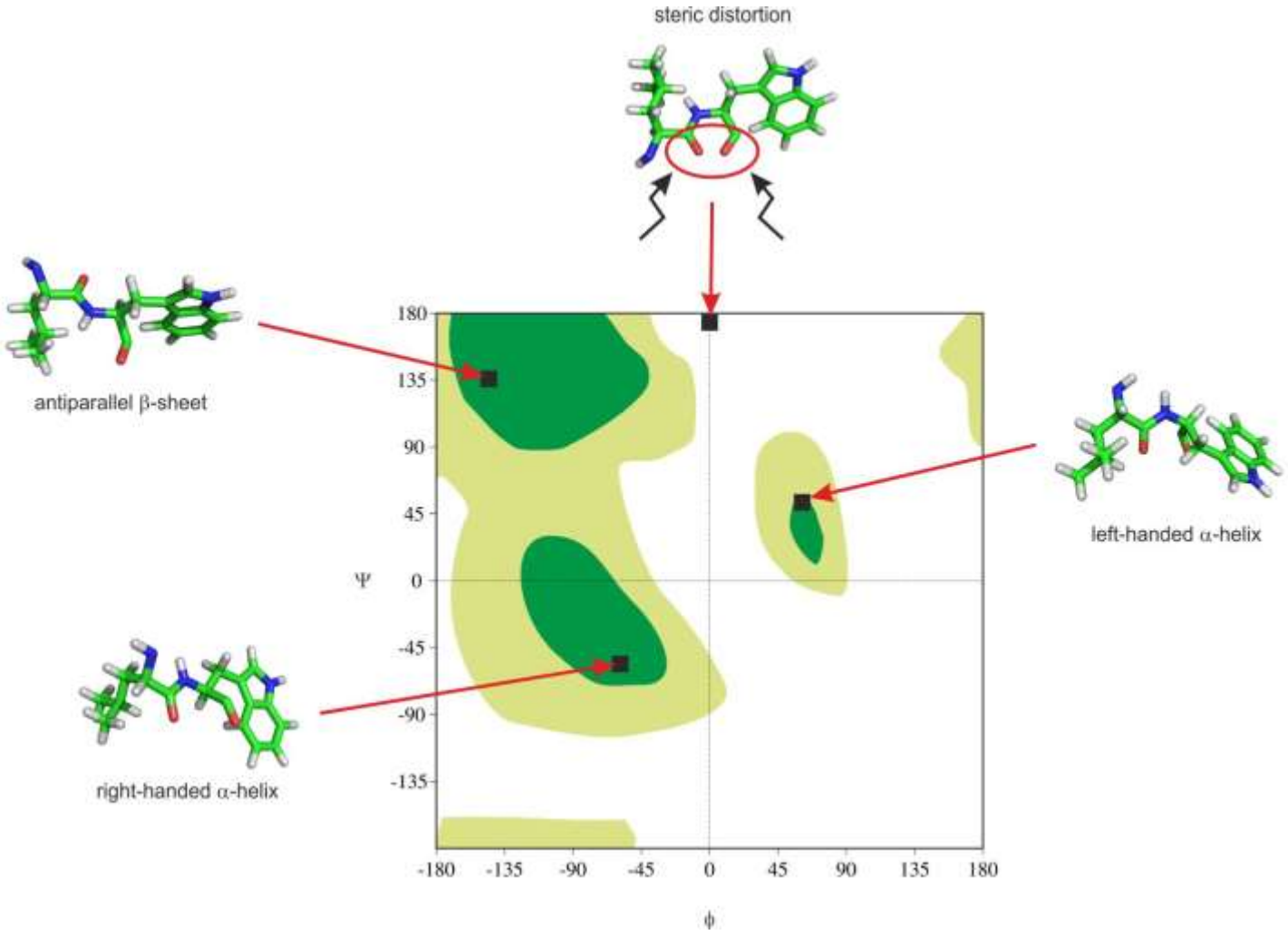


2° Structure Related to Peptide Backbone

- Double bond nature of peptide bond cause planar geometry
- Free rotation at N - αC and αC-carbonyl C bonds
- Angle about the C(alpha)-N bond is denoted phi (ϕ)
- Angle about the C(alpha)-C bond is denoted psi (ψ)
- The entire path of the peptide backbone is known if all phi and psi angles are specified



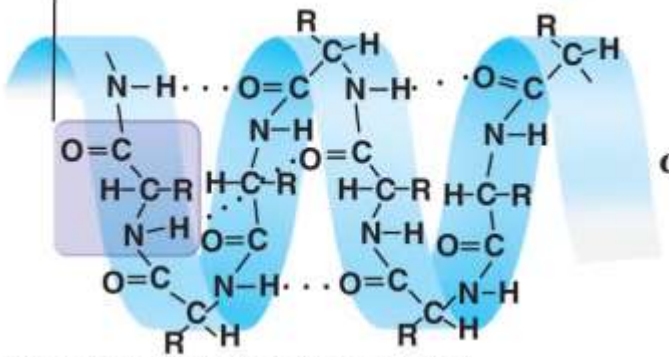
Ramachandran plot



Secondary Structure

β pleated sheet

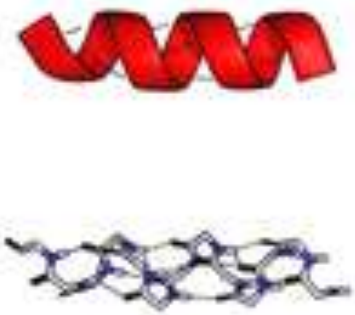
Examples of amino acid subunits



α helix

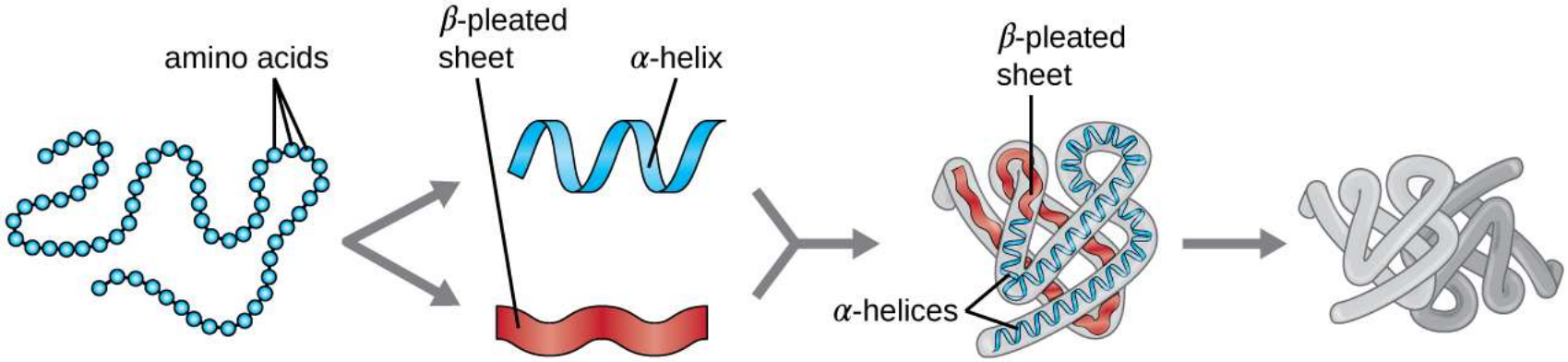
Secondary

β-Sheet (3 strands)



α-helix

Spatial structure



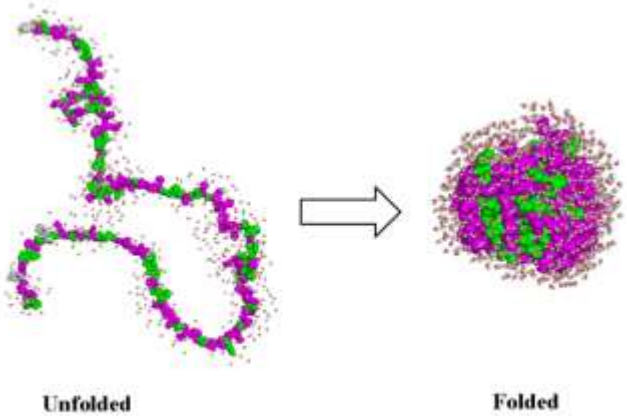
Primary Protein Structure
Sequence of a chain of amino acids

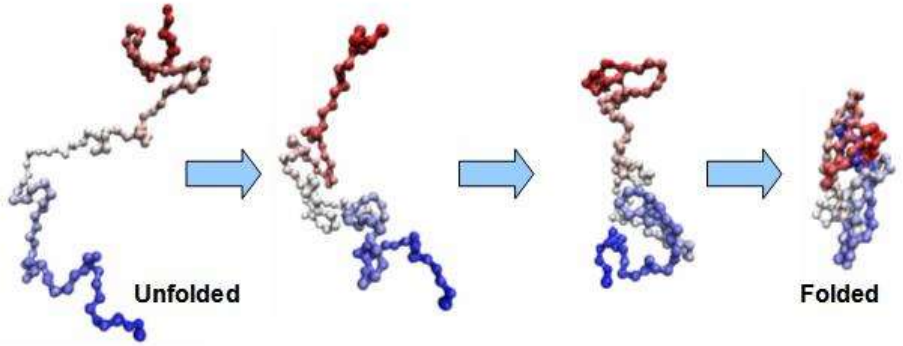
Secondary Protein Structure
Local folding of the polypeptide chain into helices or sheets

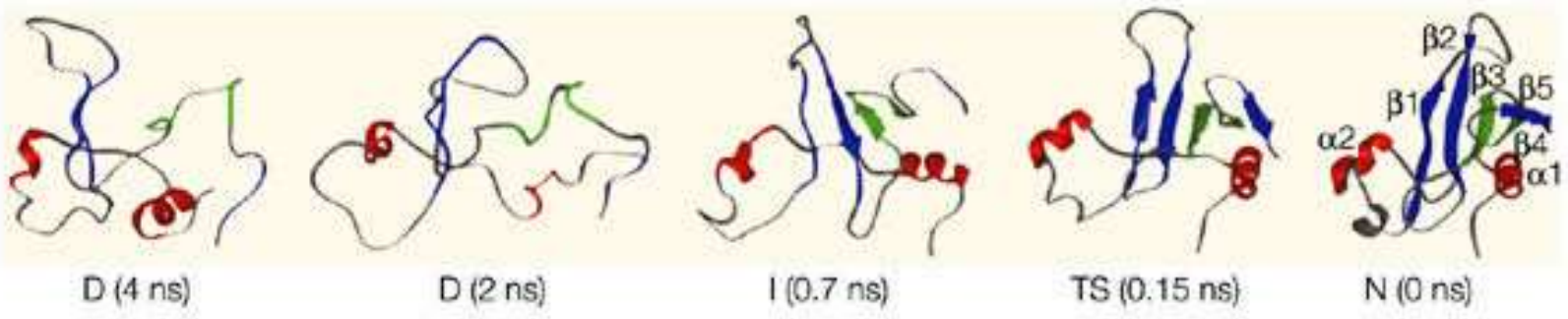
Tertiary Protein Structure
three-dimensional folding pattern of a protein due to side chain interactions

Quaternary Protein Structure
protein consisting of more than one amino acid chain

protein folding problem

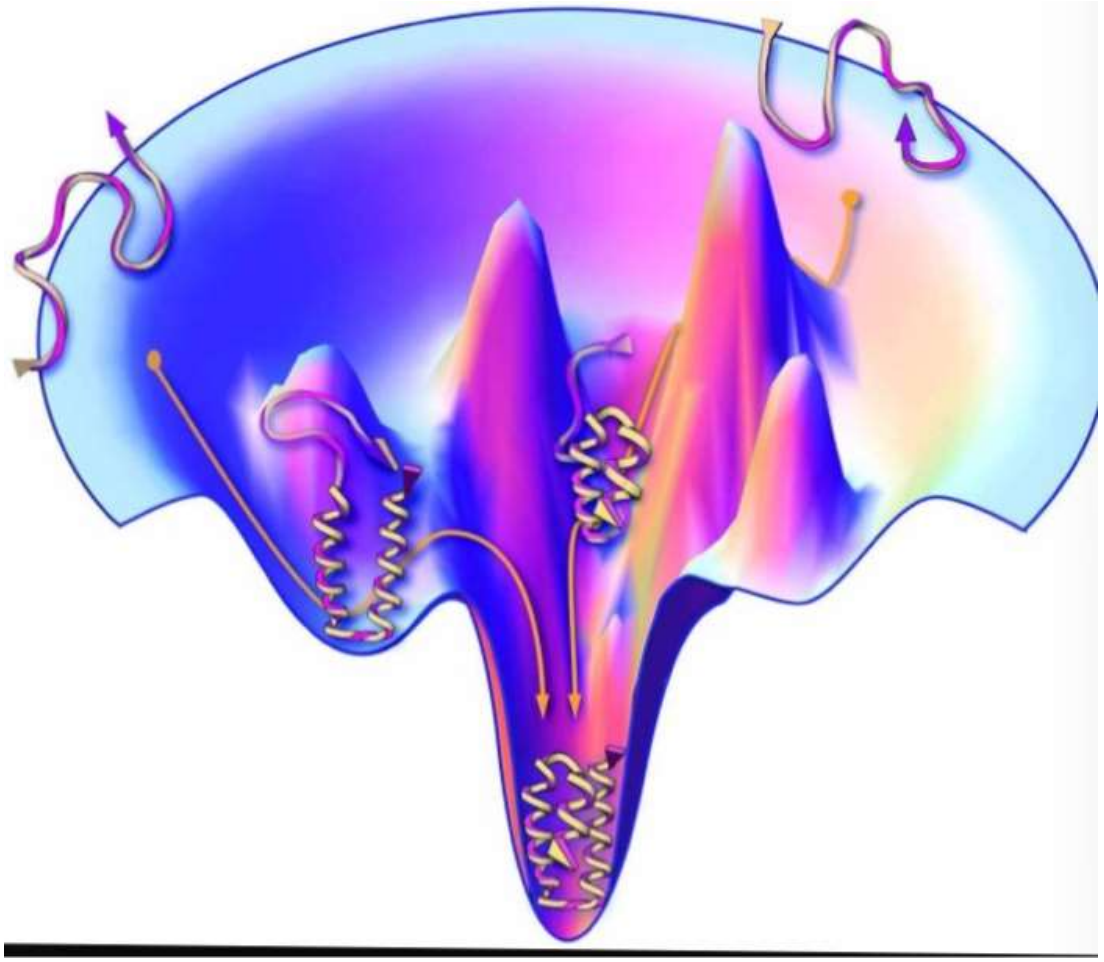






Nature Reviews | Molecular Cell Biology

Protein folding funnel



Tertiary Structure and the "Hydrophobic Effect"

What would this protein look like when properly folded?

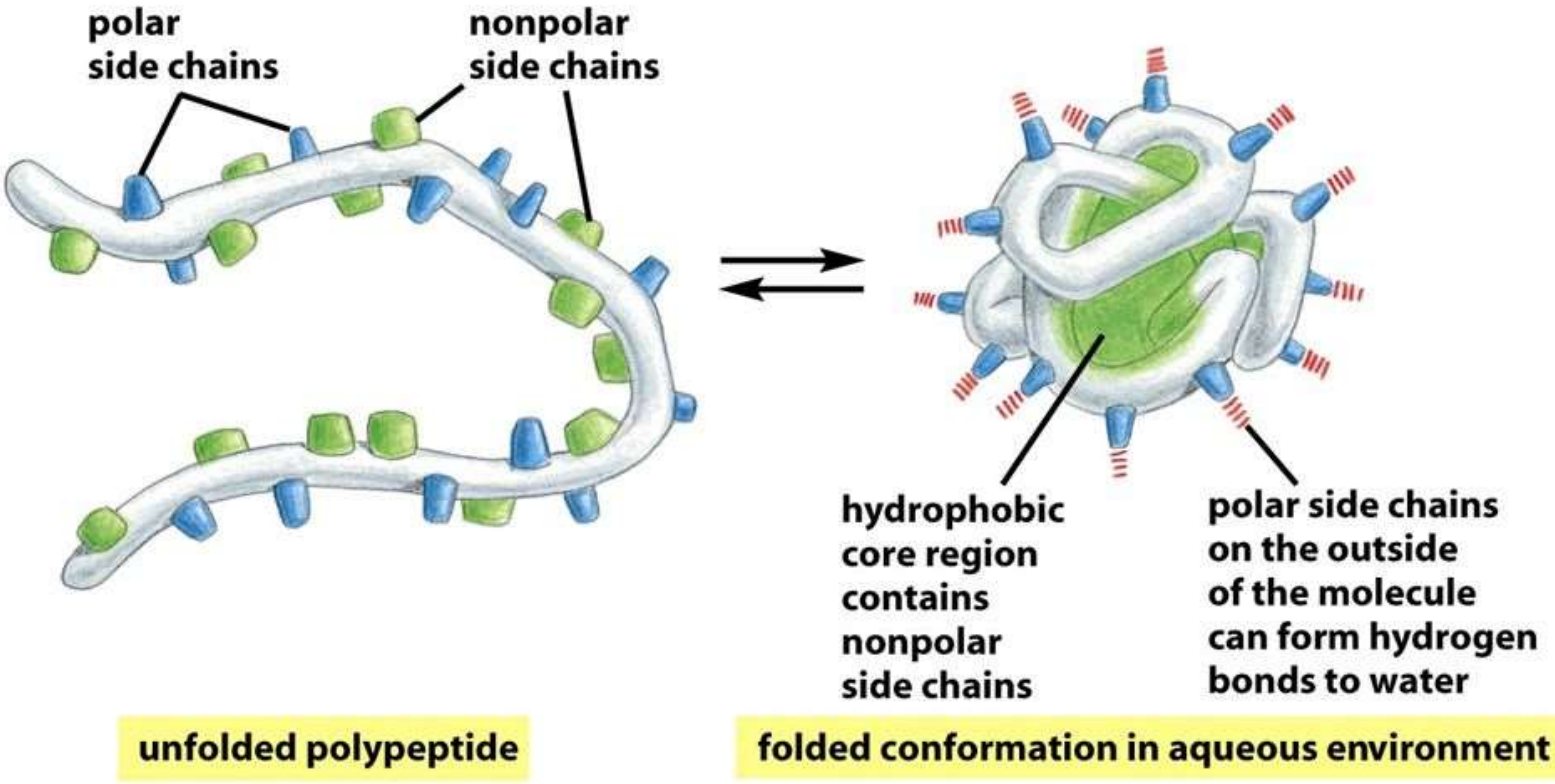
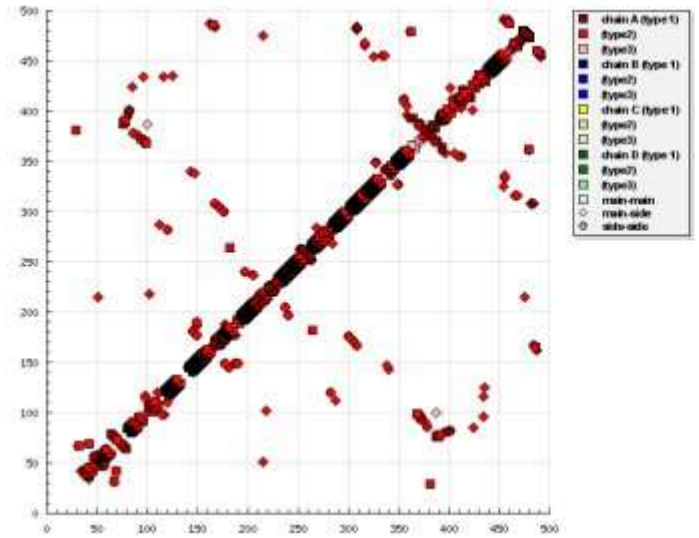
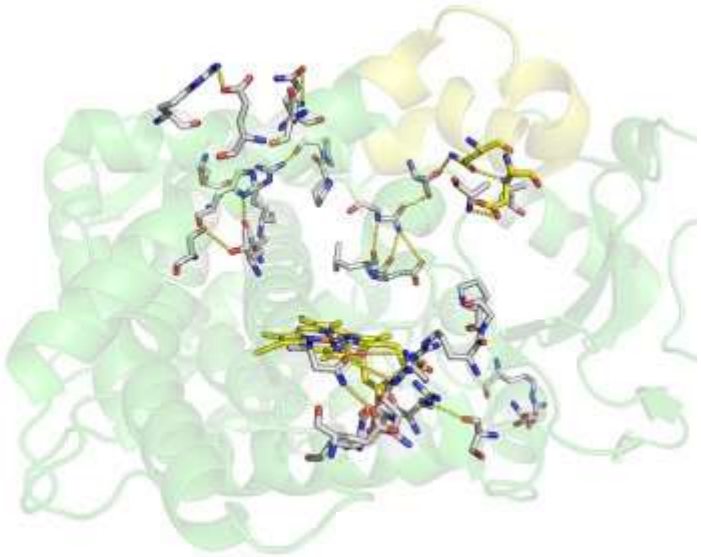
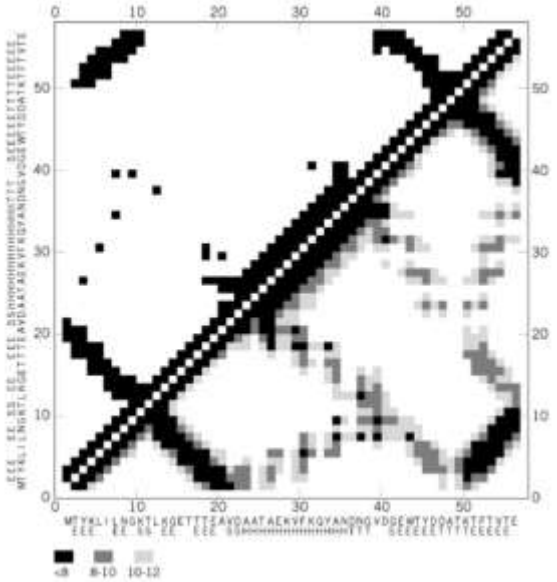


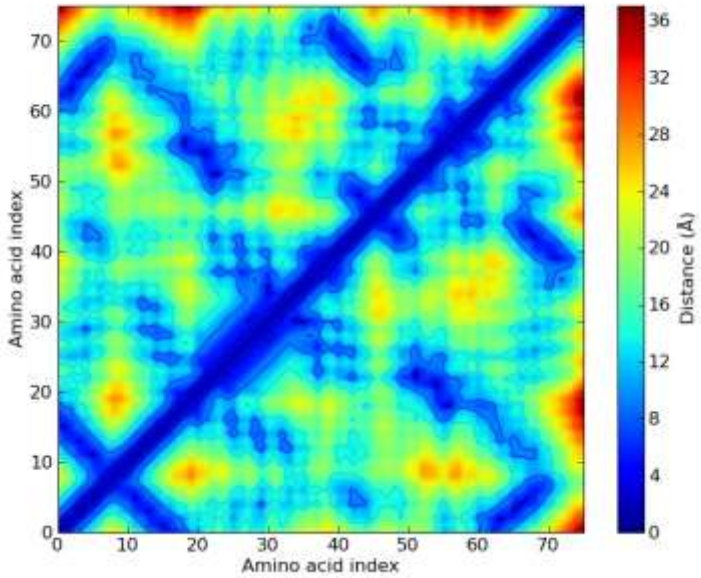
Figure 3-5 *Molecular Biology of the Cell* (© Garland Science 2008)

Side chain packing

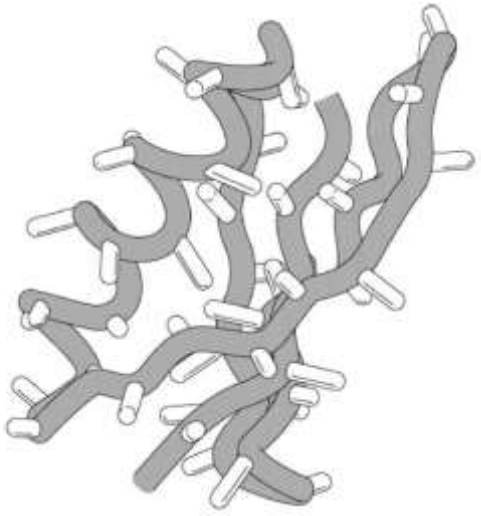


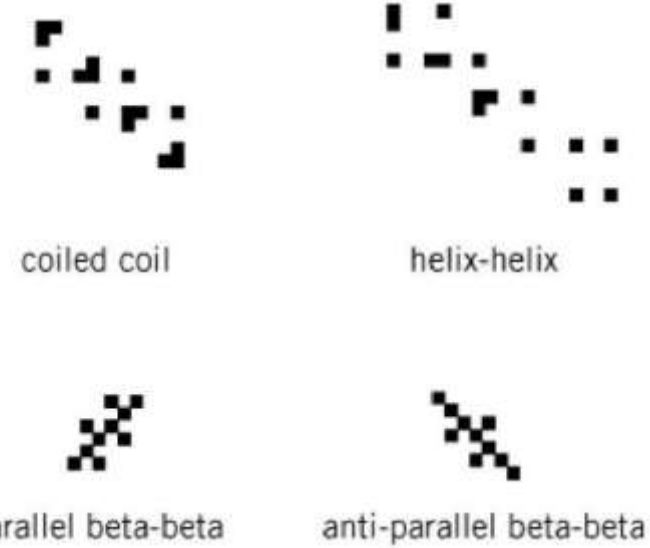


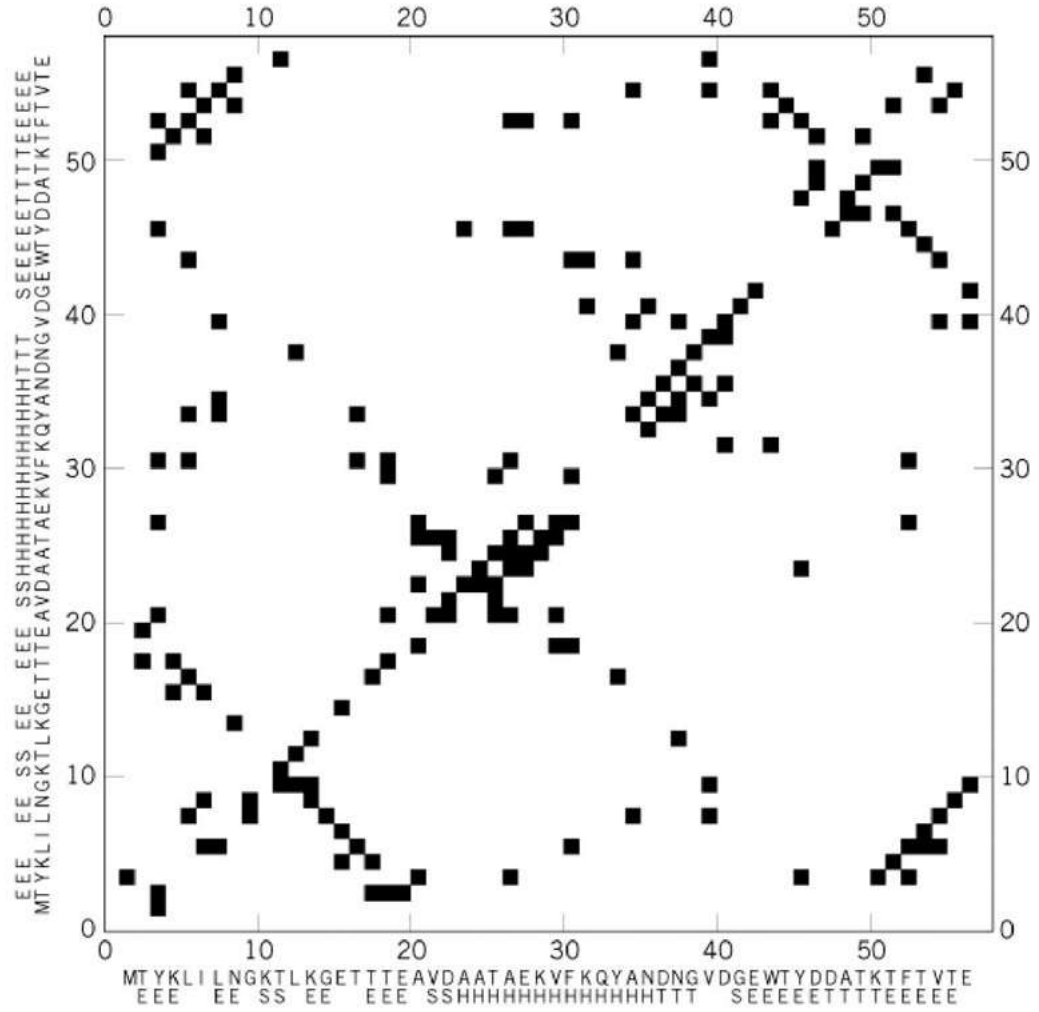




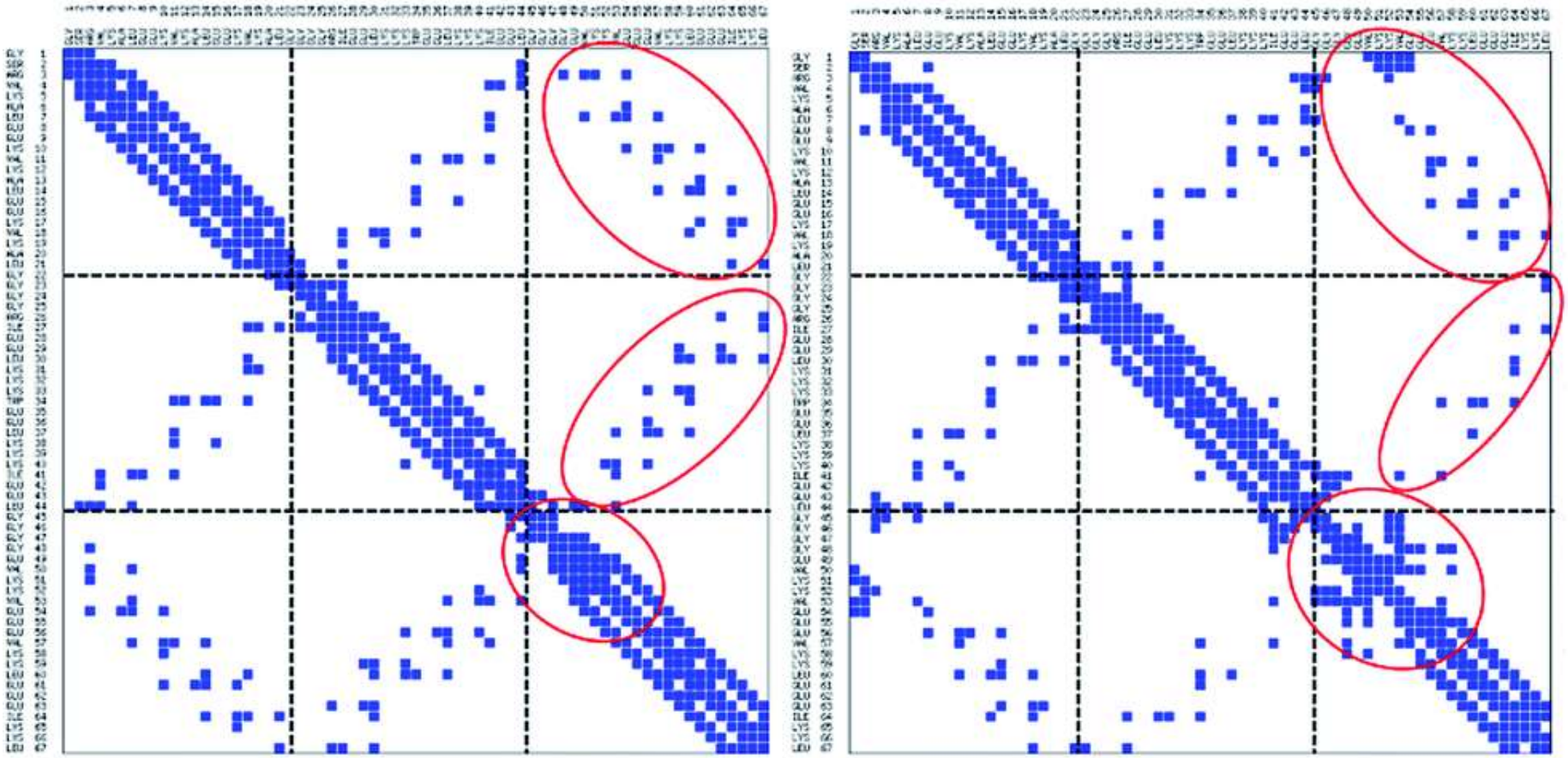
Side chain packing







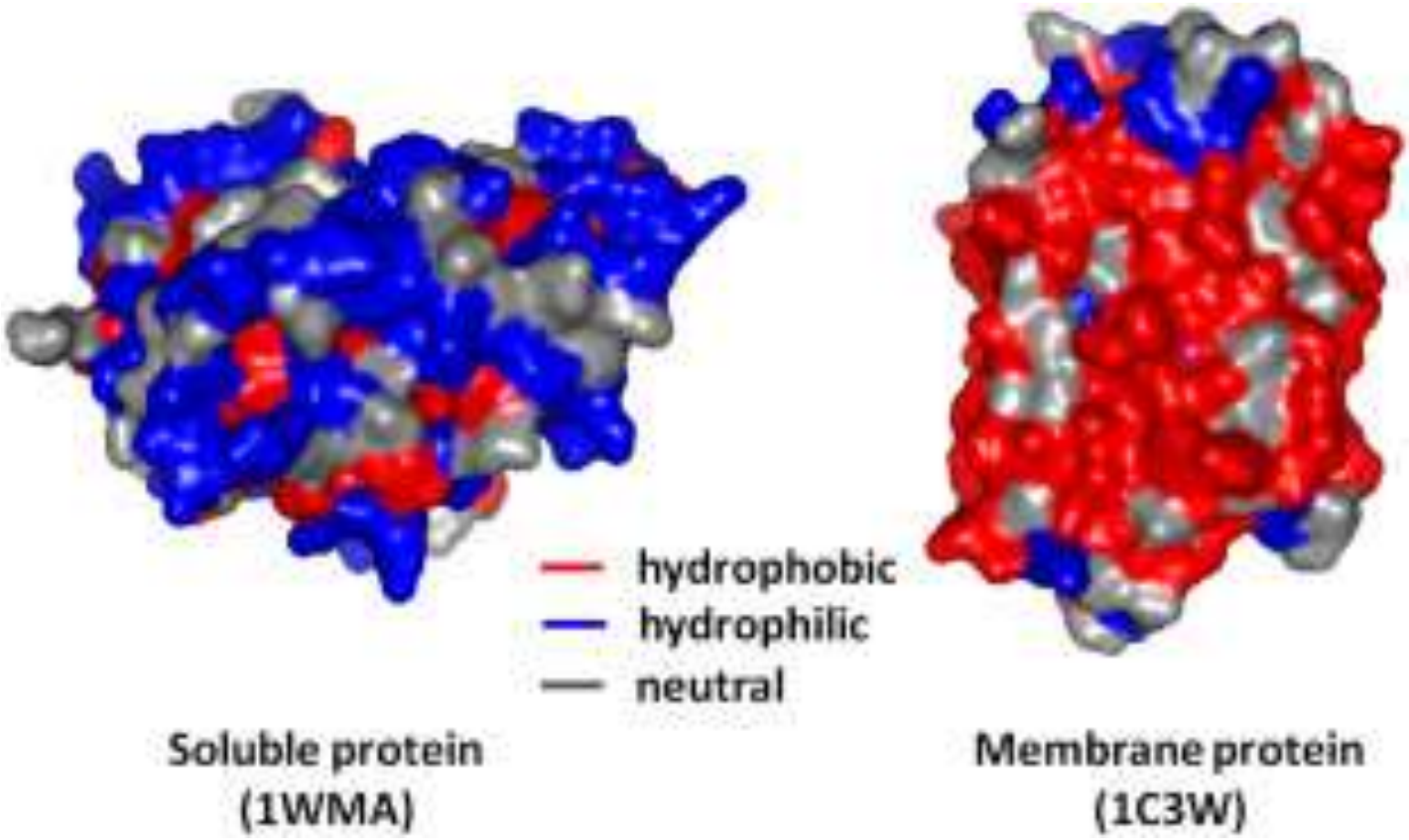
Protein structure-side chain packing



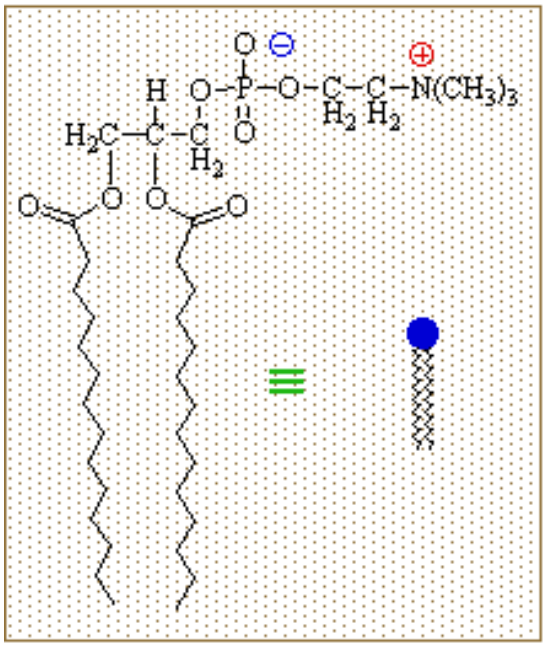
(a) Left-handed Conformation

(b) Right-handed Conformation

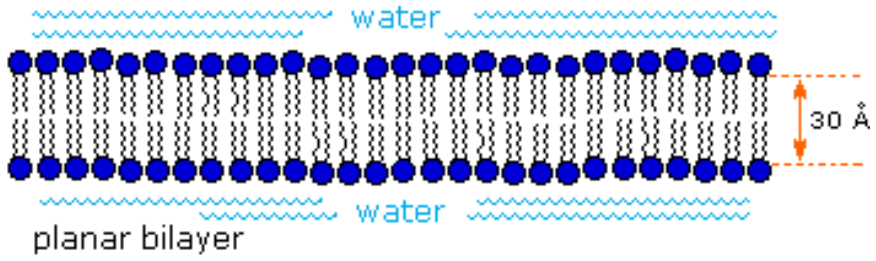
Hydrophobic effects

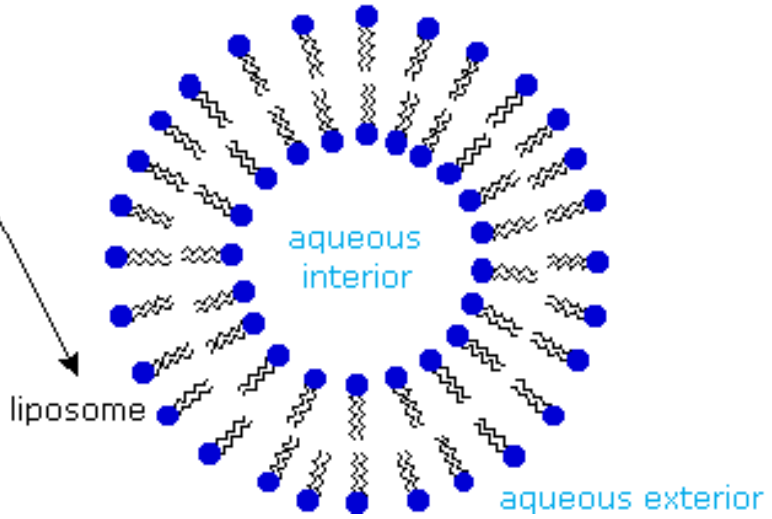


Phospholipid membranes



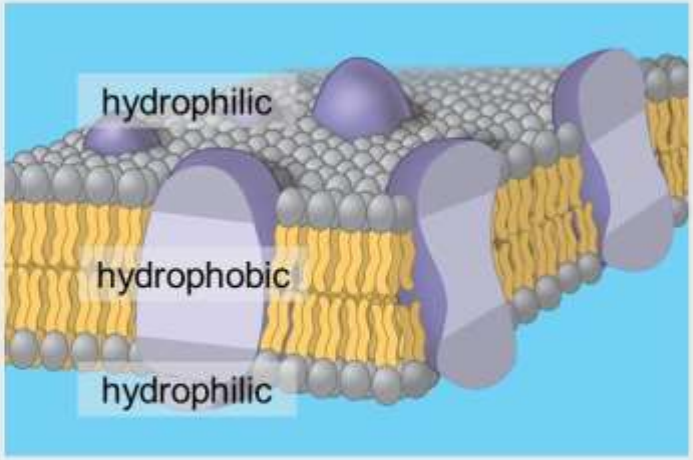
aggregation
in water



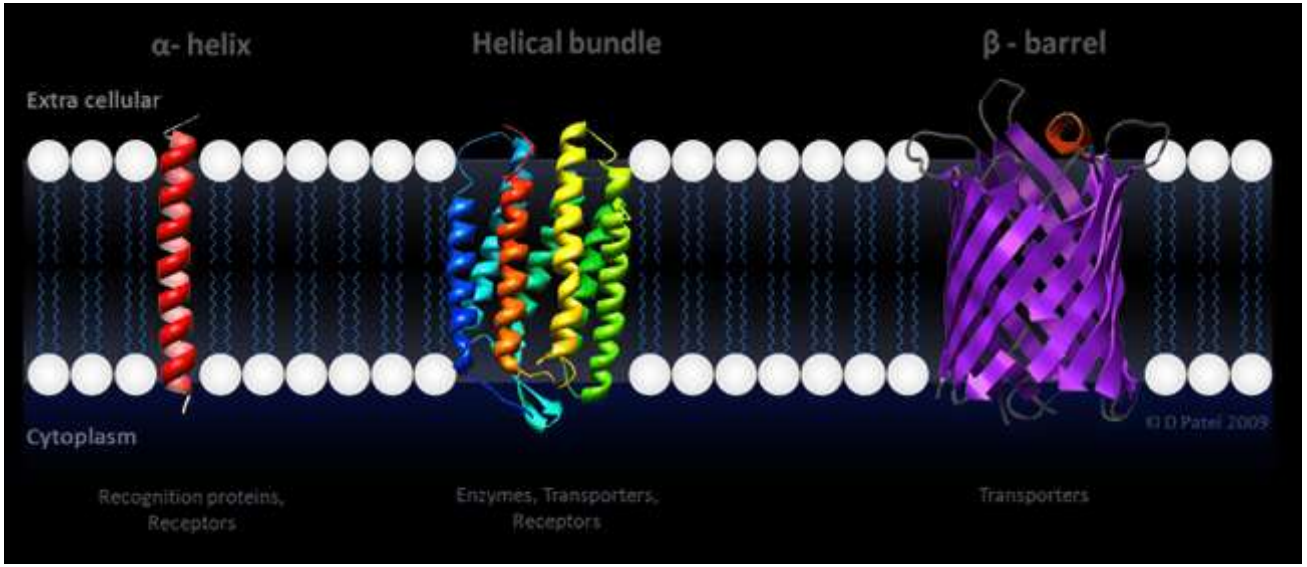


Membrane structure

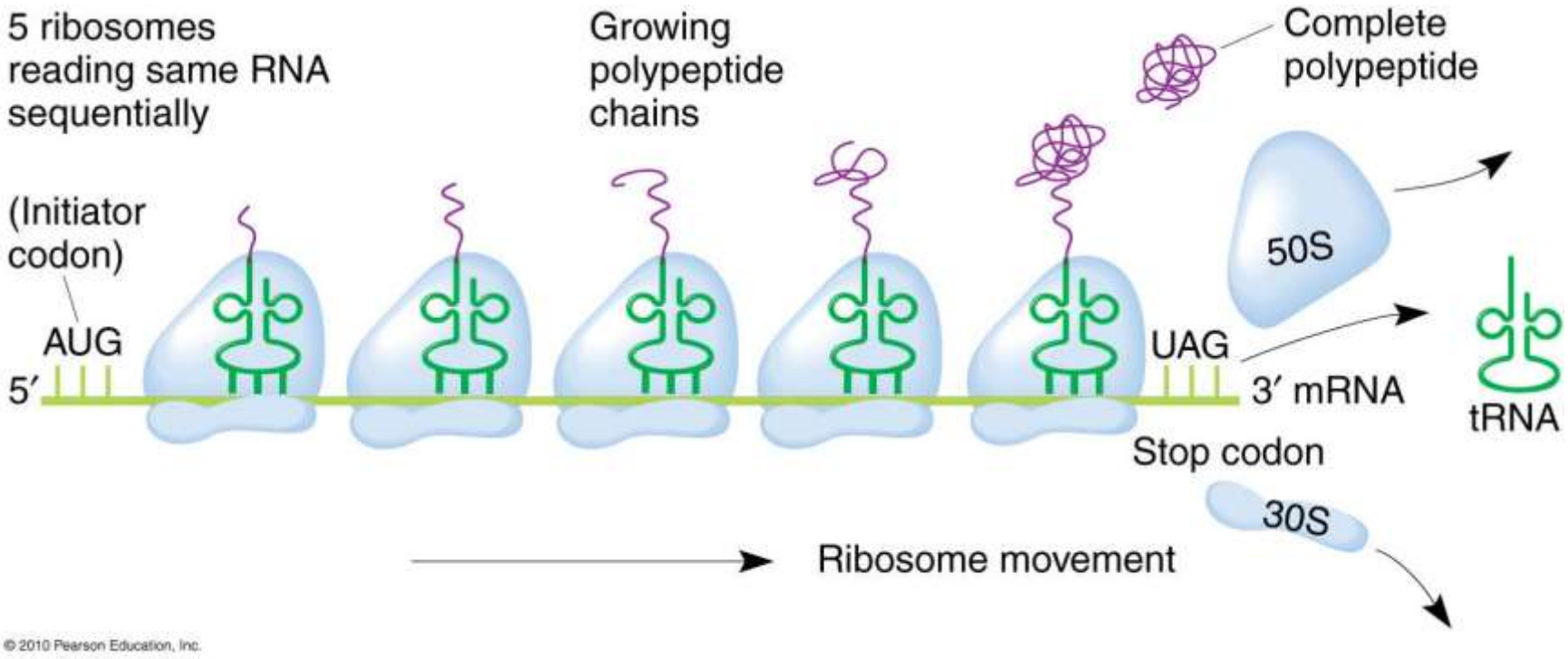
• cell membrane – amphipathic - hydrophilic & hydrophobic



• membrane proteins that are inserted, also amphipathic

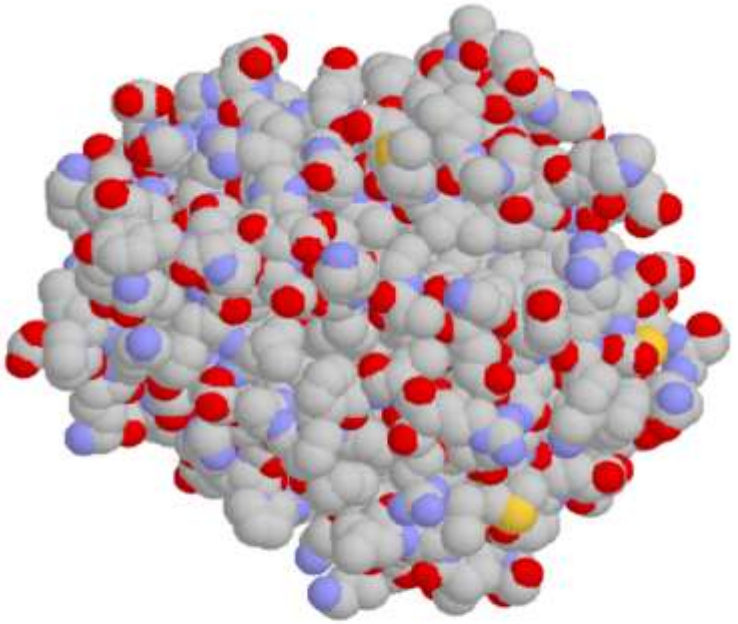


Post-translational or co-translational

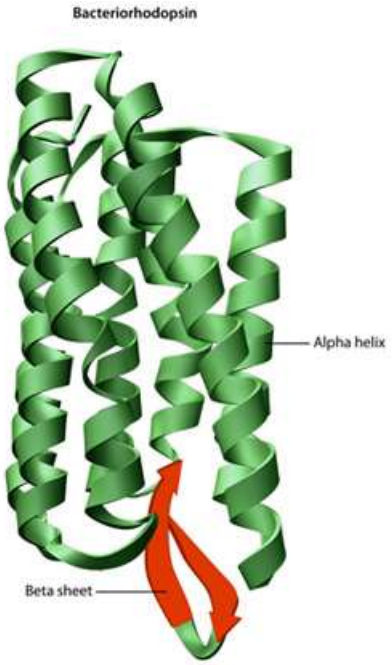


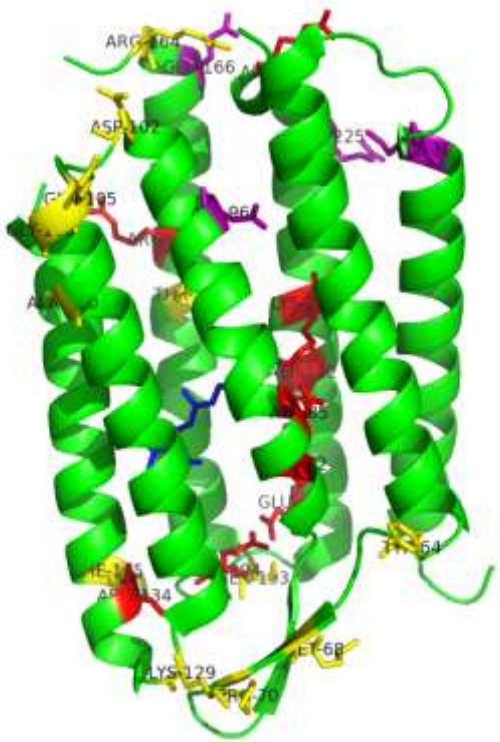
Fold-structure - visualisation

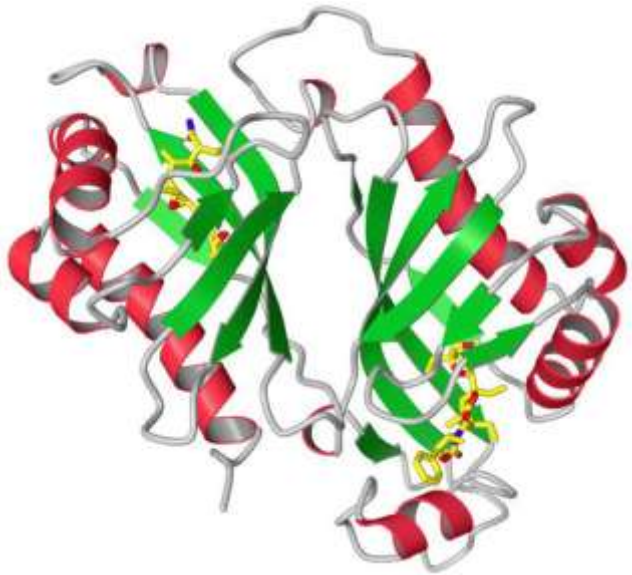
folded protein chain
(main chain view)

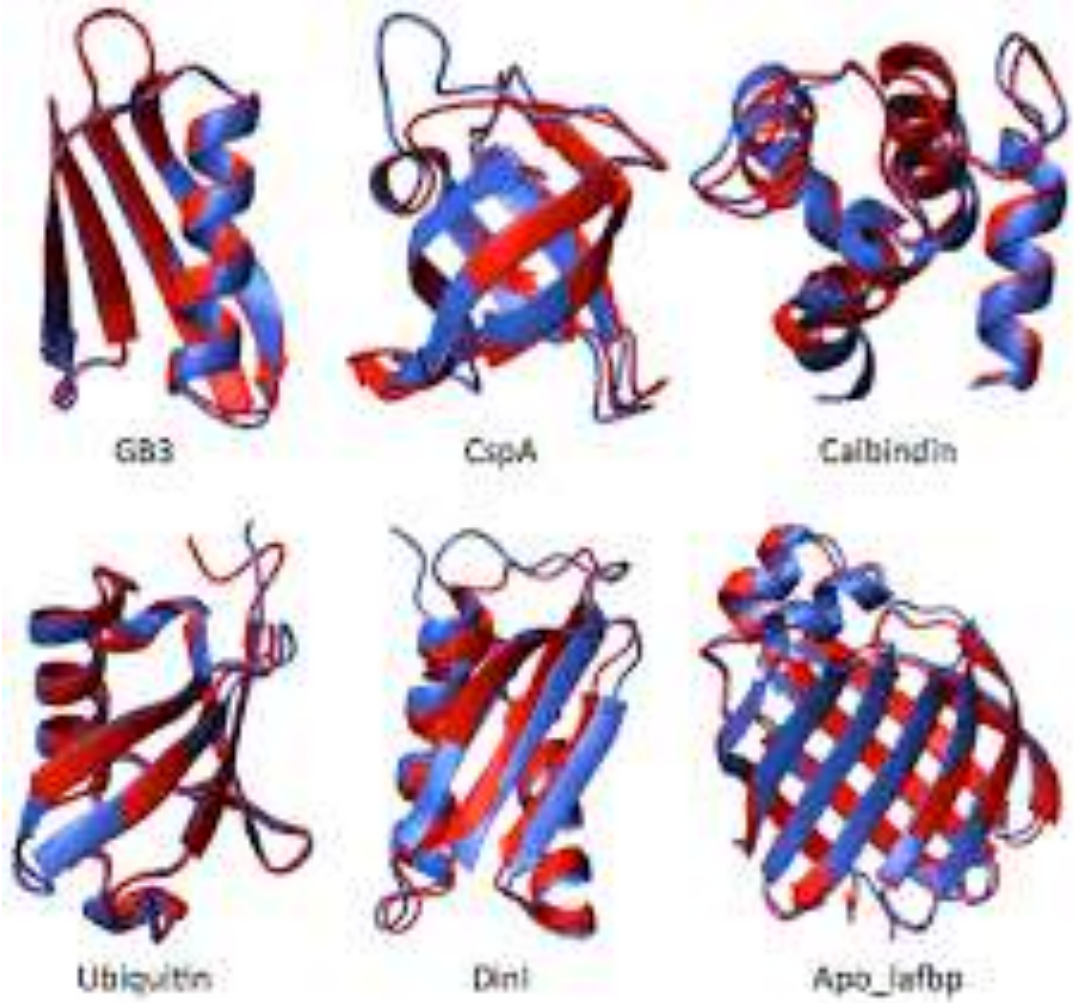


folded protein chain
('space-filling' view)


Bacteriorhodopsin
Alpha helix
Beta sheet

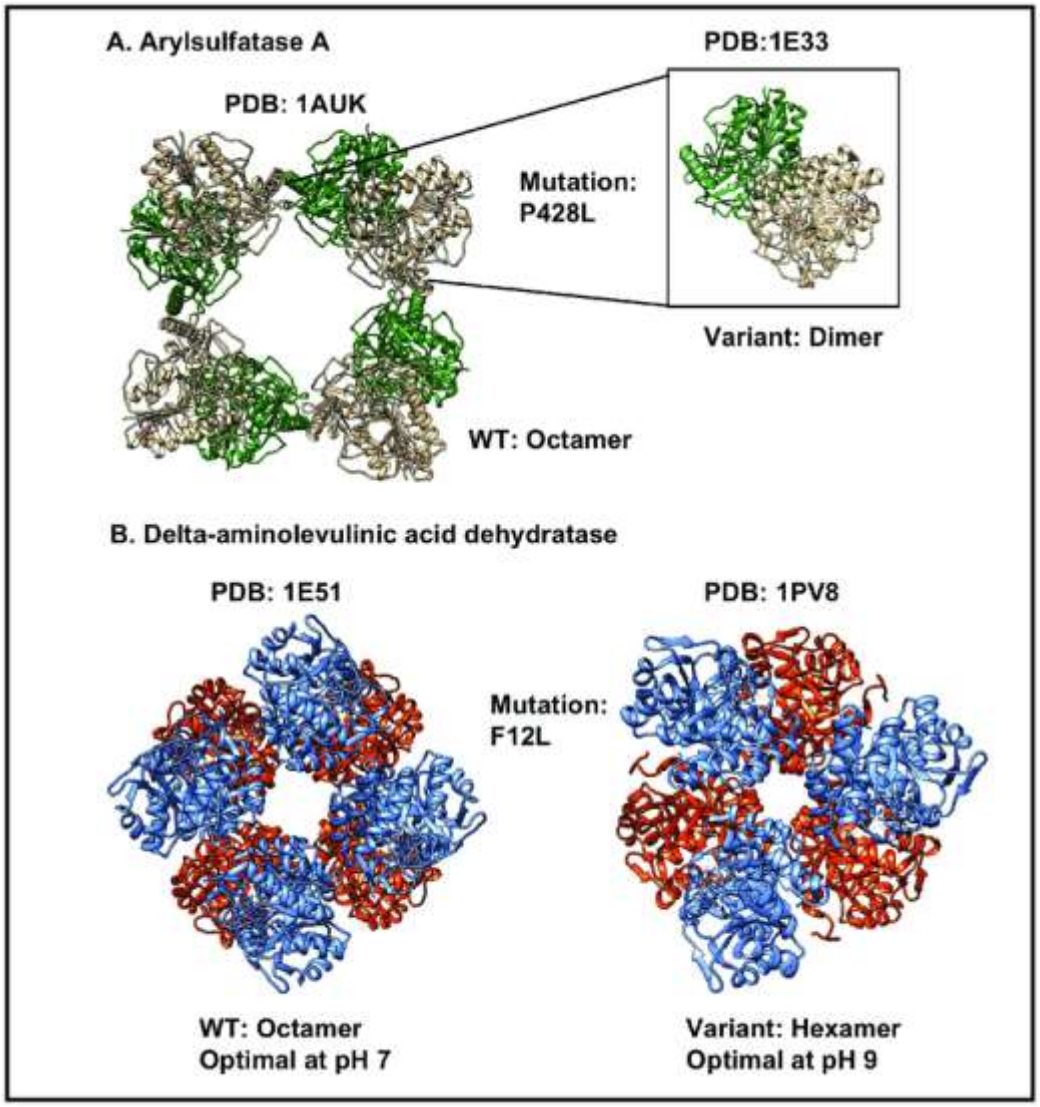


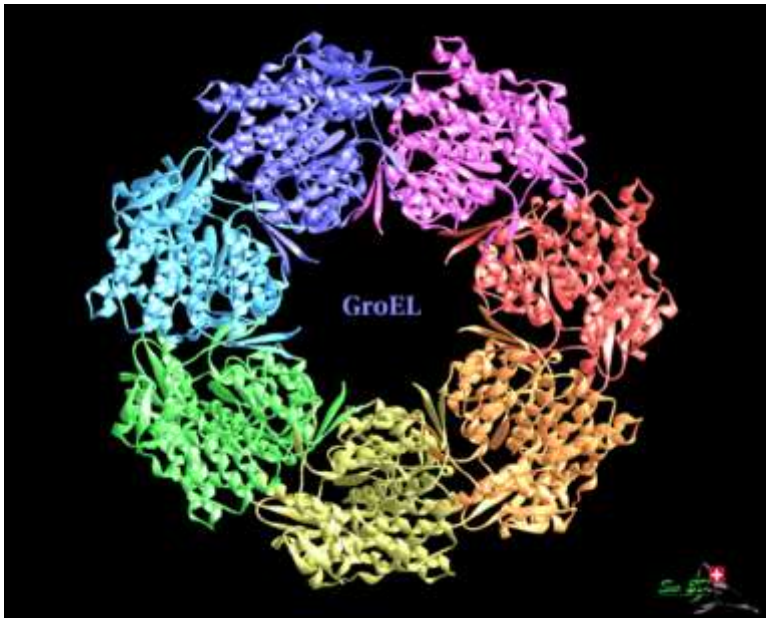


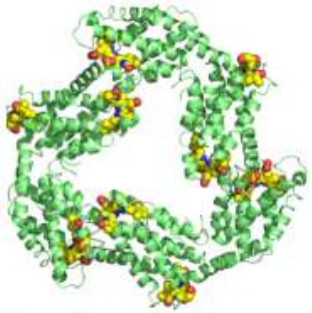


Shen et al. PNAS 2008;105:4685-4690

Struktura przestrzenna białek

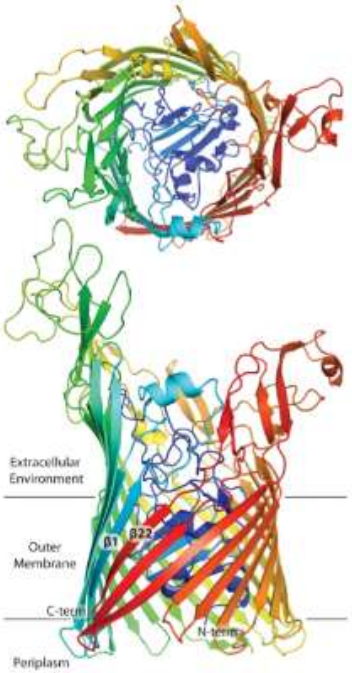




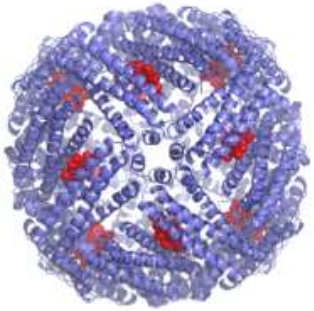


Phycobiliprotein C-Phycocyanin
Richard Cogdell Lab

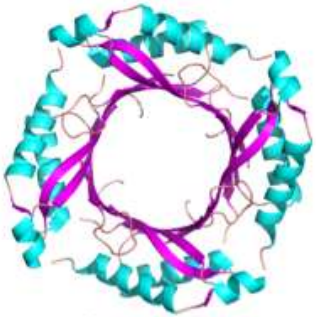
Peroxiredoxin IV
Neil Bulleid Lab



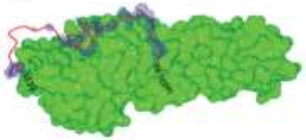
FusA, TonB-dependent outer membrane receptor, Dan Walker Lab



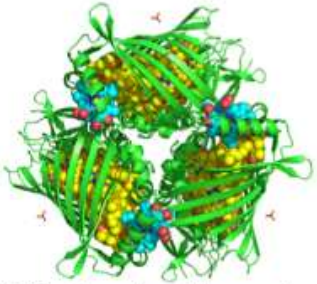
Bacterioferritin, heme binding protein, Richard Cogdell Lab



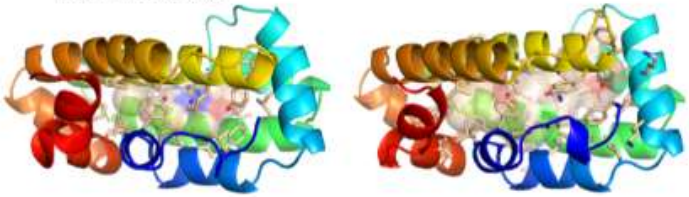
Epimerase FolX, Andy Roe Lab



Bacteriocin Syringacin M
Dan Walker Lab

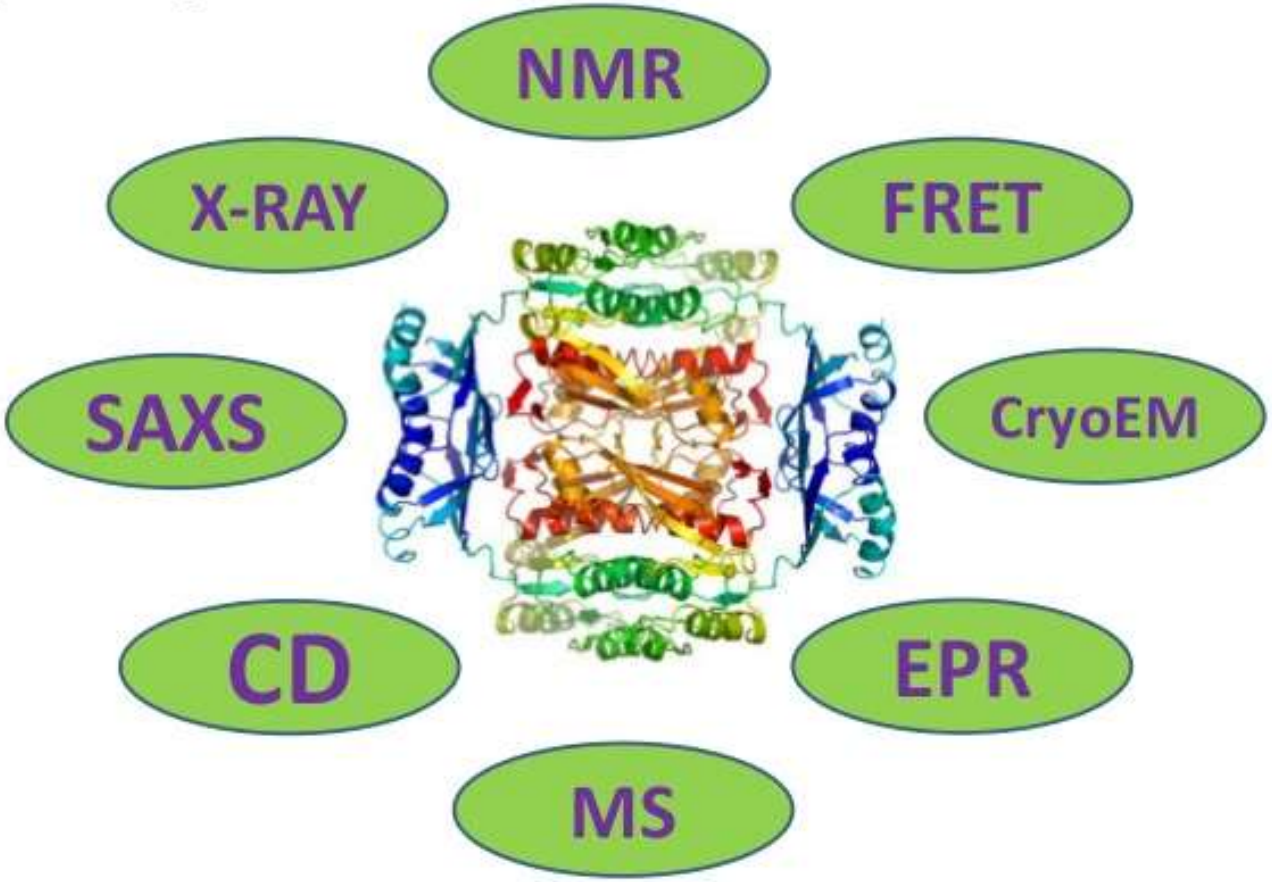


FMO, pigment-protein complex
Richard Cogdell Lab



Na-FAR-1, nematode fatty acid binding protein
Brian Smith Lab

Hybrid protein structure determination

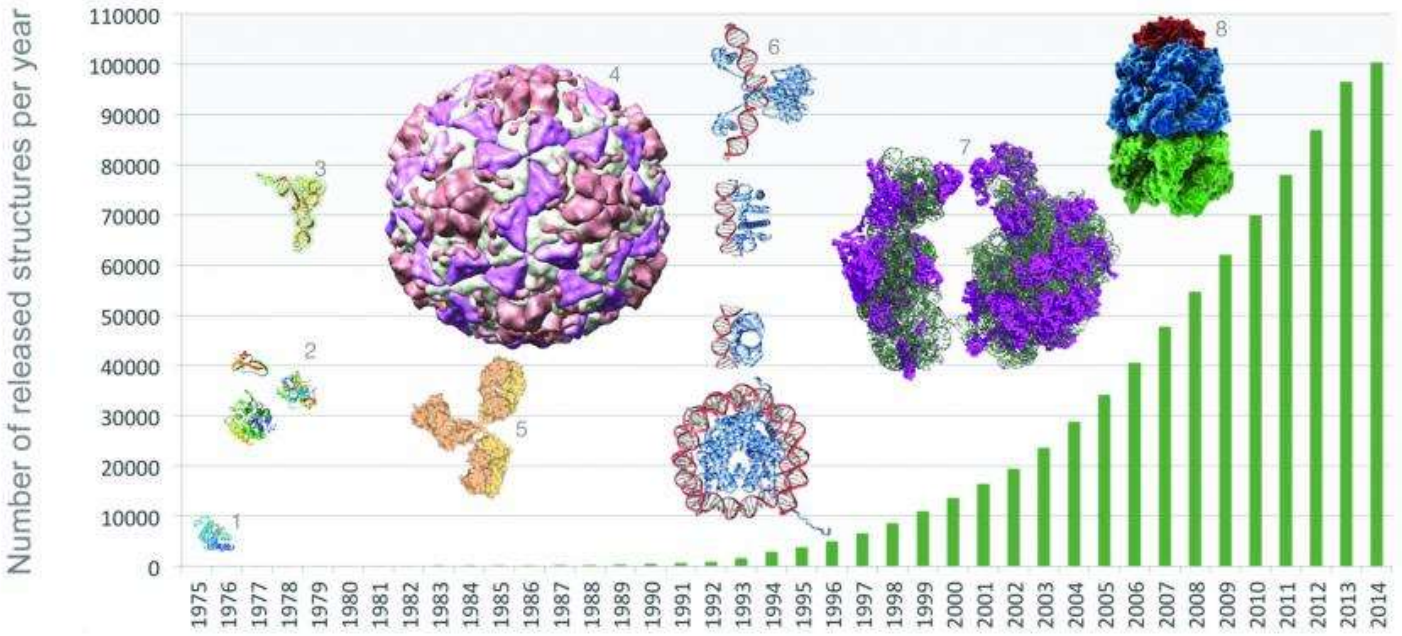


Mark Berjanskii, Edmonton, July 2015

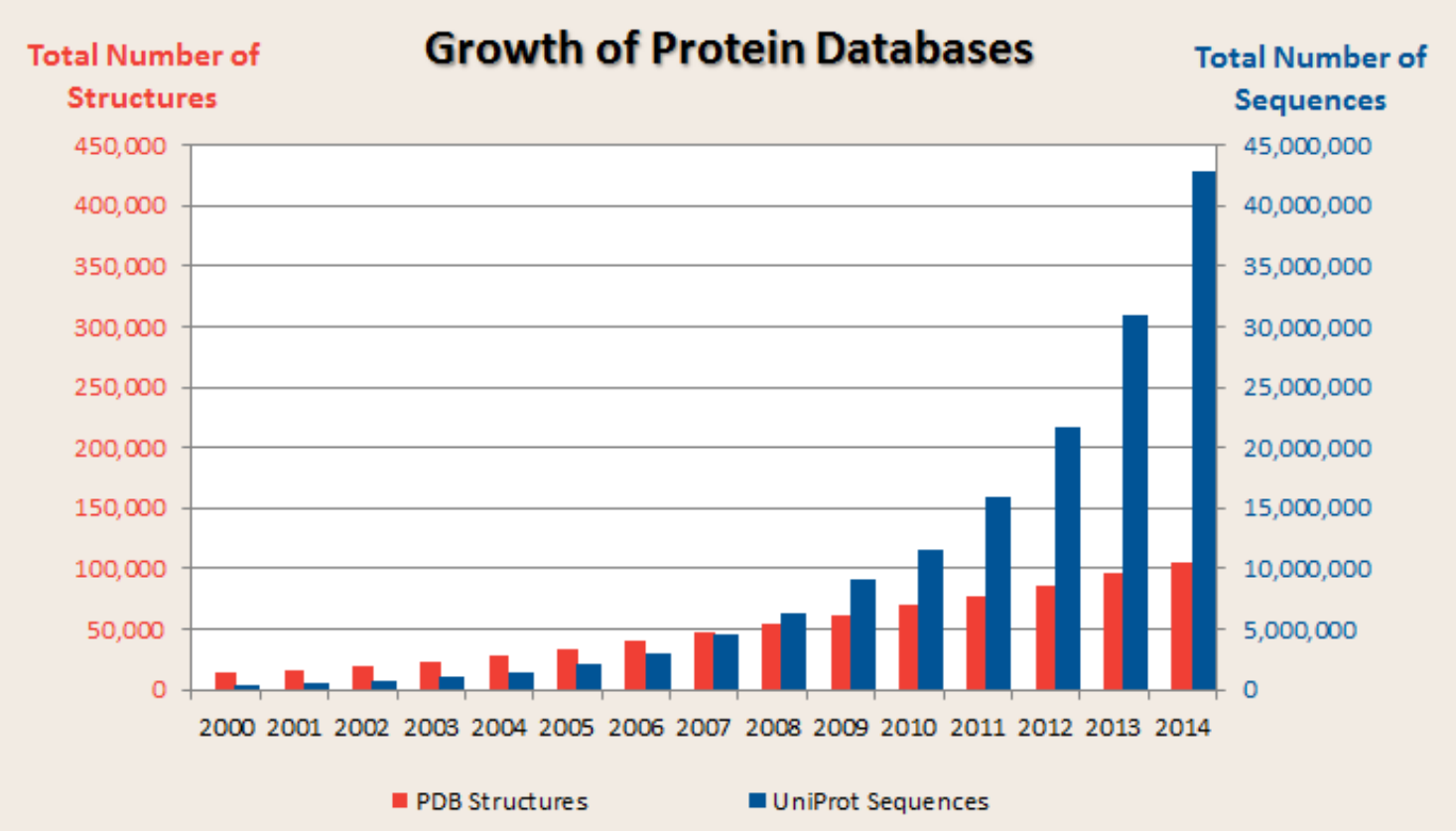
RCSB PDB
PROTEIN DATA BANK

WORLDWIDE PDB
PROTEIN DATA BANK

Research Collaboratory for Structural Bioinformatics:
Rutgers and UCSD/SDSC



Sequence - structure



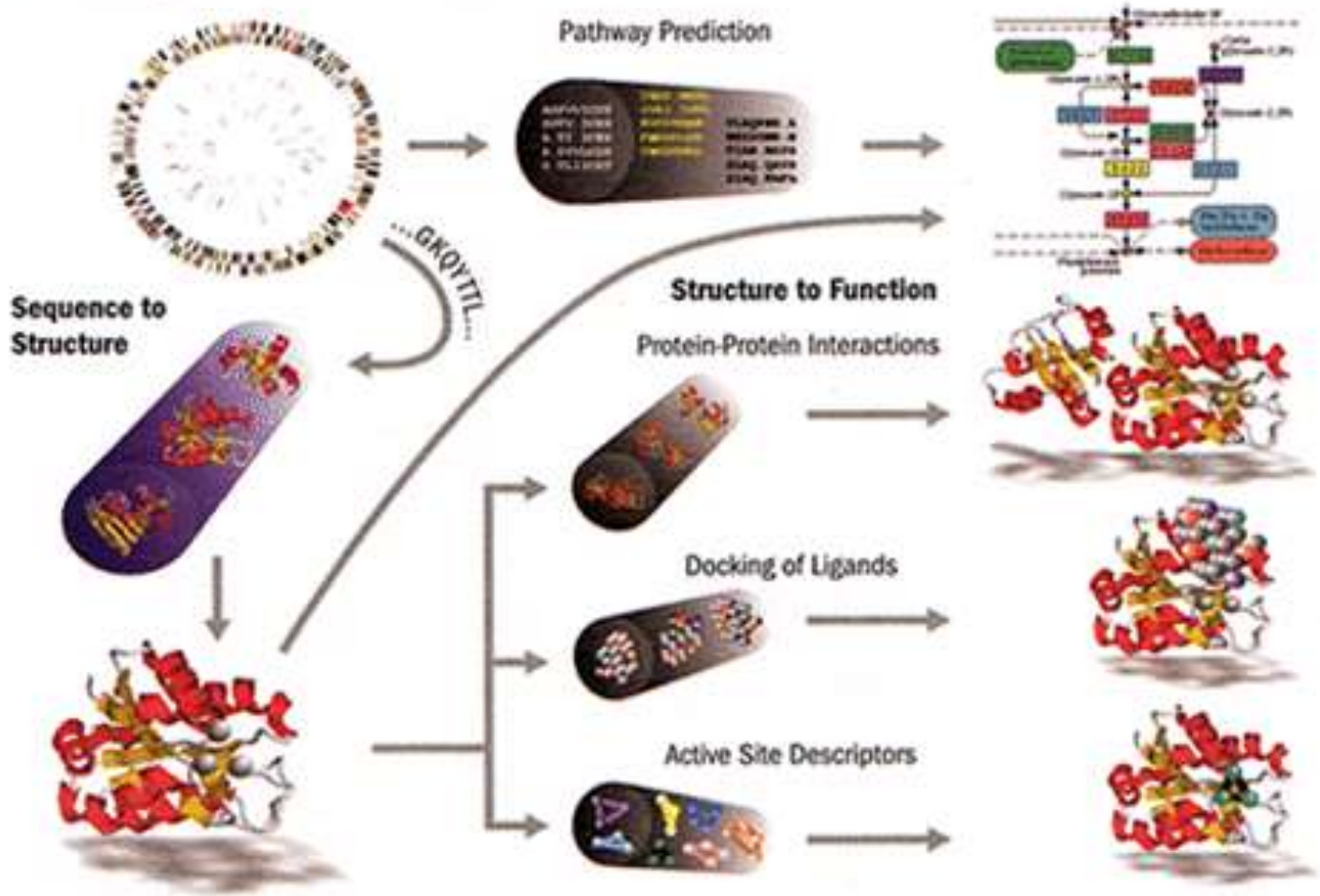
Protein Data Bank (PDB) - 140 000 protein structures

UniProtKB/TrEMBL sequence database - 133 507 323 nonredundant entries . Nov. 2018

Integrated Microbial Genomes & Microbiomes(IMG/M)database of 51 775 423 466 genes

(Coding genes *E. coli* - 4000, yeast – 6000, human, about -20000)

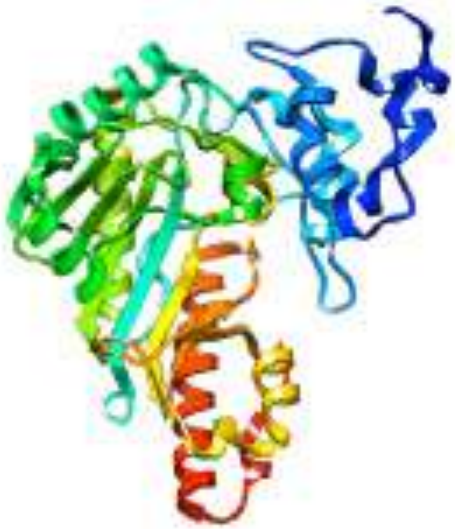
The Sequence-to-Structure-to-Function Paradigm

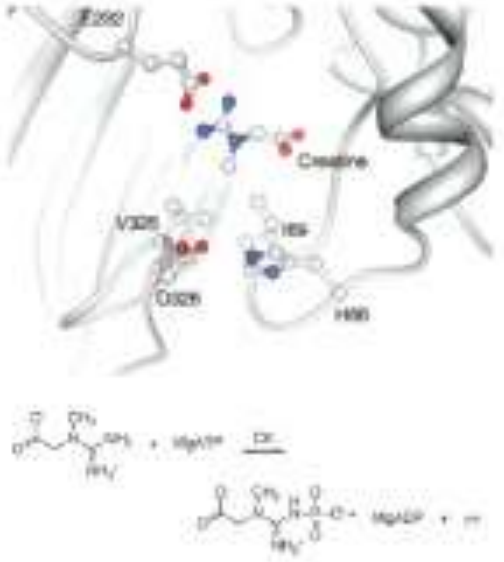


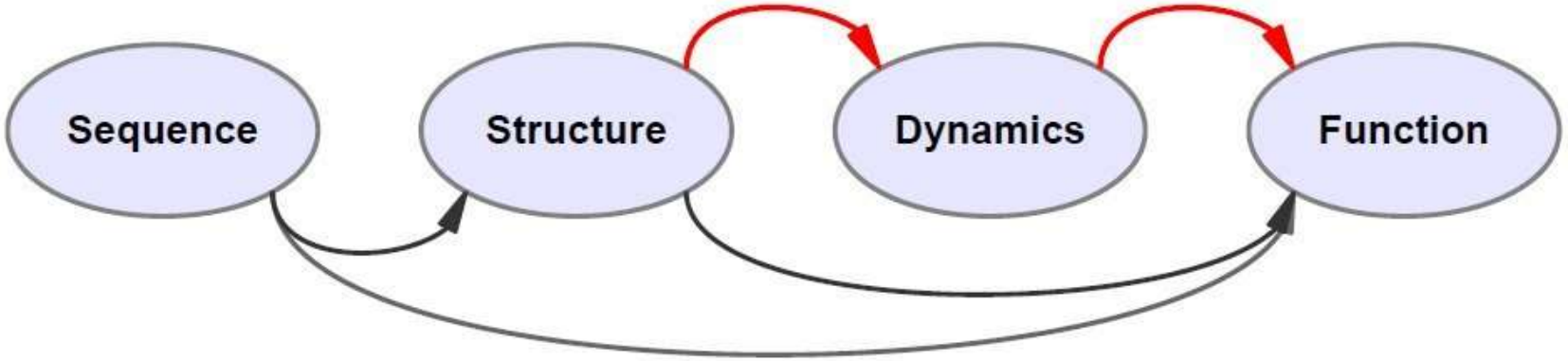
All the potential open reading frames (ORFs) in a protein sequence are threaded through a library of previously solved template protein structures. If a template is found, the structure is scanned for a match to a known active site. Alternatively, ligands can be virtually docked to identify the active site. Threading can also be used to identify potential interacting partners in the genome, or assist ORF pathway assignment.

Sequence → Structure → Function

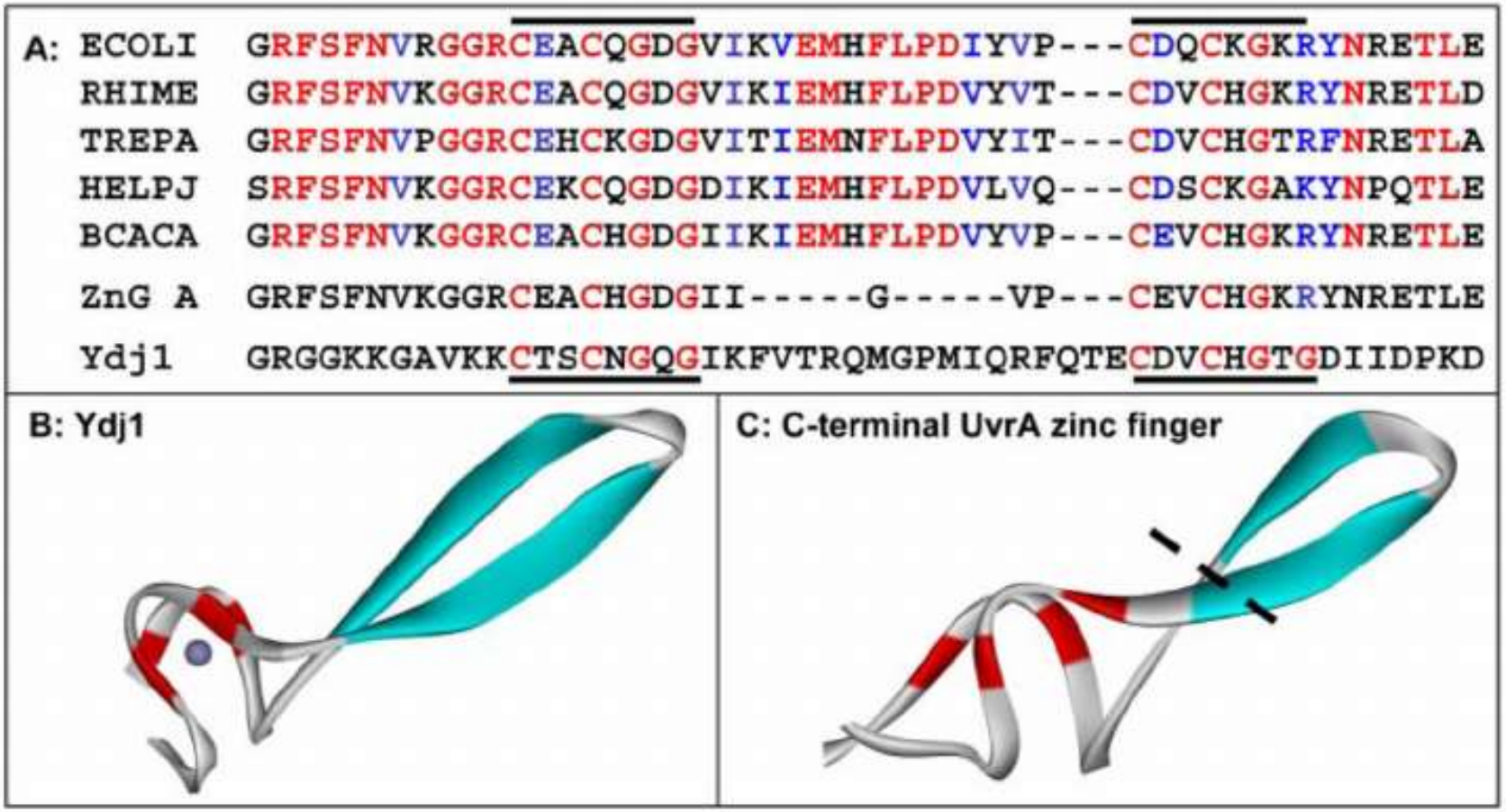
MPFGNTHNKFKL
NYKPEEEYPDLSK
HNNHMAKVLTLE
LYKKLRDKETPSGF
TVDDVIQTGVDNP
GHPFIMTVGCVAG
DEESYEVFKELFDPI
ISDRHGGYKPTD...







Structure – Comparative modeling



Local Alignment

Pairwise Sequence Alignment

Target Sequence

```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
              |||| |||||| ||||||||||||||||
Query Sequence 5' TACTCACGGATGAGGTACTTTAGAGGC 3'
```

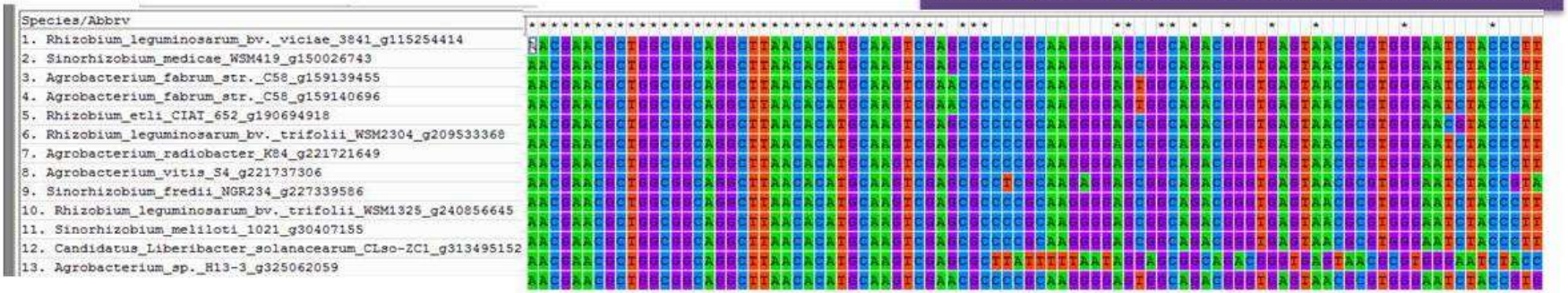
Global Alignment

Target Sequence

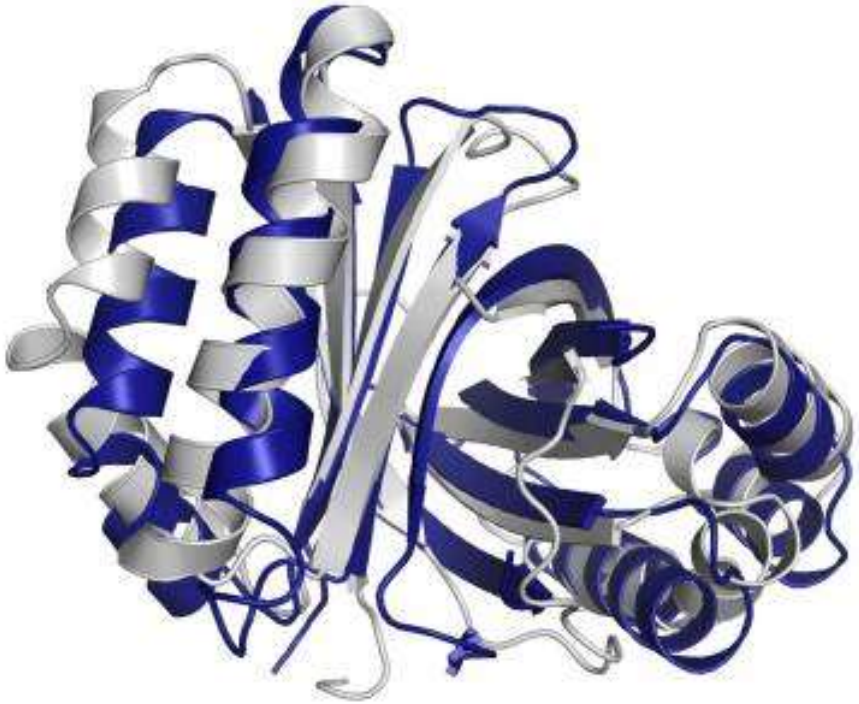
```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
   ||||||||||||   |||||||  |||||||||||||| |||||||
5' ACTACTAGATT----ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'
```

Query Sequence

Multiple Sequence Alignment (MSA)



Comparative (homology) modeling



Both cases (A,B) represent extremely distant homologies with sequence identity on the level of 10–12%

A

B

Comparative Modeling--Basic Protocol

1. Identification of homologue for target sequence
2. Alignment of target sequence to template sequence and structure
3. Side-chain modeling, copy the backbone of the template and model the new side chains onto this backbone
4. Loop modeling, for insertions and deletions in the alignment
5. Refinement of model -- moving template closer to target
6. Assessment of (predicted) model quality
7. Using the model to explain experiments and guide new ones

David F B, Charlotte M D, Hampapathalu A N, Nuria C, An Iterative Structure-Assisted Approach to Sequence Alignment and Comparative Modeling, PROTEINS: Structure, Function, and Genetics Supplementations, 3, pp. 55-60.

Sources of errors

- experimental errors and uncertainties in X-ray, NMR

1Å
100%

Applications

- studying catalytic mechanism / function

- side-chain packing
- mis-placed side-chains

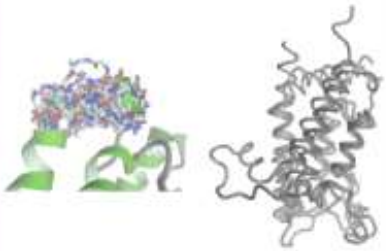
1.5Å
95%

- structure-based drug design, ligand docking

- modeling of loop regions (insertions and deletions)

- distorsions of aligned regions

60%

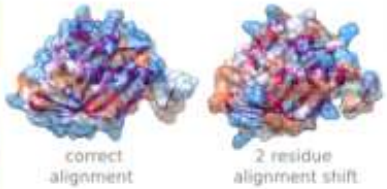


- structural support for mutagenesis studies

- molecular replacement

- alignment errors

3Å
40%



- integrative modeling

- modeling into low-resolution density maps

- sub-optimal template selection

- model may even have the wrong fold

>3Å
<30%

- domain boundaries

- identification of structural motives

Protein folding problem

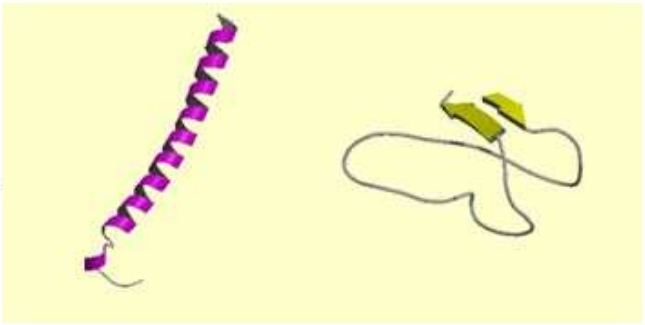
PRIMARY STRUCTURE (amino acid sequence)

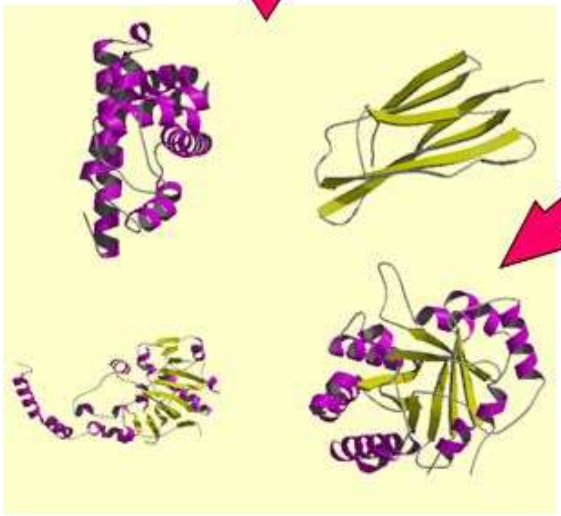
VHLTPEEKSAVTALWGKVNVDE
VGGEALGRLLVVYPWTQRFFE
SFGDLSTPDAVMGNPKVKAHG
KKVLGAFSDGLAHLDNLKGTFA
TLSELHCDKLHVDPENFRLLGN
VLVCVLAHHFGKEFTPPVQAAY
QKVVAGVANALAHKYH

1-step process

2-step process

SECONDARY STRUCTURE (helices, strands)





TERTIARY STRUCTURE (fold)

Each protein sequence "knows" how to fold into its tertiary structure. We still do not understand how and why

The 1-step process is based on a hydrophobic collapse; the 2-step process, more common in forming larger proteins, is called the framework model of folding

How to solve the Holy Grail problem