

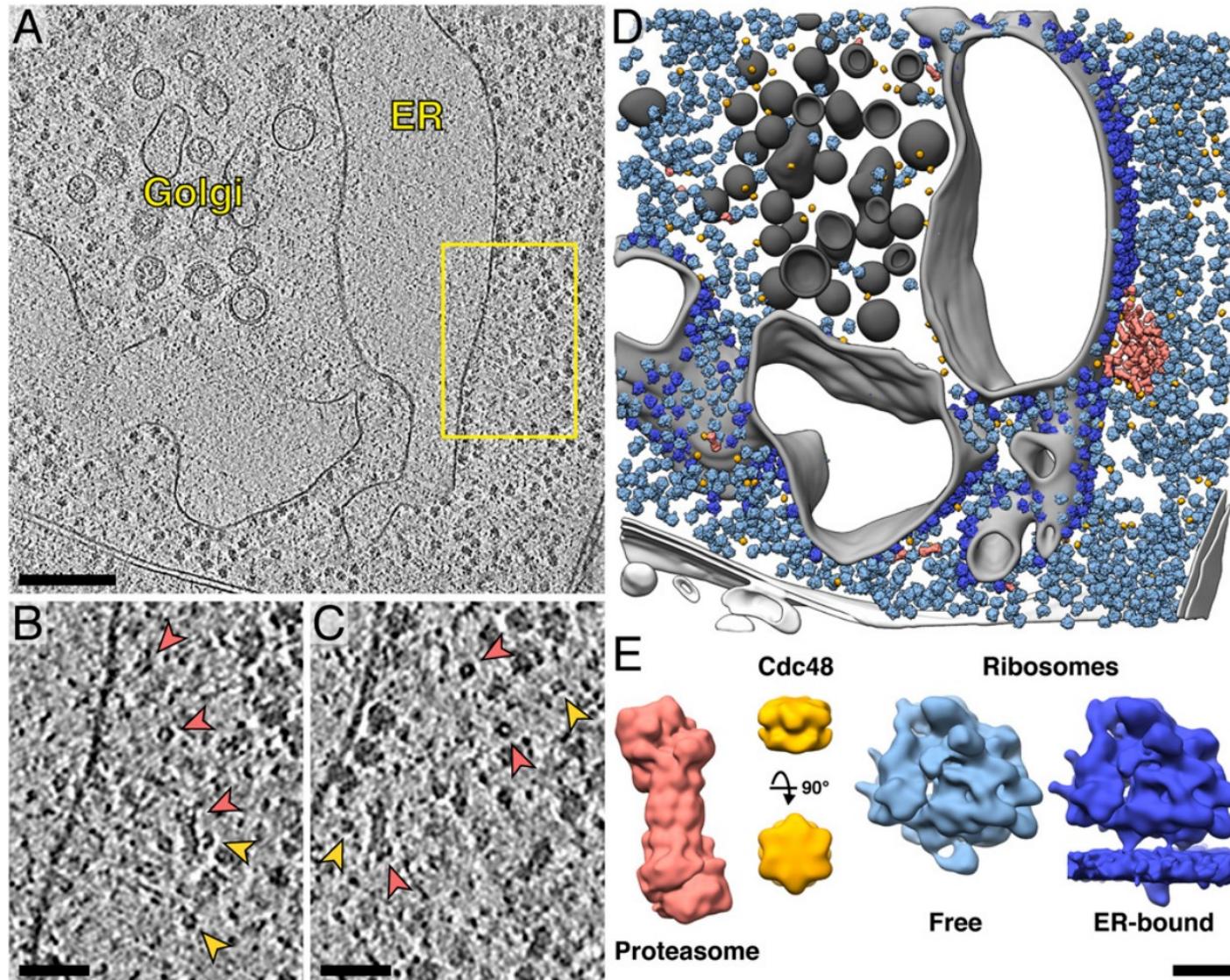
Bioinformatics III

Analysis and prediction of 3D macromolecule structures

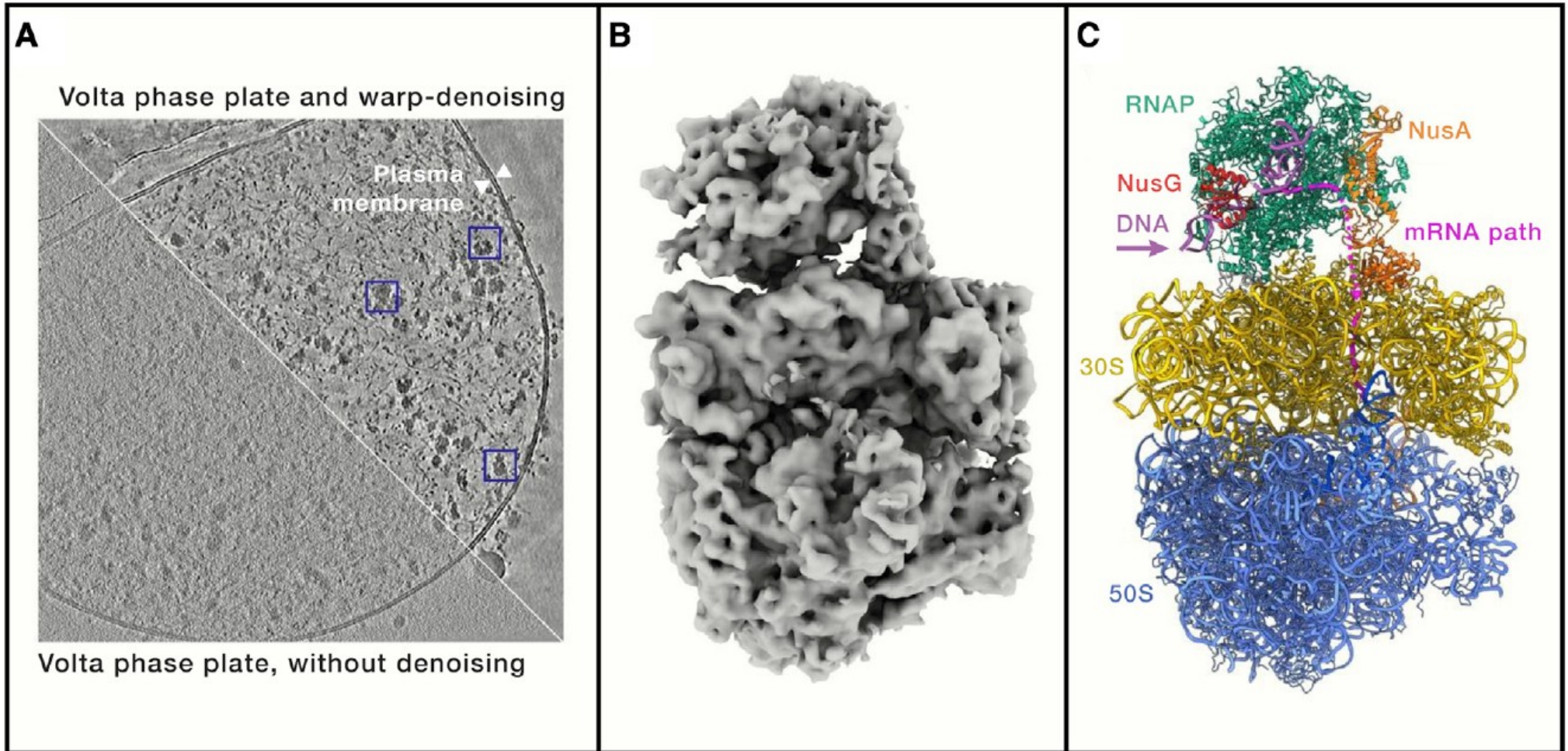
Lecture 1
structural organisation of proteins

Saulius Gražulis
2023 m.

Life at a glance



Resolution levels



Function and structure of living organisms

- The functions of life depend on the interactions of biological **(macro)molecules**:
 - **proteins**
 - **nucleic acids**
 - **polysaccharides**
 - lipids
 - small organic molecules
 - ions (salts)
 - **water**
- Diagram illustrating the classification of biological molecules:
- Proteins, nucleic acids, and polysaccharides are grouped as **polymers**.
 - Proteins and nucleic acids are further classified as **hetero-polymers**.
 - Polysaccharides are further classified as **hetero- and homo-polymers**.

The importance of protein structures

- 3D protein structure depends on their sequence
- **Protein function depends on their 3D structure**
- For most proteins, their 3D structure is practically unique in its native state
- Forces that shape protein 3D structures are weak non-covalent interactions

Levels of protein organisation

- Primary structure (a.a. sequence)
- Secondary structure (alpha-helices, beta-strands, beta sheets, loops and turns)
- Ternary structure (protein globule, domains)
- Quaternary structure (larger complexes of protein molecules).

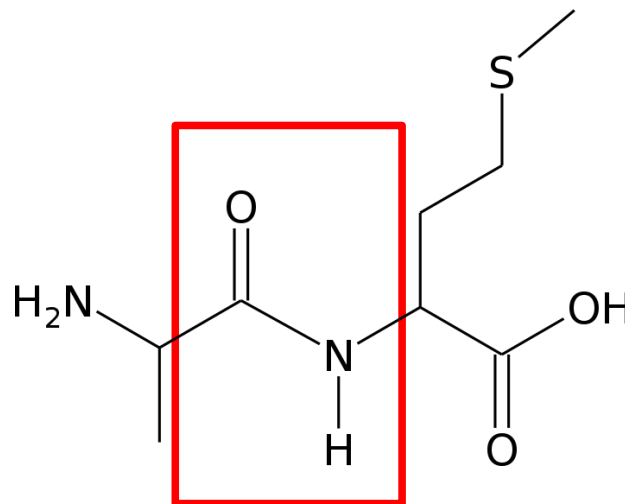
Proteins have hierarchical structure that is important to understand their function.

Primary structure

- Primary structure = sequence of amino acid (a.a.) residues E.g.: P61823.fasta:

```
>P61823|RNAS1_BOVIN Ribonuclease pancreatic - Bos taurus (Bovine).  
MALKSLVLLSLLVLVLLLVRVQPSLGKETAAAKFERQHMDSSSTAASSSNYCNQMMKSRN  
LTKDRCKPVNTFVHESLADVQAVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPN  
CAYKTTQANKHIIIVACEGNPYVPVHFDASV
```

Peptide
bond:

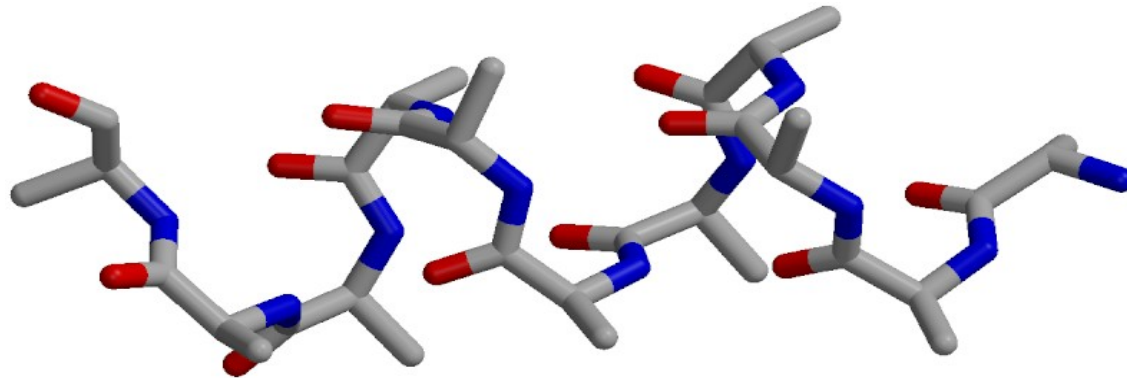


Secondary structure

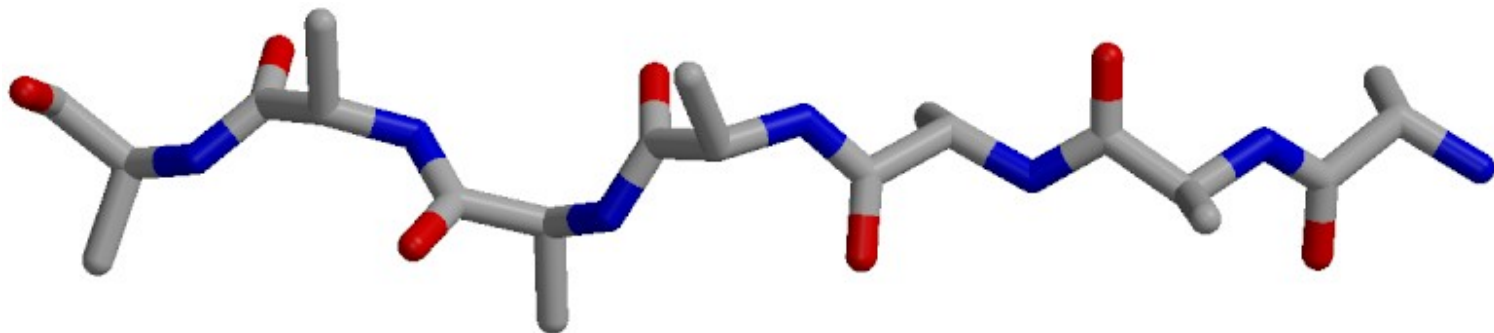
- Local structural elements of proteins:
 - alpha-helices,
 - beta-strands,
 - beta-sheets,
 - loops and turns.

Alpha-helices and beta-strands

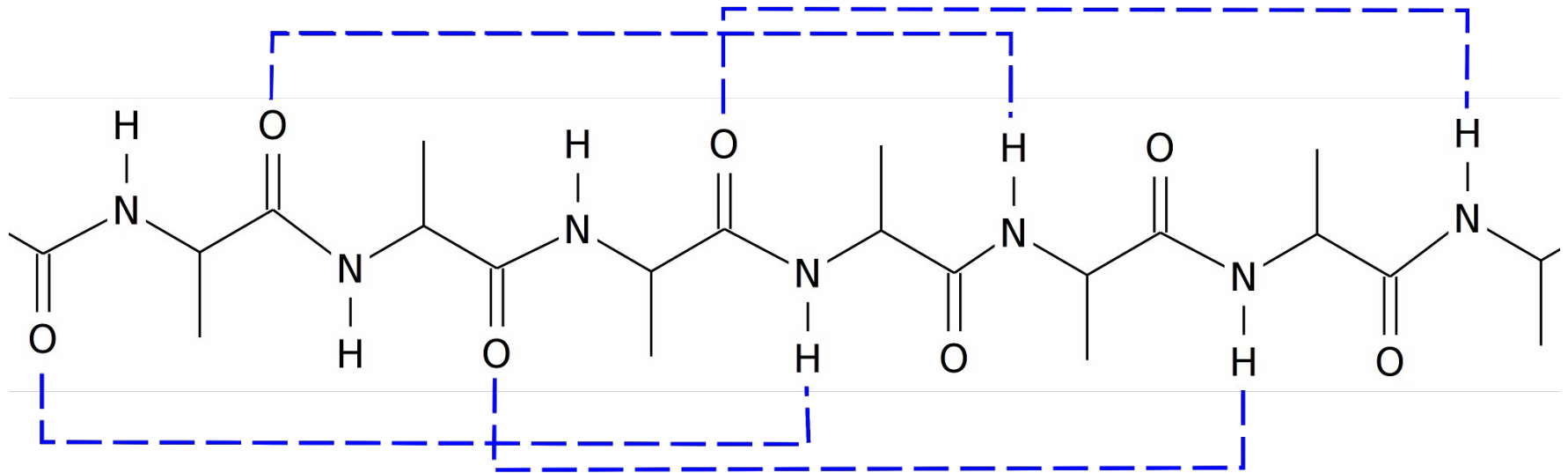
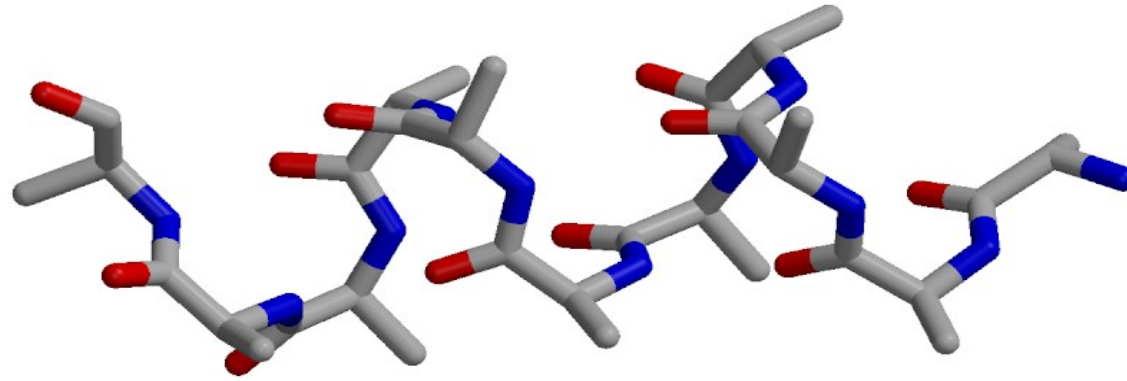
Alpha-helix



Beta-strand

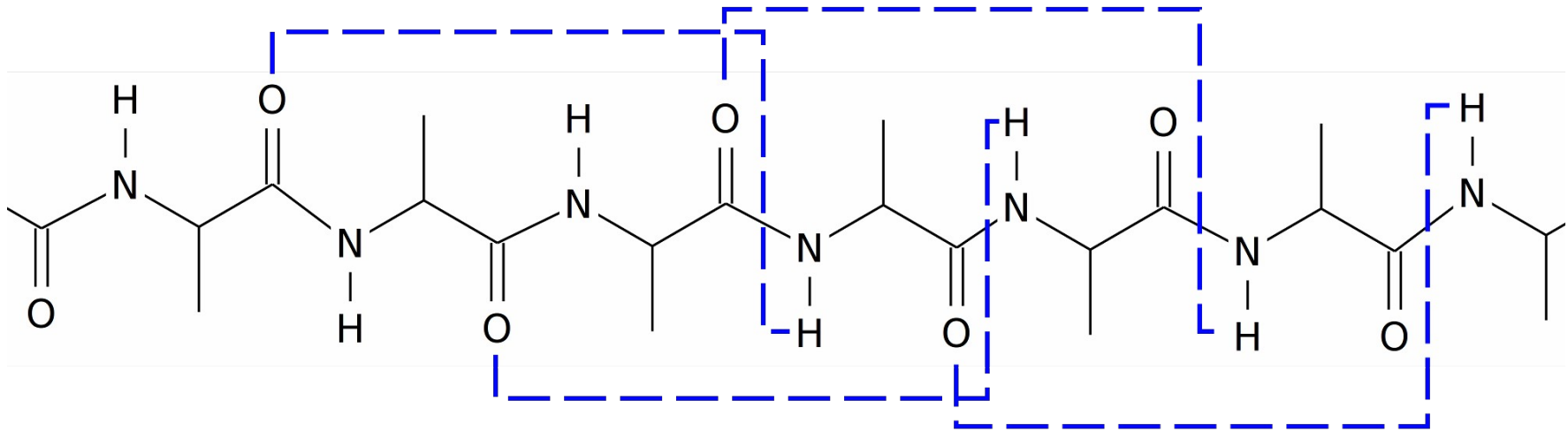


Structure of alpha-helices



α -helix == 4_{13} helix

Other helices



3_{10} helix

We also encounter in nature: 2_7 ; 5_{16} , or π – helices

(Финкельштейн & Птицын, Физика белка (2005), стр. 87)

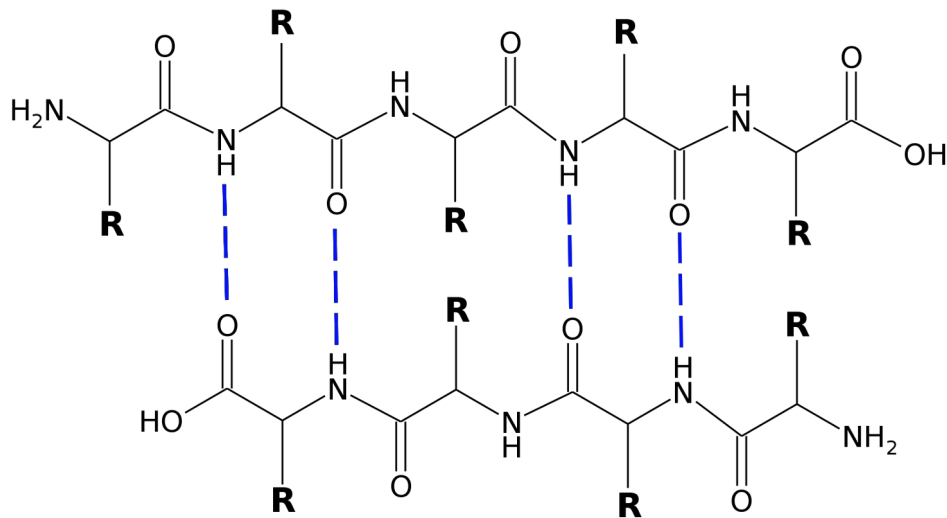
http://en.wikipedia.org/wiki/Alpha_helix

http://en.wikipedia.org/wiki/3_10_helix

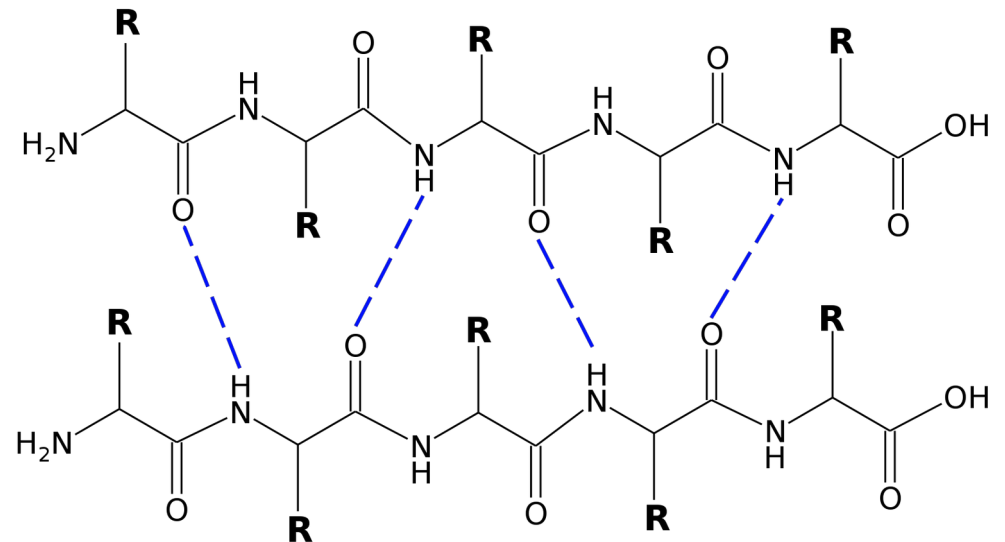
http://en.wikipedia.org/wiki/Pi_helix

Structure of beta sheets

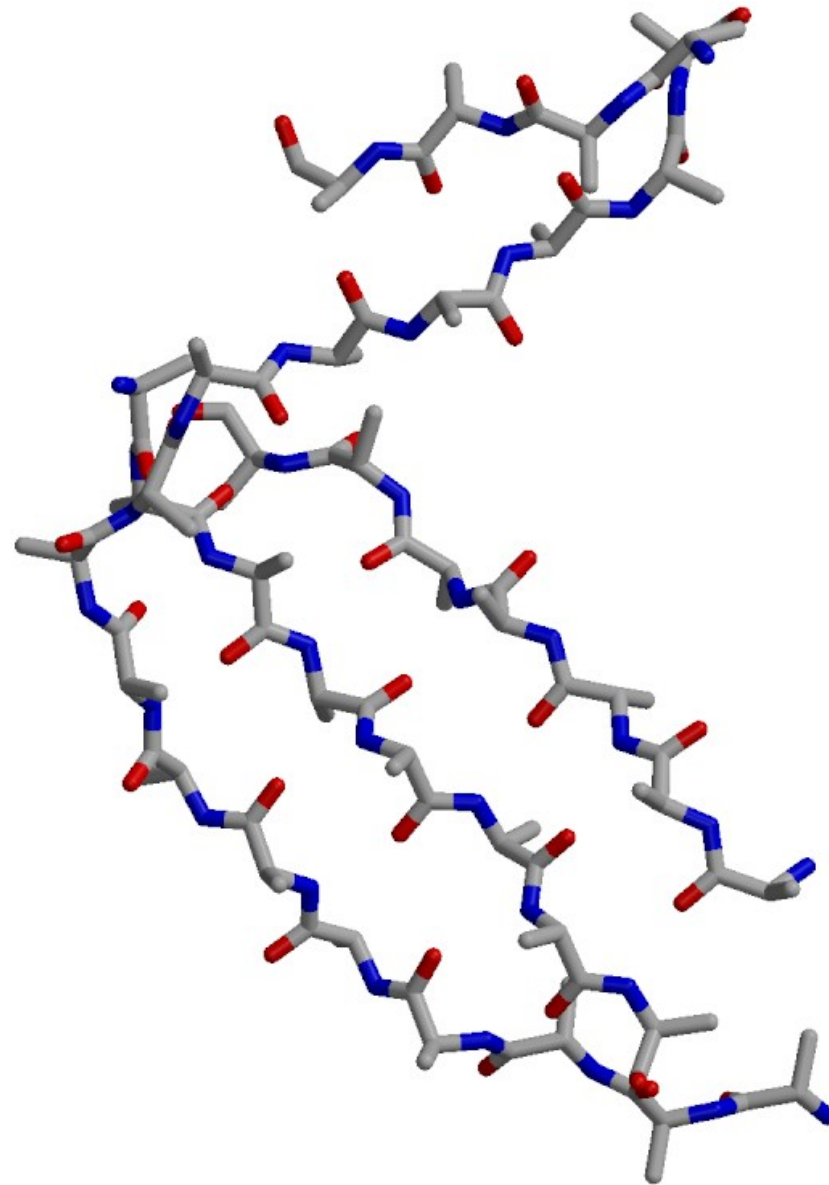
- Anti-parallel β -sheets



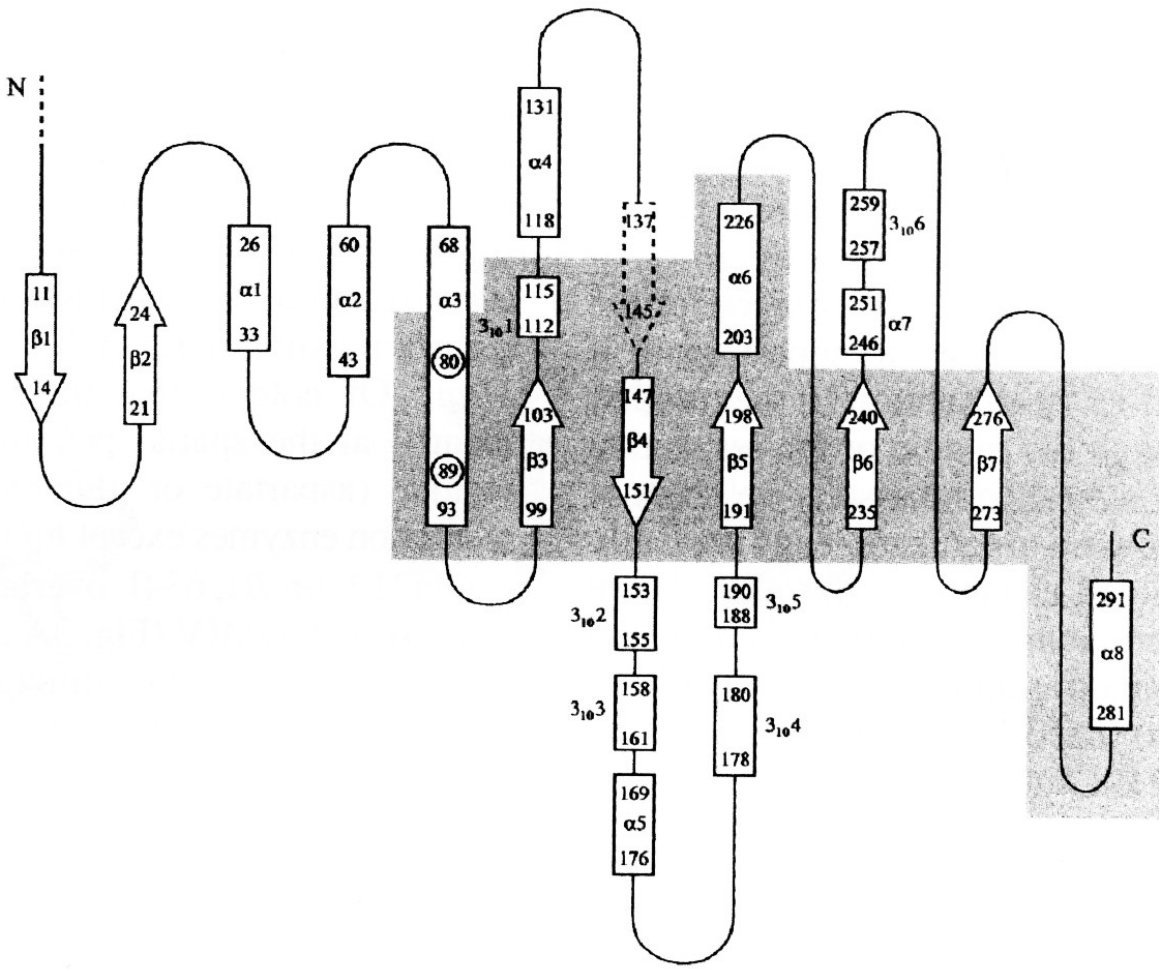
- Parallel β -sheets



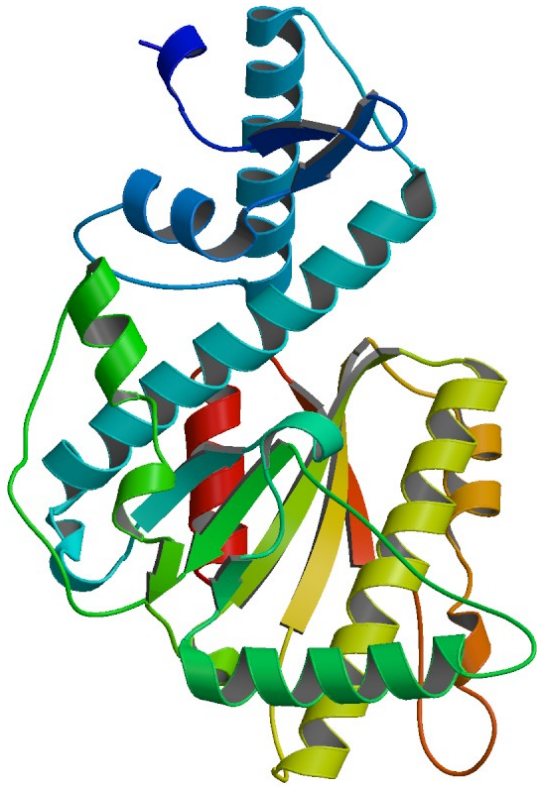
Beta-sheets



Topology

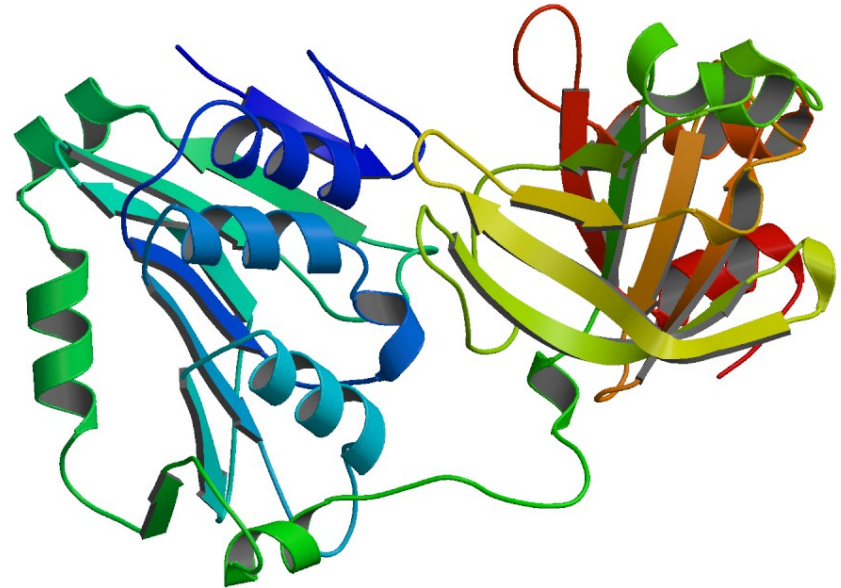


Tertiary Structure



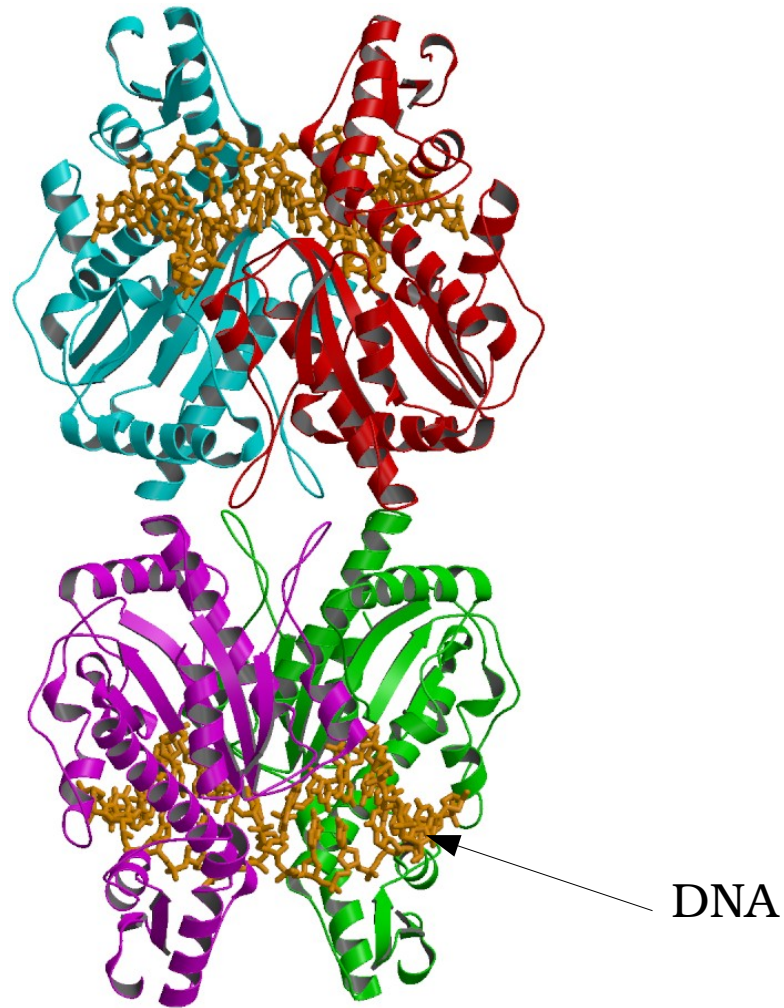
Bse634I restriction endonuclease
Gražulis *et al.* NAR 2002 p.876

BfiI restriction endonuclease
Gražulis *et al.* PNAS 2005 p.15797



Quaternary structure

Bse634I REase
tetramer in a crystal

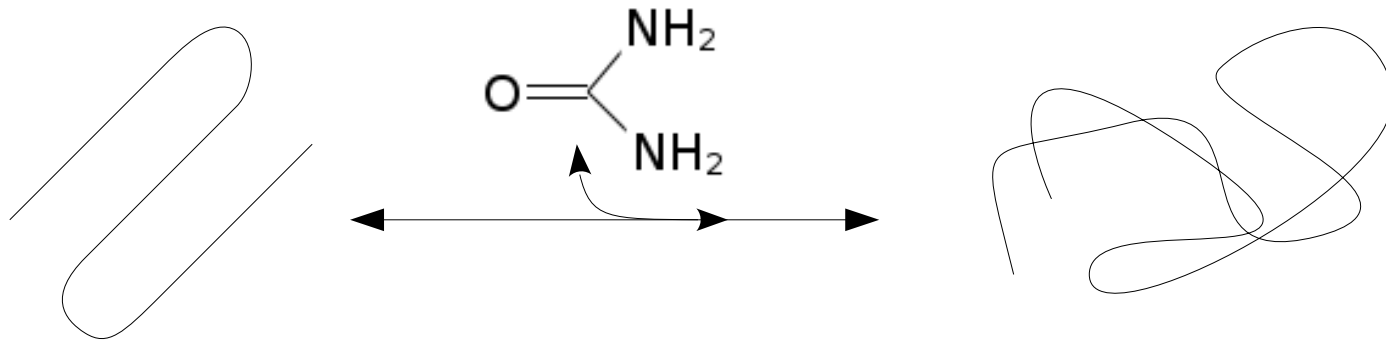


Main tenet of the (philosophy) of Science

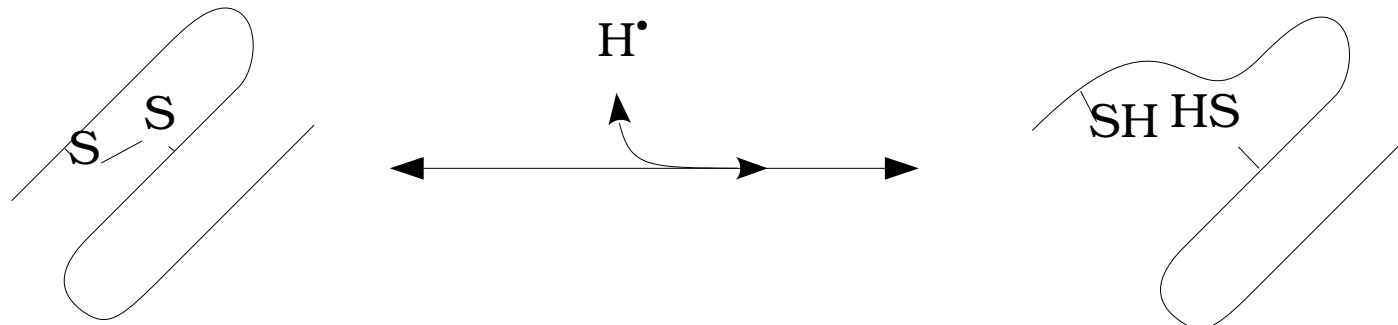
- Theories must be consistent, elegant, simple; however...
- ... their correctness/applicability is only validated by observations (experiments).

Investigations of sequence-structure relations in proteins

1. Urea denatures proteins:



2. Oxidising creates, and reduction breaks Cys S-S bridges:



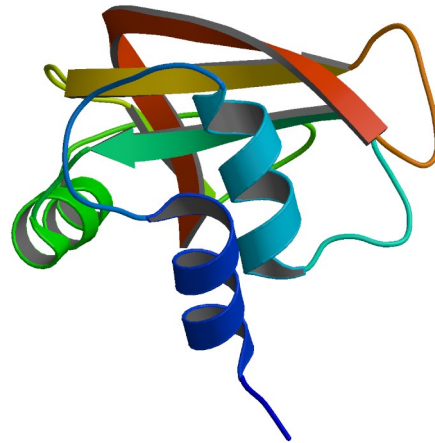
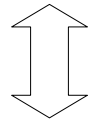
Anson & Mirsky JGP 1931 p.341, Anfinsen PNAS 1961 p.1309

Sequence-structure relations in proteins

- Anfinsen (Christian B. Anfinsen) experiments:
 - RNase A is fully unfolded in an 8M solution;
 - if the enzyme is oxidised **before** reducing urea, 1% of its activity is restored;
 - if the urea is removed and **after** that the protein is oxidised, 90% of activity is restored;
 - the same results are obtained if a chemically synthesised protein chain is used.

Sequence-structure relation for proteins

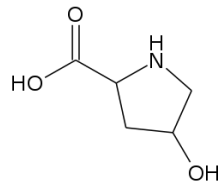
KETAAAKFERQHMDSSSTAASSSNYCNQMMKSRNLTKDRCKPVNTFVHE
SLADVQAVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKT
TQANKHIIVACEGNPYVPVHFDASV



Conclusion: protein structure and protein function are essentially determined by the a.a. sequence.

Other protein structures: fibre (fibrous) proteins

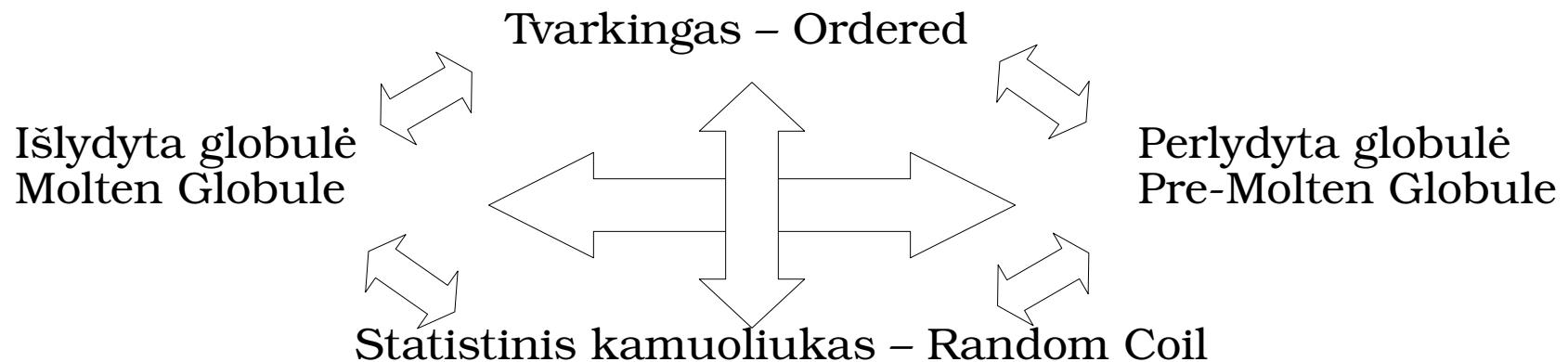
Protein	Structure	Cross-links
α -keratin	long α -helix	-Cys-Cys-
collagen (GXP GXH) _n , H = hydroxyproline	triple (!) left (!) helix (only encountered in collagen)	-Nor ¹ -Lys-
elastin	„Gly-helix“ (typical for elastin)	-Nor-Lys-, Lys- } Des ² { -Lys Lys- } -Lys
phage „steam“ protein	„Beta-helix“	



¹(Nor = norleucine) ²(Des = desmosine)

Naturally unfolded proteins

- Naturally unfolded proteins:
 - molten globules
 - pre-molten globules
 - random coils



Levinthal paradox

- Number of possible conformations of a short protein chain: 100 a.r., $\sim 2^{100}$ (assuming 2 possibilities per residue); for an exhaustive search we need:
- if 1 conformation is checked per 1 ps,
- π s = 1 nano-century (3.14 s = 10^{-7} years)
- 2^{100} ps $\approx 10^{30}$ ps $\sim \mathbf{10^{10}}$ **years.**

Exhaustive search is infeasible for both real protein and for a computer

Methods of protein structure determination

- Experimental:
 - X-ray diffraction, NMR, EM
- Theoretical:
 - *Ab initio*
 - **Machine learning**
 - Homology modelling
 - Fold recognition
 - Fragment methods
 - Protein complex prediction

Other modes structure prediction

- Sparse data predictions
 - secondary structure
 - contact analysis
- Membrane protein prediction
- Complex prediction
- Small molecule docking, rational drug design

Resources on the Web

Protein Data Bank (PDB)

<http://www.rcsb.org/>

<https://www.ebi.ac.uk/pdbe/>

<https://pdbj.org/>

Nucleic acid database

<http://ndbserver.rutgers.edu/>

Crystallography Open Database (COD)

<http://www.crystallography.net/>