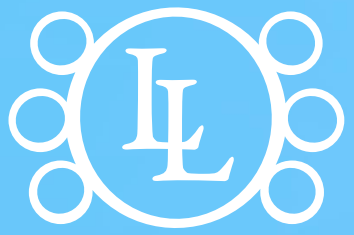


Introduction to the structure of macromolecules

Course information

- 11 lectures ≈ 22 h

1. Introduction to the structure of macromolecules
2. Structure of biomolecules
3. Bioinformatics databases
4. Structure prediction
5. Models of structures
6. Stability and dynamics of macromolecules
7. Analysis of protein structures
8. Protein-ligand complexes
9. Macromolecular complexes and interactions
10. Engineering of protein structures
11. Applications of structural biology and bioinformatics

Course information

- Lecturers
  - Sérgio Marques, PhD
    → Main lecturer

  - David Bednář, PhD
    → Lecture 6

  - Joan Planas, PhD
    → Lectures 3-5

  - Anthony Legrand, PhD
    → Lecture 9

Course information

- Examination
  - Written exam, multiple choices, 25 questions, 25 points
    - A: 25-22
    - B: 21-19
    - C: 18-16
    - D: 15-13
    - E: 12-10
    - F (fail): < 10
- 3 exam dates; you can attend them all
  - 10/17 Dec. 2024 (to be voted)
  - Jan. 2025
  - Feb. 2025
- Slides with essential information have the sign:

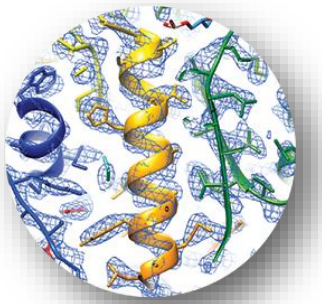
Course information

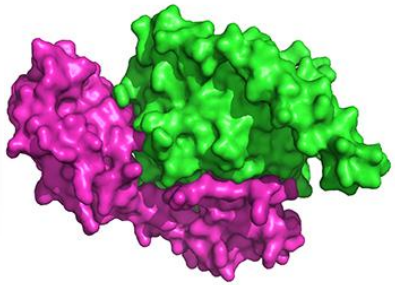
- Literature (provided)
  - Petsko, G. A. & Ringe, D. (2004). **Protein Structure and Function**, New Science Press, London.
  - Gu, J. & Bourne, P. E. (2009). **Structural Bioinformatics**, 2nd Edition, Wiley-Blackwell, Hoboken.
  - Widłak, W. (2013). **Molecular Biology - Not Only for Bioinformaticians**. Springer Berlin, Heidelberg
  - **Lecture slides** (uploaded every week)
  - Journal articles (not essential)
- Alternative literature (not provided)
  - Claverie, J-M., & Notredame, C. (2006), Bioinformatics for Dummies. Wiley Publishing, Hoboken
  - Xiong, J. (2006), Essential Bioinformatics, Cambridge University Press, New York.
  - T. Schwede & M. C. Peitsch (2008), Computational Structural Biology: Methods and Applications, World Scientific Publishing Company
  - Liljas, L. Liljas, J. Piskur, G. Lindblom, P. Nissen, M. Kjeldgaard (2009), Textbook Of Structural Biology, World Scientific Publishing Company

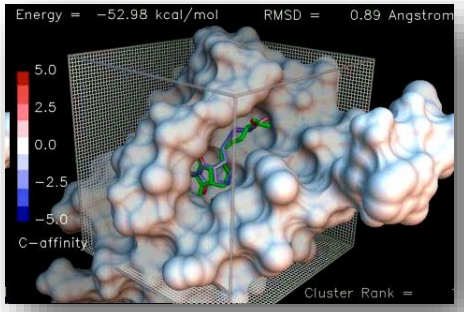
Structural biology - practice - **Bi9410cen**

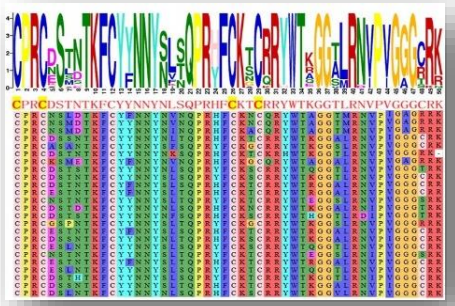
EN

- Semester: autumn
- Exercises: 2 hours/week
- Tutors: MUDr. J. Mičan, Mgr. J. Horáčková, Dr. S. Eyrilmez, Dr. A. Legrand
- Outline:
  - **Visualize** 3D structure of biomolecules
  - Obtain structures and relevant information from **databases**
  - **Analyze** function, stability and dynamics of biomolecules
  - **Predict** the structures of proteins and their complexes
  - Predict the **effects of mutations** and engineer protein properties





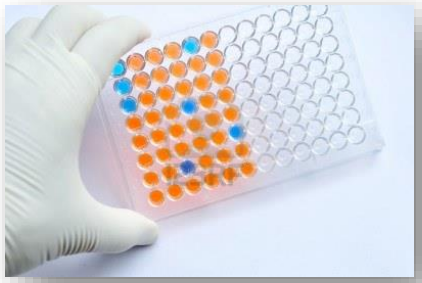
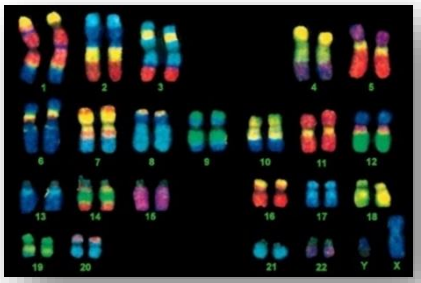
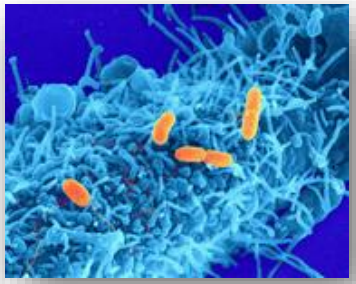




Molecular biotechnology - **Bi7430**

CZ

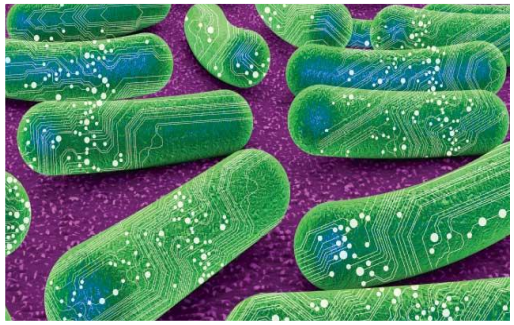
- Semester: autumn
- Lectures: 2 hours/week; exercises: 2 hours/week
- Lecturers: Dr. Z. Prokop, Dr. M. Marek, Dr. P. Dvořák, Dr. Š. Nevolová
- Outline:
  - Protein and metabolic engineering
  - Molecular diagnostics and modern vaccines
  - Cell and gene therapy and regenerative medicine
  - Molecular biotechnology in industry and agriculture

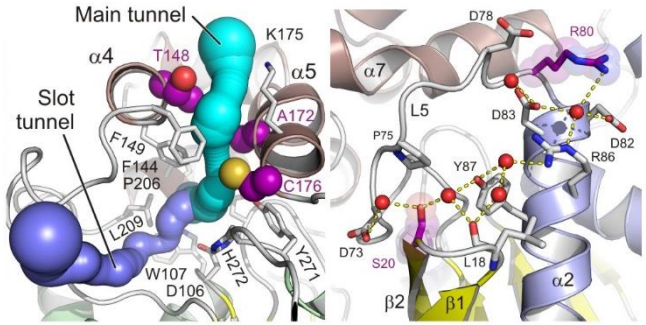





Synthetic biology - **S2015**

CZ

- Semester: autumn
- Lectures: 2 hours/week
- Lecturers: Dr. M. Marek, Dr. K. Říha
- Outline:
  - Engineering concepts in synthetic biology
  - From genetic engineering to synthetic genomes
  - Protein engineering and design, from proteins to nanomachines
  - Metabolic engineering, artificial organelles



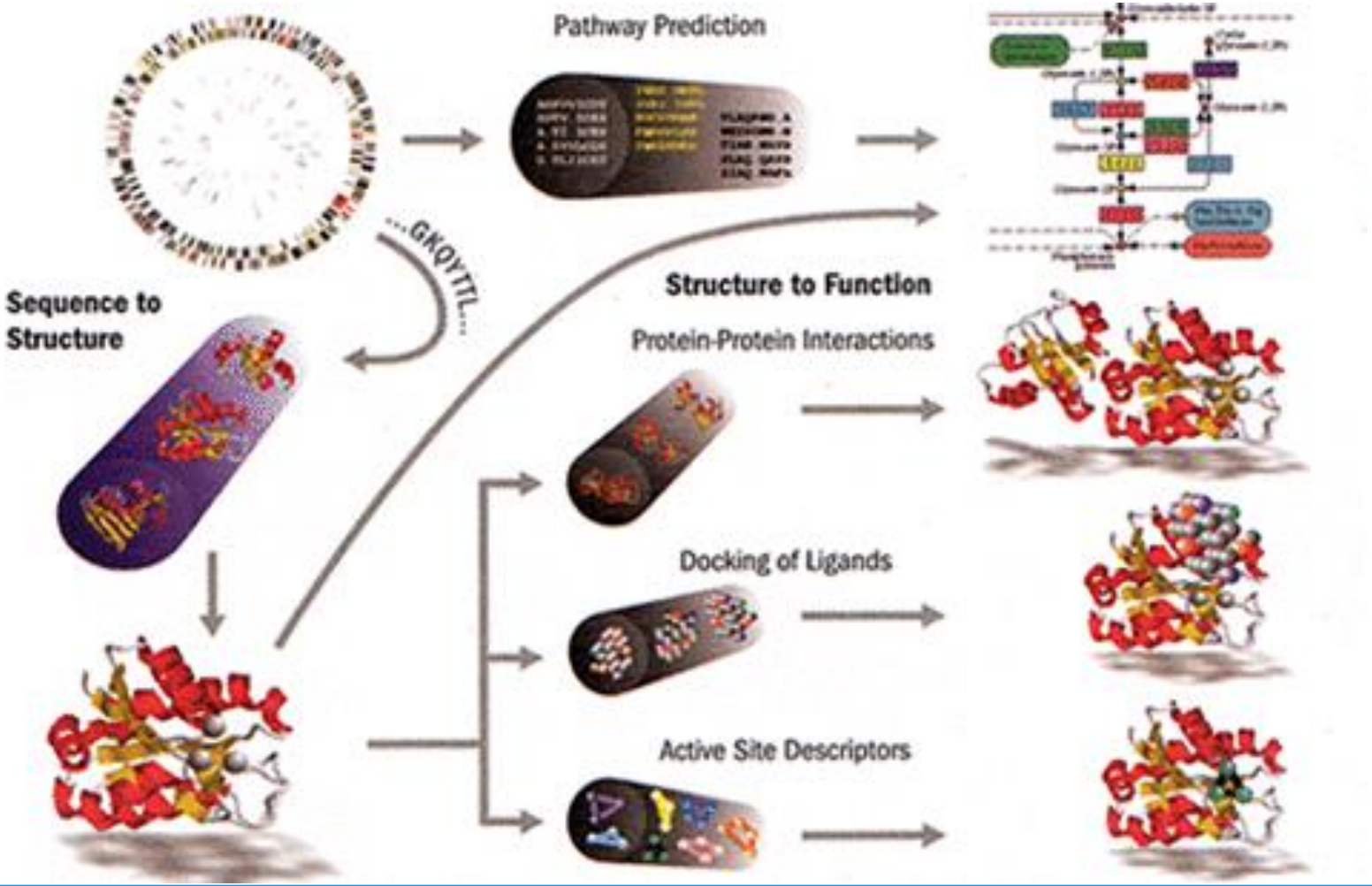


Outline

- Motivation
- What is Structural Biology and Bioinformatics
- Visualization of structure
- Energetics of structures
- Molecular interactions
- Determination of structure

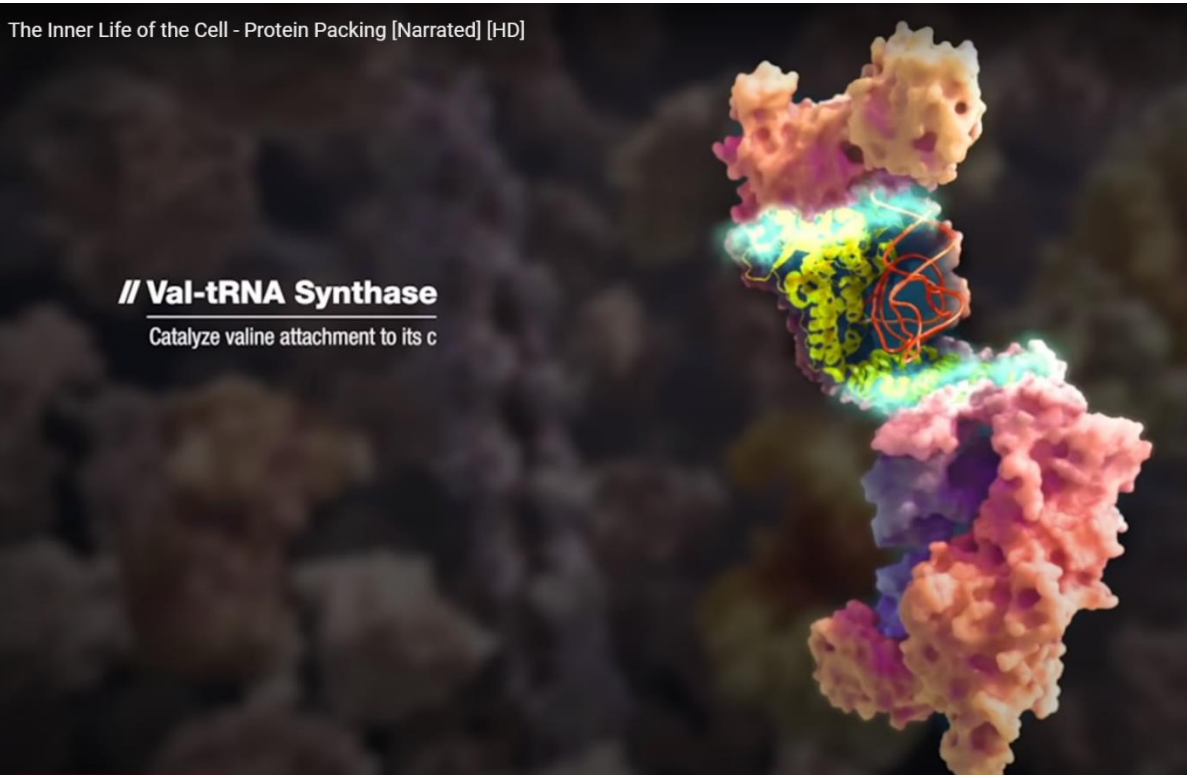
Motivation

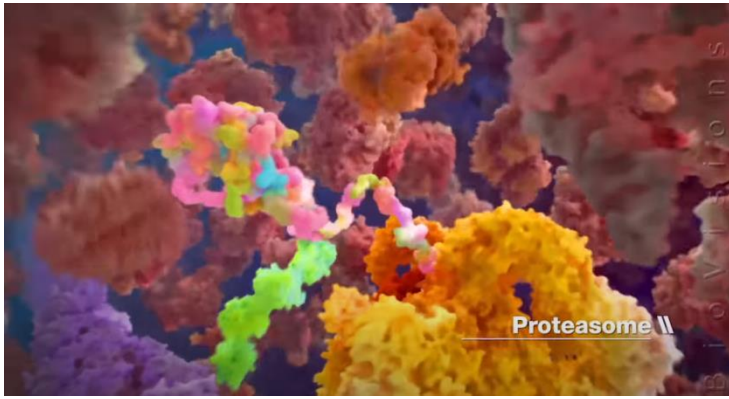
- **Sequence-structure-function** paradigm

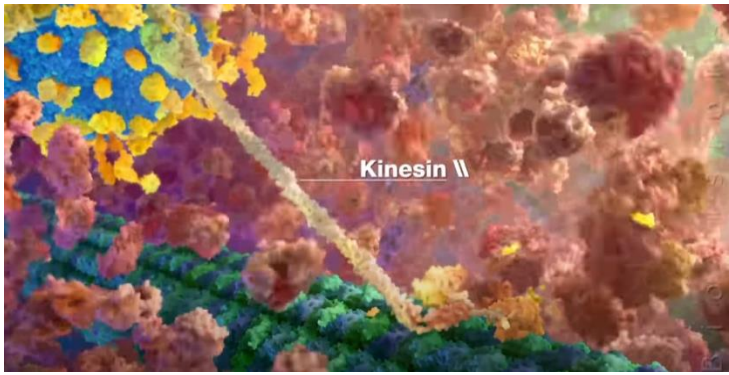


Motivation

- 3D structure ⇔ biological function







The inner life of the cell - XVIVO & Harvard University:
https://youtu.be/XOaiWl-nW1k

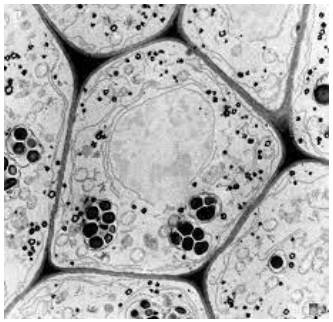
What is Structural Biology

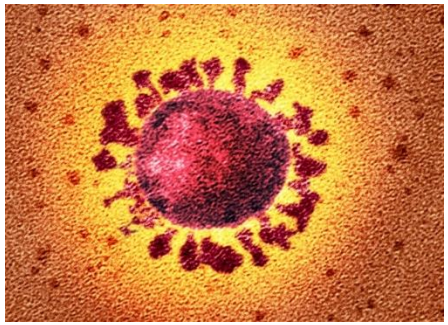
- **Structural biology** is the study of the molecular structure and dynamics of biological macromolecules, particularly proteins and nucleic acids, and how alterations in their structures affect their function

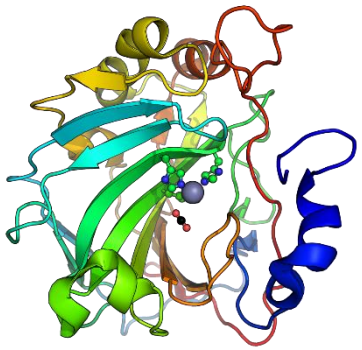
What is Structural Biology

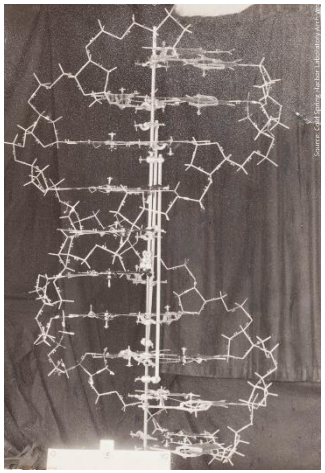
- Focused on the three-dimensional arrangement of biomolecules – the 3D structure – and their mutual interactions to understand their functions in the cell.

- Makes biological objects visible and understood
  - "Seeing is believing"
  - To understand, we need to see









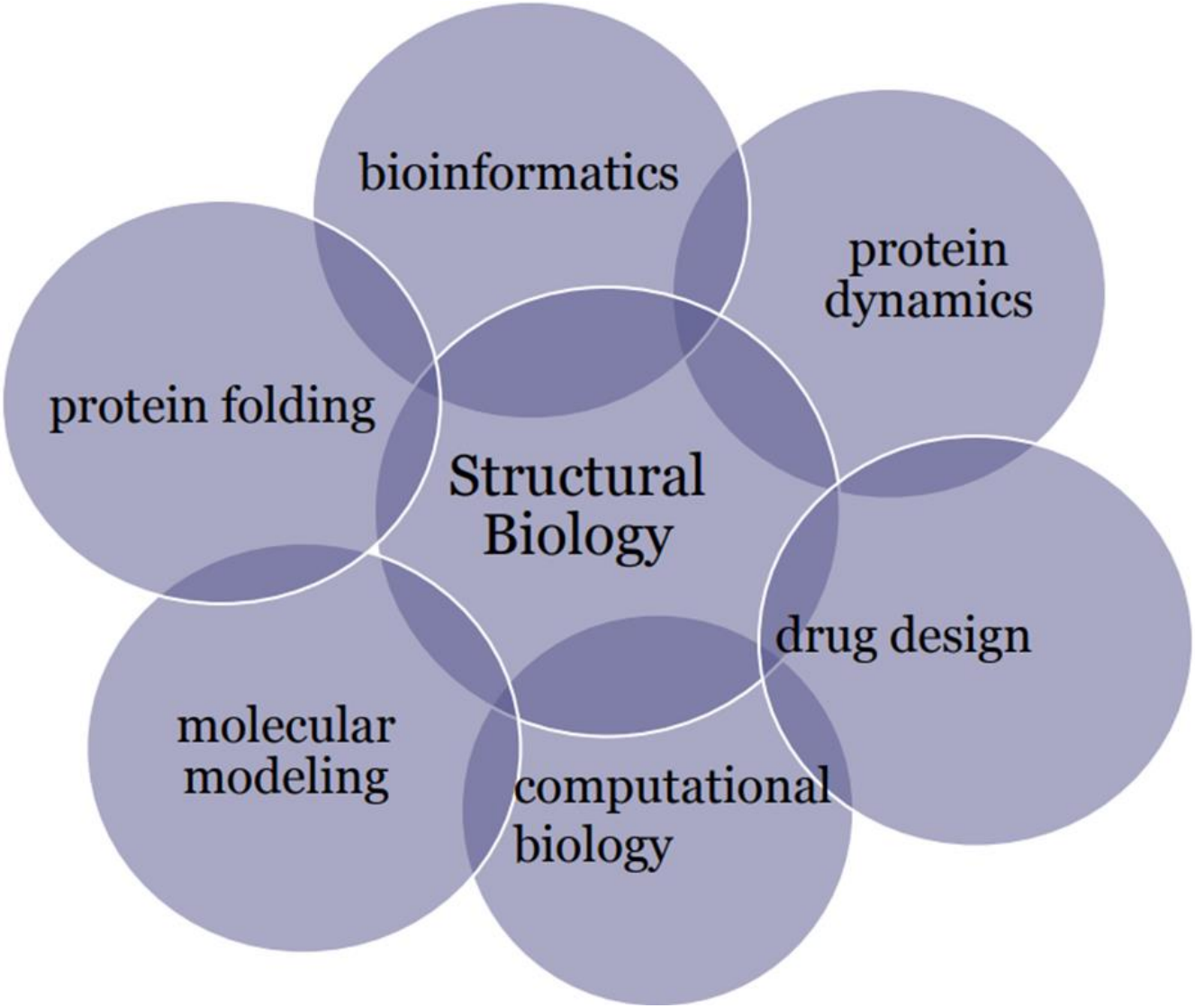
What is Structural Biology

- "Unfortunately, we cannot accurately describe at the chemical level how a molecule functions unless we first know its structure"

  James Watson, 1964

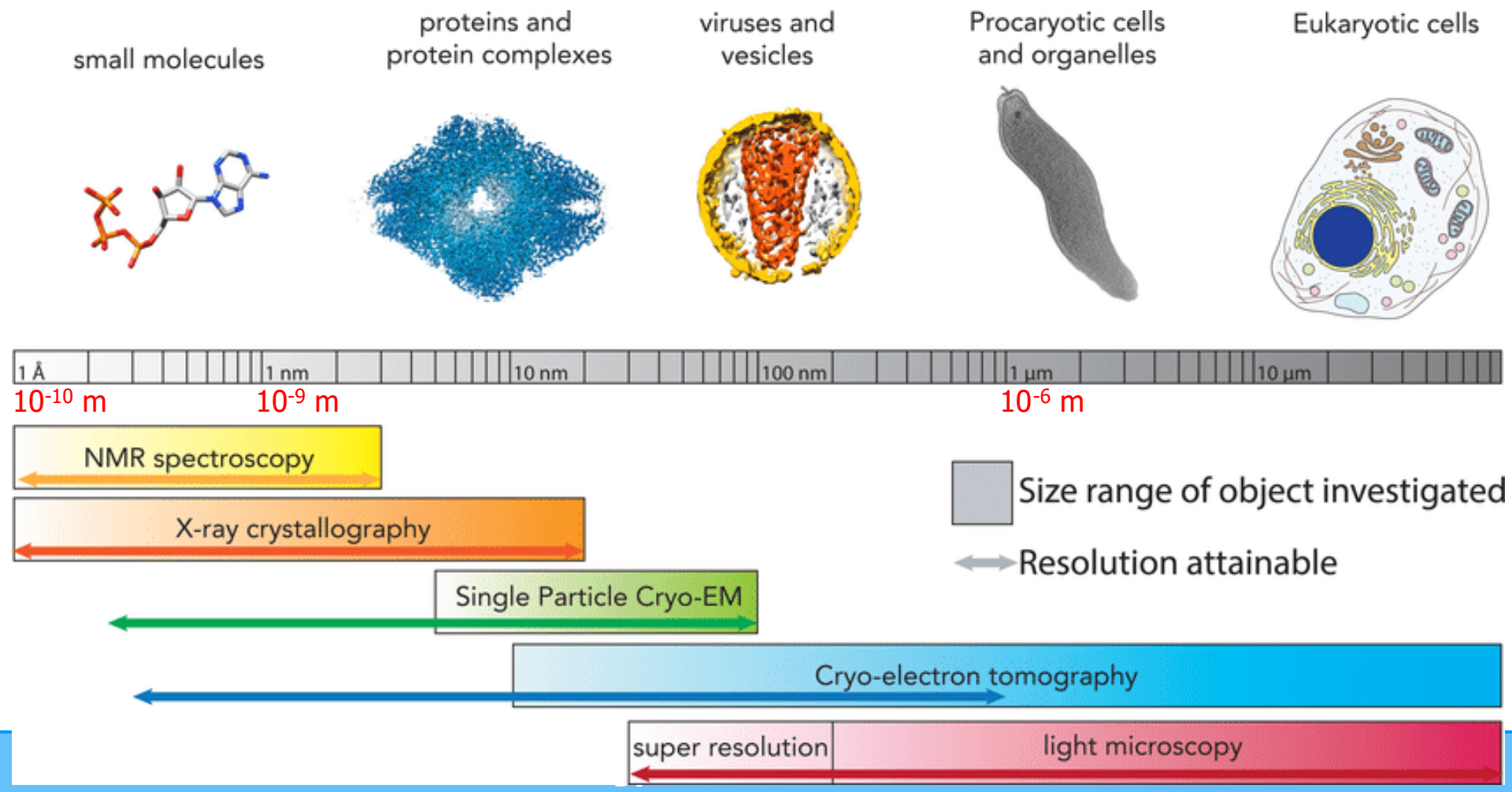
- Important milestones
  - 1838 – Protein discovery - Gerardus Mulder
  - 1869 – DNA discovery - Friedrich Miescher
  - 1953 – DNA structure - James Watson and Francis Crick
  - 1958 – Myoglobin crystal structure - John Kendrew
  - 1959 – Hemoglobin crystal structure - Max Perutz

What is Structural Biology



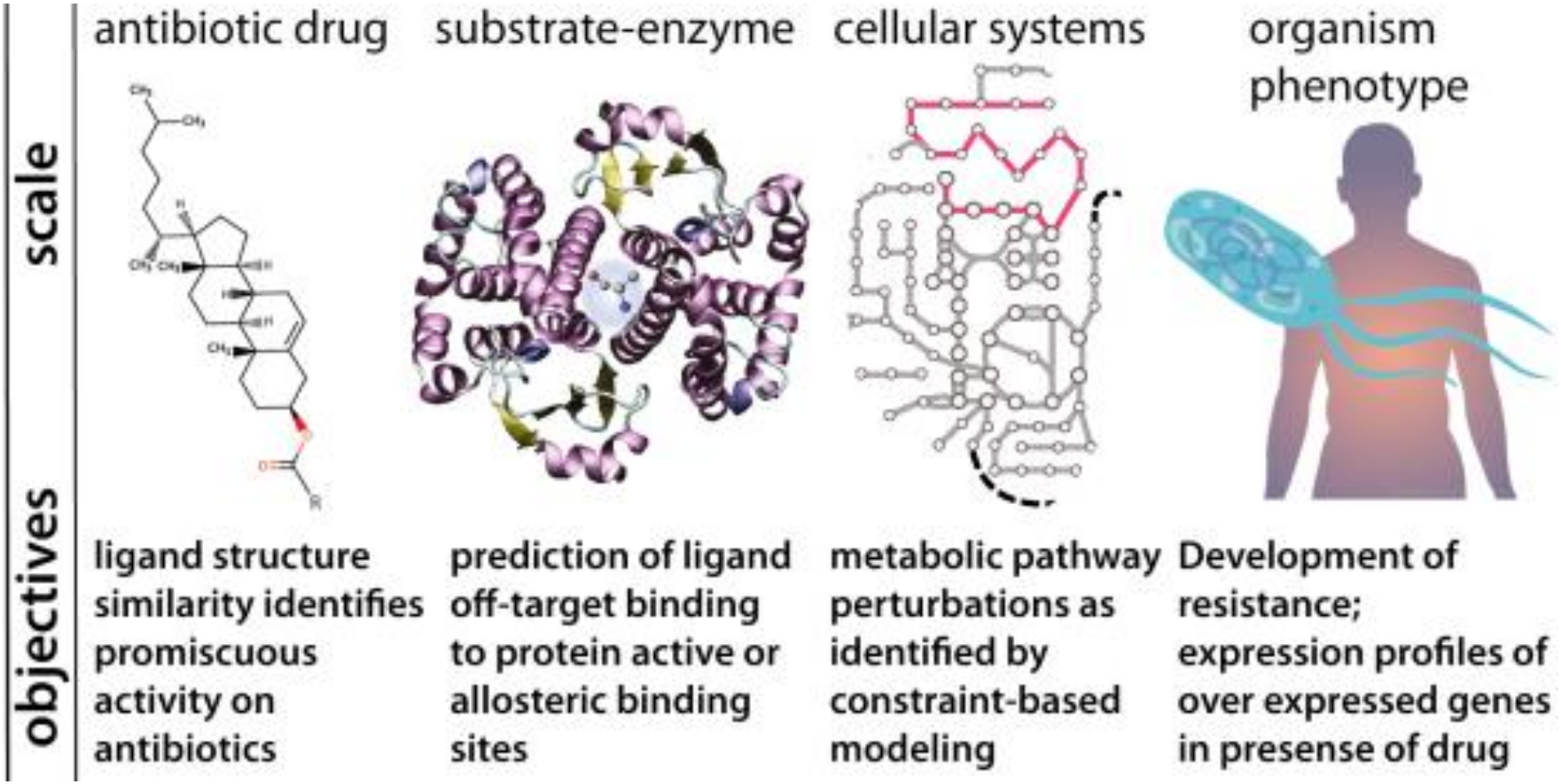
What is Structural Biology

- Several different scales



What is Structural Biology

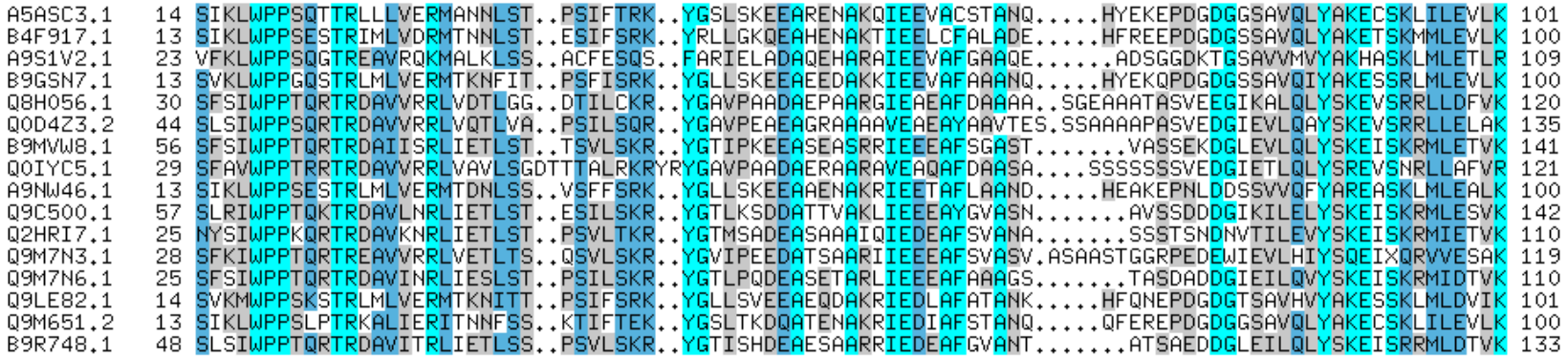
- Several different scales



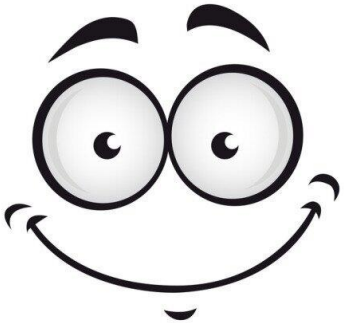
What is Bioinformatics

- **Bioinformatics** is an interdisciplinary field that develops methods and software tools for understanding biological data, in particular when the data sets are large and complex.

- Sequence analysis, genomics, proteomics, systems biology, structural bioinformatics



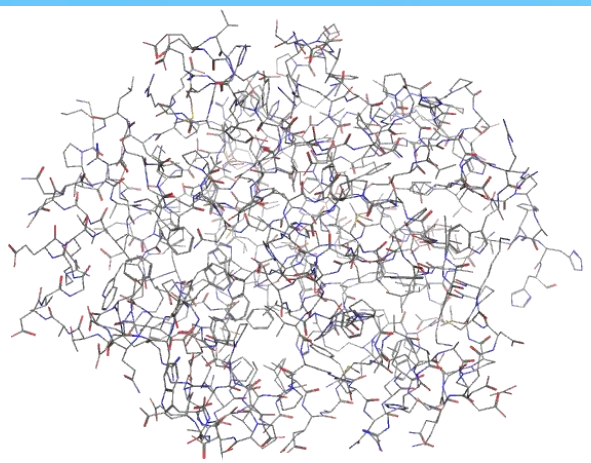
The students



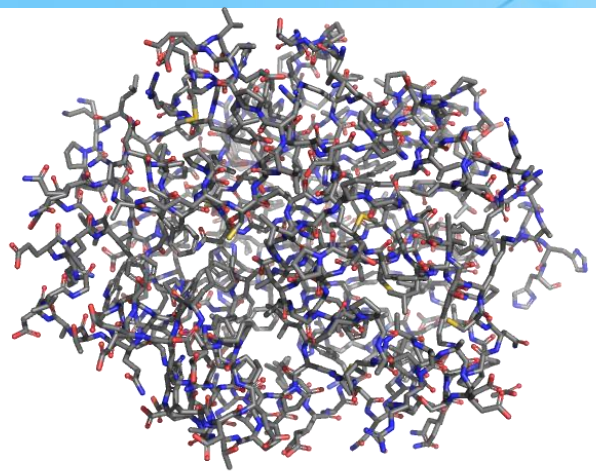
Structure visualization

- Some widespread-used programs
  - PyMOL – http://www.pymol.org/
  - Chimera – http://www.cgl.ucsf.edu/chimera/
  - VMD – http://www.ks.uiuc.edu/Research/vmd/
- Various representation
  - Bond-based
  - Backbone-based
  - Surface-based
- Seeing is believing, but …
  - Beware of misinterpretations and over-interpretations!

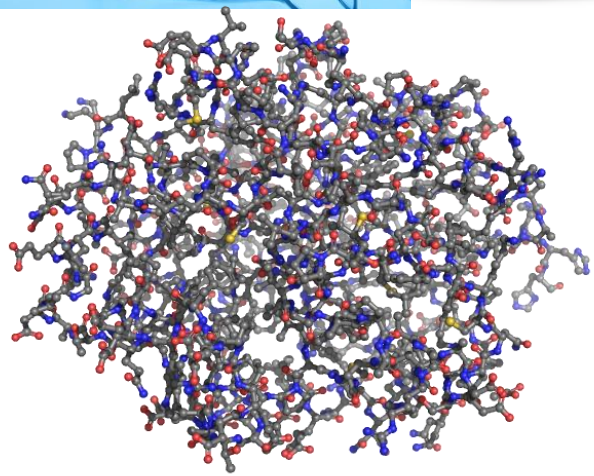
Structure visualization



Wire



Stick

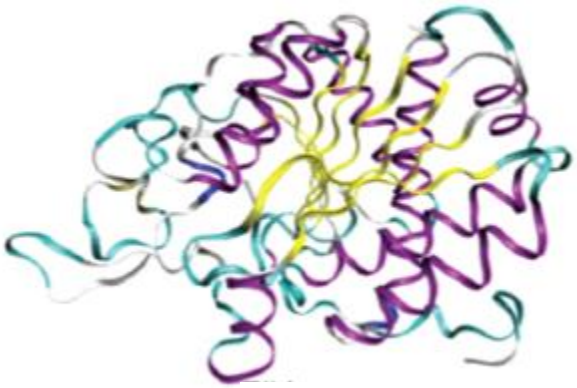


Ball and stick

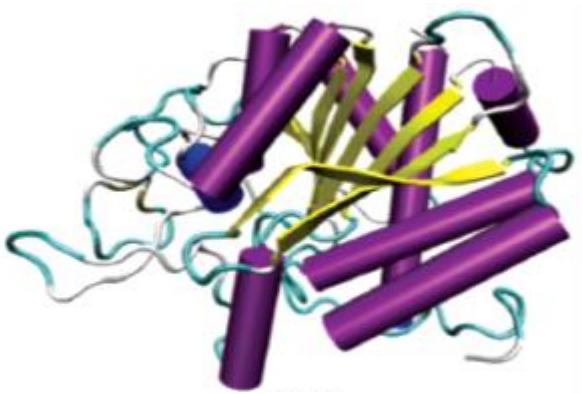
	hydrogen (H)	white
	carbon (C)	black
	nitrogen (N)	blue
	oxygen (O)	red
	fluorine (F), chlorine (Cl)	green

- Bonds-based representation
  - Fast, little resource-demanding
  - Suitable for detailed analysis
  - Incorrect impression about atom packing (empty space) and interatomic distances
- Hydrogen atoms are often omitted for simplicity

Structure visualization



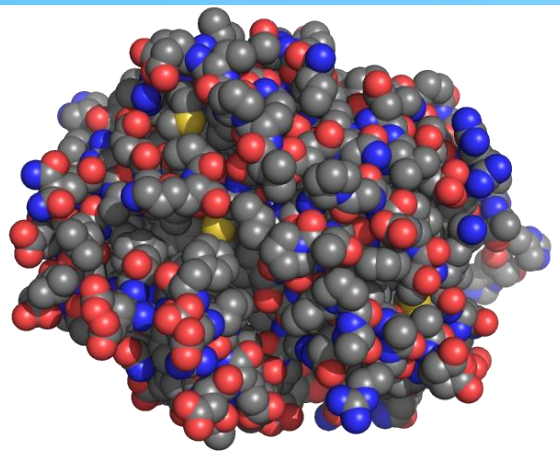
Ribbon



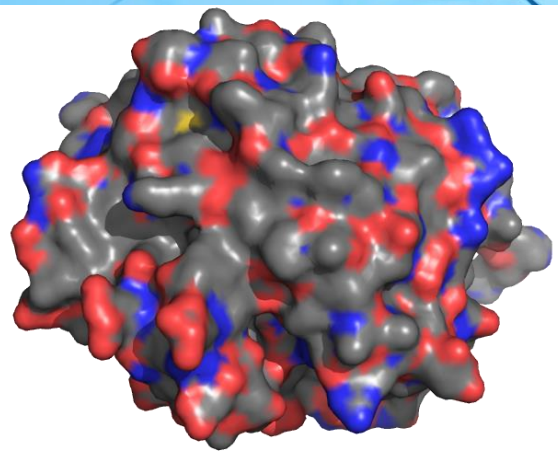
Cartoon

- Backbone-based representation
  - Moderately fast, not very resource-demanding
  - Suitable to investigate secondary structure and protein folds
  - Shows main landmarks; good for overall orientation in the structure

Structure visualization

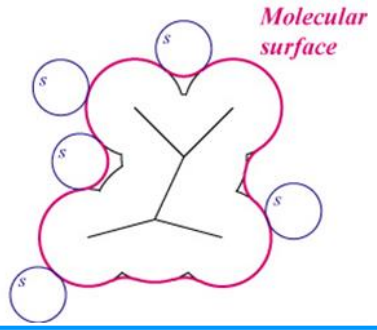


CPK/ spheres



Surface

- Surface-based representation
  - Very slow, very resource-demanding
  - Suitable to study shapes, volume, cavities and molecular contacts



Energetics of structures

- Energy
- Entropy
- Free energy
- Energy landscape

Energetics of structures

- Energy
  - Internal energy **U** (const. V); enthalpy **H** (constant P), …
  - Total energy often inaccessible -> differences in energy
  - Convention: negative energy is favorable, positive is unfavorable

  - Potential energy E_p – interactions of atoms in a system
  - Kinetic energy E_k – movement of atoms

$$U = E_p + E_k$$

$$H = U + P.V$$

Energetics of structures

- Entropy
  - Related to the thermal disorder or conformational availability (degrees of freedom)
  - Total entropy **S** > 0
  - Higher entropy is more favorable

Energetics of structures

- Free energy
  - Helmholtz **A** or **F** (const. V), Gibbs **G** (const. P)
  - Combination of internal energy or enthalpy and entropy **S**

$A = U - TS$; $G = H - TS \rightarrow \Delta G = \Delta H - T\Delta S$ (T = temperature)

  - Negative change of free energy ($\Delta G < 0$) is favorable

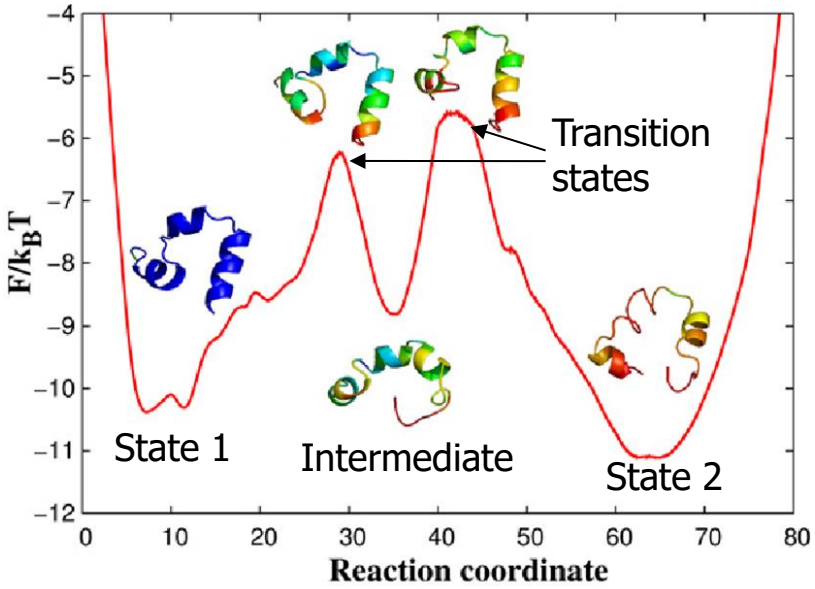
H↓ S↑

H↑ S↓

Energy landscape

- Relationship between structure and its potential energy
  - Structure dictates potential energy – how strong are the individual interactions
  - Potential energy reflects probability of finding the different structures – lower energy ➔ more frequently occurrence

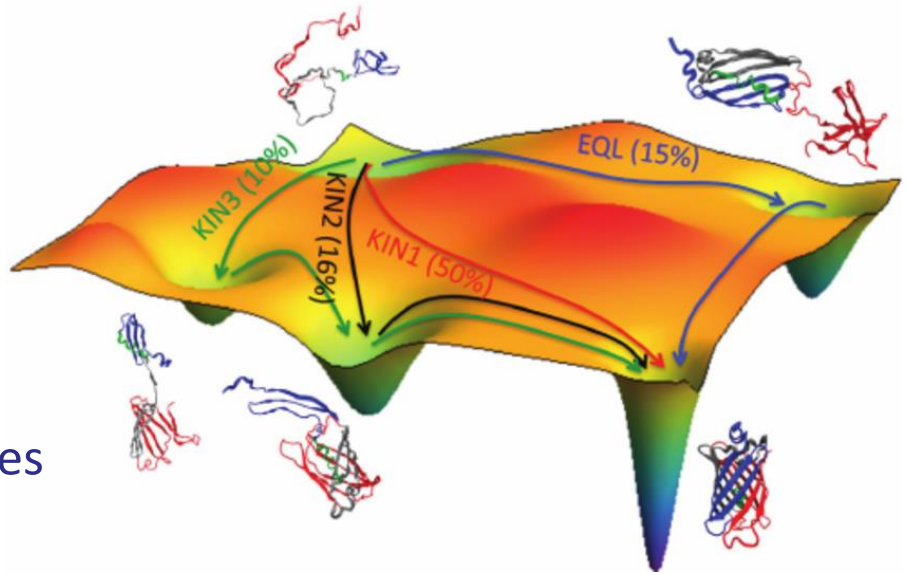
- Potential/free energy surface
  - Minima – stable structures
  - Saddle points – transient
  - Maxima – unstable structures
  - Energy barriers



Energy landscape

- Relationship between structure and its potential energy
  - Structure dictates potential energy – how strong are the individual interactions
  - Potential energy reflects probability of finding the different structures – lower energy ➔ more frequently occurrence

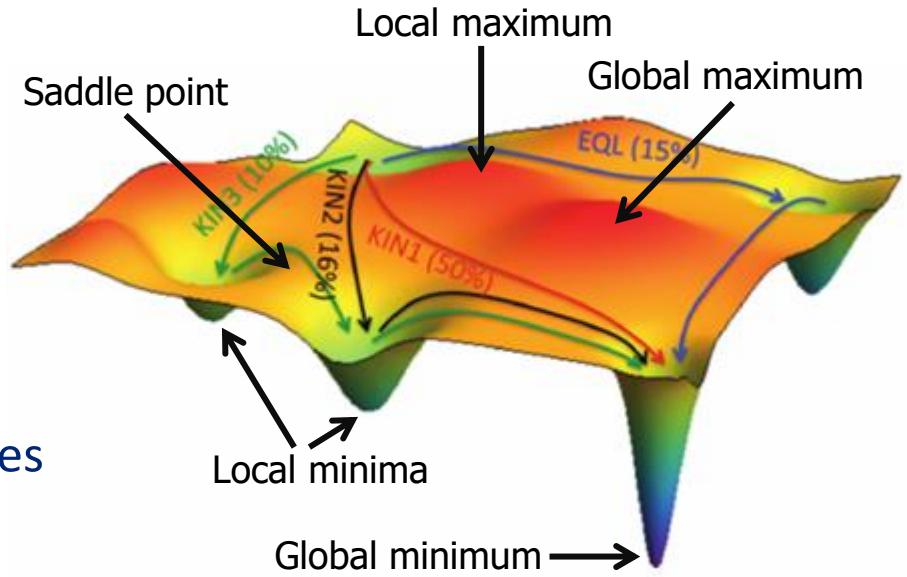
- Potential/free energy surface
  - Minima – stable structures
  - Saddle points – transient
  - Maxima – unstable structures
  - Multidimensional surface



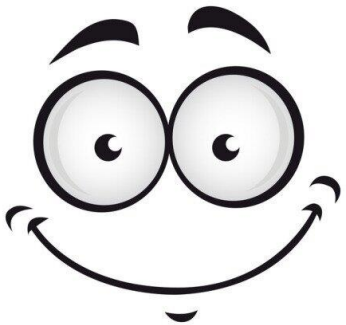
Energy landscape

- Relationship between structure and its potential energy
  - Structure dictates potential energy – how strong are the individual interactions
  - Potential energy reflects probability of finding the different structures – lower energy ➔ more frequently occurrence

- Potential/free energy surface
  - Minima – stable structures
  - Saddle points – transient
  - Maxima – unstable structures
  - Multidimensional surface



Molecular interactions



Molecular interactions

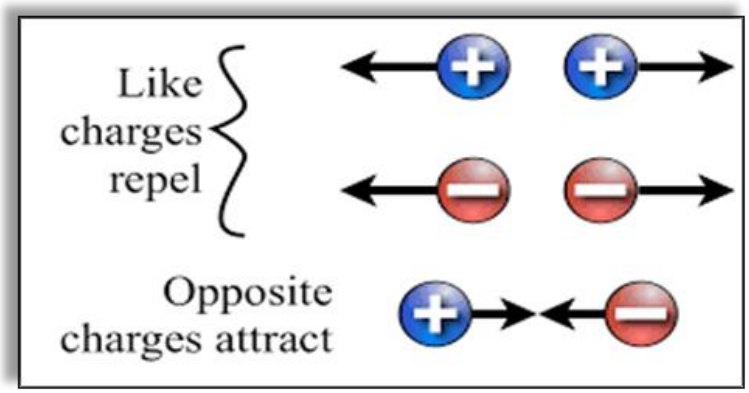
- Covalent interactions (chemical bonds)
  - Between two atoms sharing electrons
  - Very stable under standard condition
- Non-covalent interactions
  - Much weaker than covalent bonds
  - Electrostatic interactions
  - Polar interactions
  - Non-polar interactions

Electrostatic interactions

- Charge-charge or ionic interactions
  - Coulomb's law – between any two charges
  - Attractive (opposite signs) or repulsive (same sign)
  - Long-range interactions (up to 10 Å) – decrease with r^2

$$F = \frac{q_1 \cdot q_2}{4\pi \cdot \varepsilon \cdot r^2}$$

r = distance
ε = permittivity



Electrostatic interactions

- Charge-charge or ionic interactions

$$F = \frac{q_1 \cdot q_2}{4\pi \cdot \varepsilon \cdot r^2}$$

  - Environment-dependent
    - Permittivity

$\varepsilon = \varepsilon_0 \cdot \varepsilon_r$ $\quad\quad$ ε_0 = vacuum permittivity

    - Relative permittivity (ε_r) = dielectric constant

Non-polar

Material	ε_r
Vacuum	1.0
Air	1.0006
Teflon	2.1
Interior of proteins, membranes	2-20
Water (20 deg C)	80.1
Water (0 deg C)	88

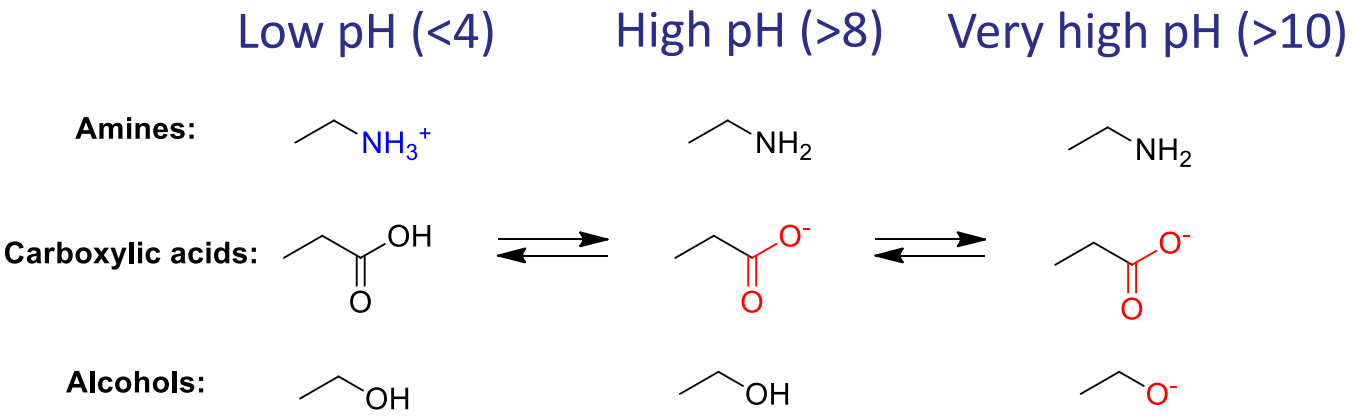
Highly polar

Stronger force

Weaker force

Electrostatic interactions

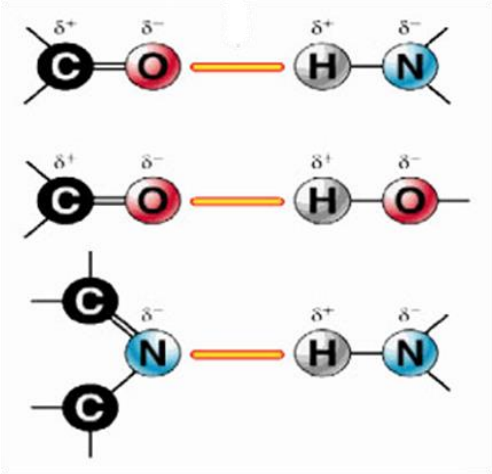
- Charge-charge or ionic interactions
  - Environment dependent
    - Salt concentration – presence of counter-ions (Na^+, K^+, Cl^-, etc.)
    - pH – may induce a change of charge



Polar interactions

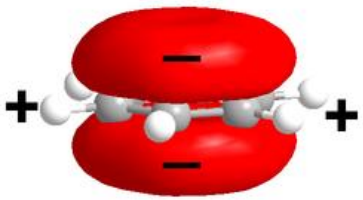
- Hydrogen bonds (H-bonds)
  - Only between highly electronegative atoms: fluorine, oxygen, nitrogen (F, O, N)
  - Donor and acceptor atoms sharing hydrogen
  - H-bond distance: 2.8 – 3.4 Å

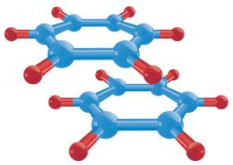
acceptor (δ^-) donor (δ^+)



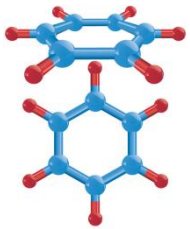
- Aromatic (π-π) interactions
  - Attractive interaction between aromatic rings
  - Distance between the center of mass of rings: ~ 5 Å

π orbitals

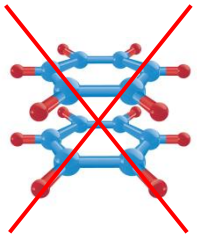




parallel displaced



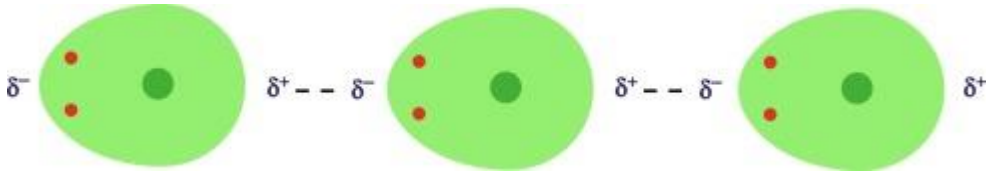
T-shaped

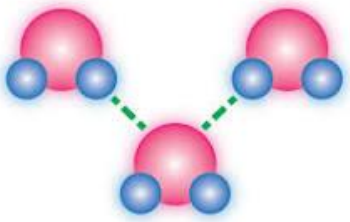


sandwich

Polar interactions

- Van der Waals (vdW) interactions
  - Between any two atoms
  - Permanent dipole-dipole (in polar molecules)





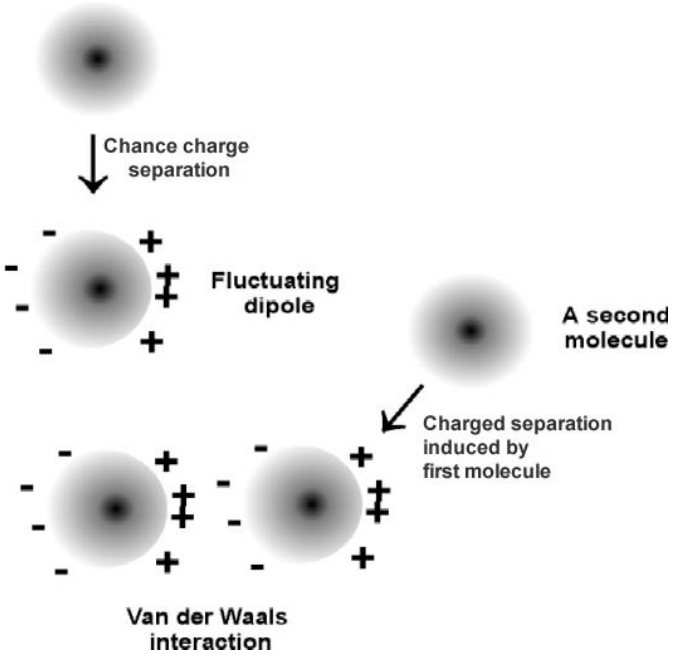
Non-polar interactions

- Van der Waals (vdW) interactions
  - Between any two atoms
  - London dispersion forces, or temporary dipole-induced dipole

    (in non-polar molecules)
  - Short-range interactions – up to 5 Å

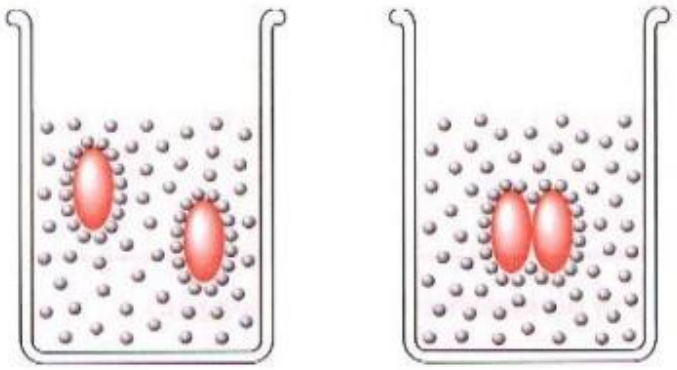
$$F_{\mathrm{VdW}}(r) = -\frac{AR_1R_2}{(R_1+R_2)6r^2}$$

R_1, R_2 – van der Waals radii
r - distance



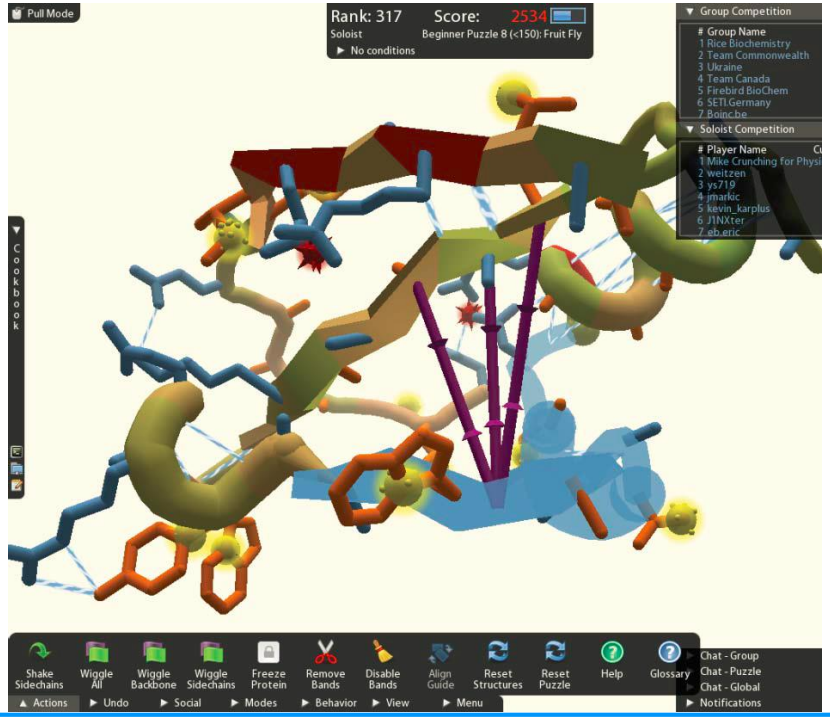
Non-polar interactions

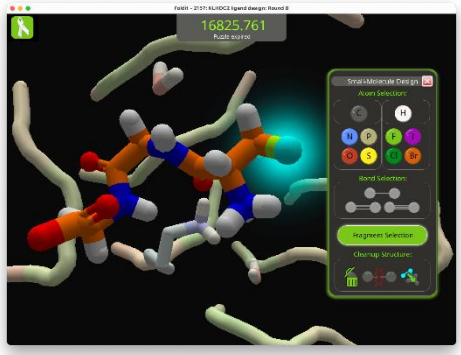
- Hydrophobic interactions
  - Entropic origin – water molecules ordered around hydrophobic moiety -> unfavorable
  - Hydrophobic packing -> favorable release of some ordered water molecules

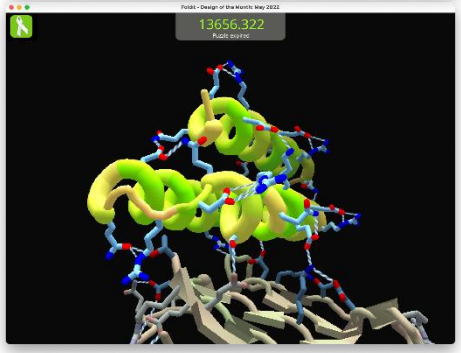


Protein folding game

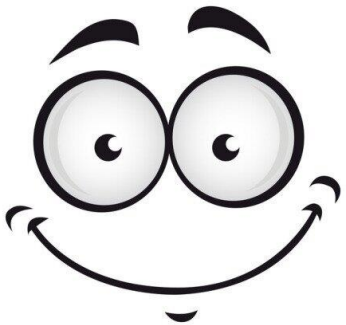
- FOLD.IT – https://fold.it/
  - Crowdsourcing computer game
  - Prediction of protein structures
  - You can contribute to help scientific research





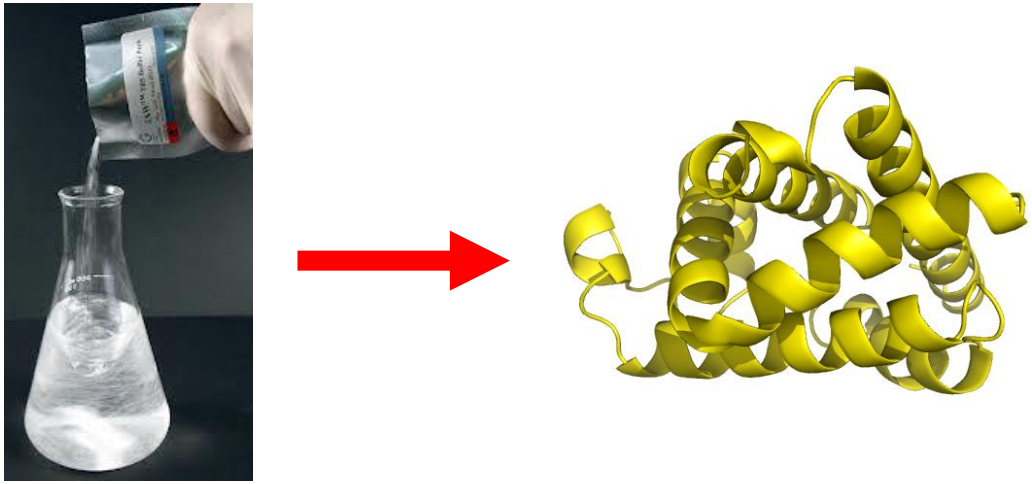


Structure determination



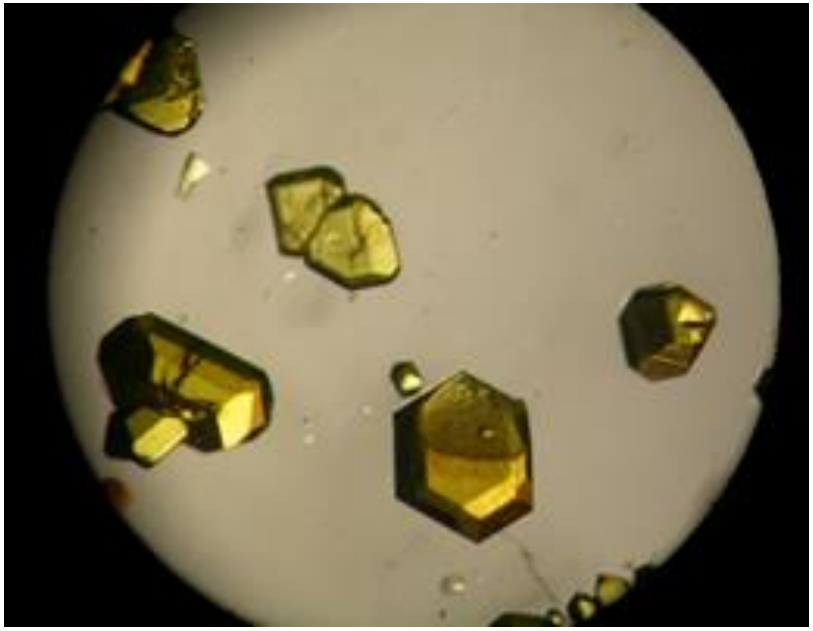
Structure determination

- Established methods
  - X-ray crystallography
  - NMR spectroscopy
  - Electron microscopy
  - Bioinformatics predictions – theoretical



X-ray crystallography

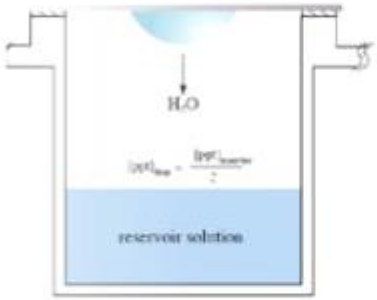
- Crystallization procedures
  - Slow (days-weeks)
  - High risk of failure



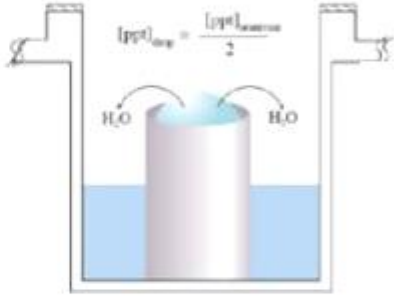
Some Crystallization Methods:

Vapor diffusion

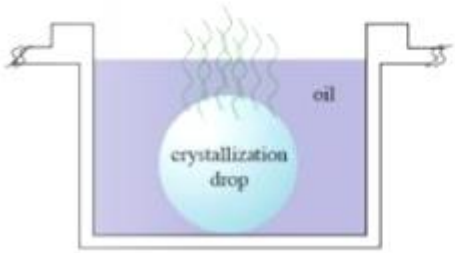
Hanging-drop



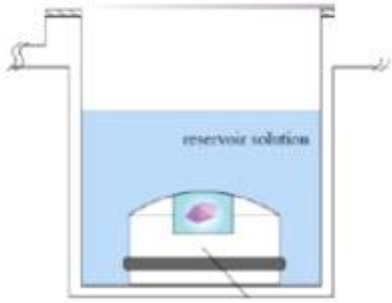
Sitting-drop



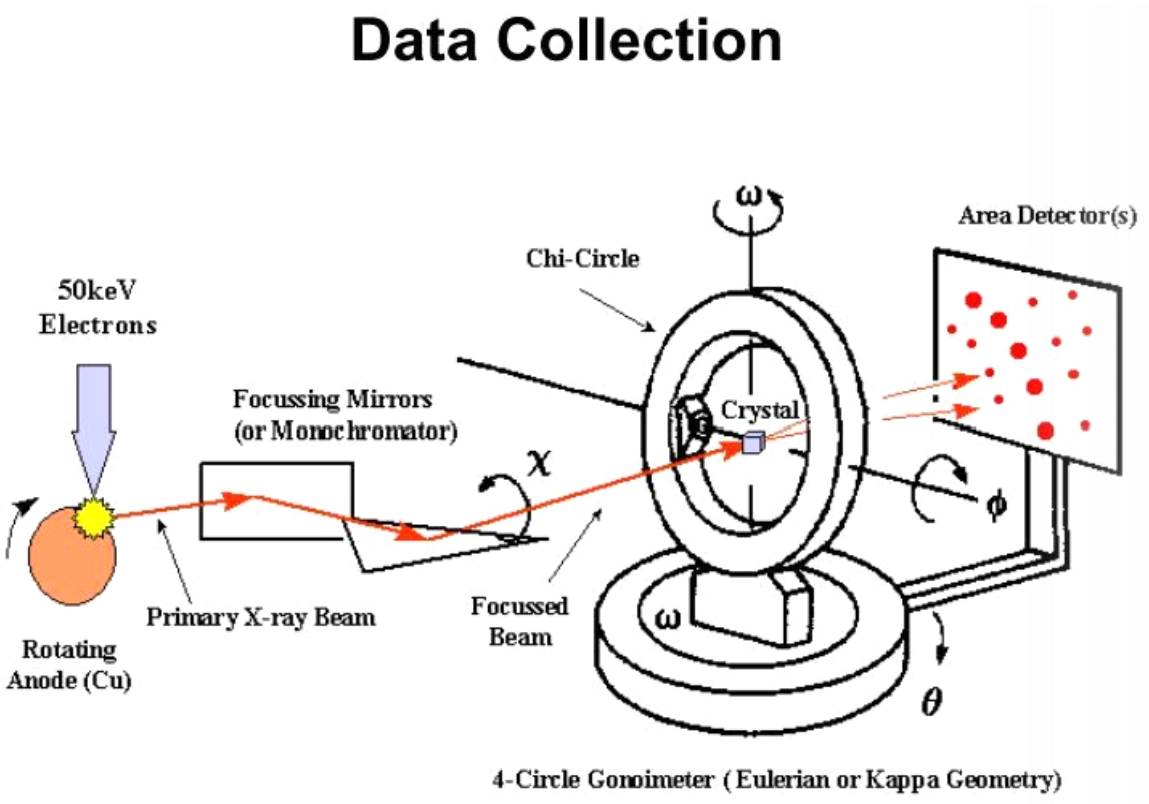
Batch:
micro batch under oil



Dialysis



X-ray crystallography



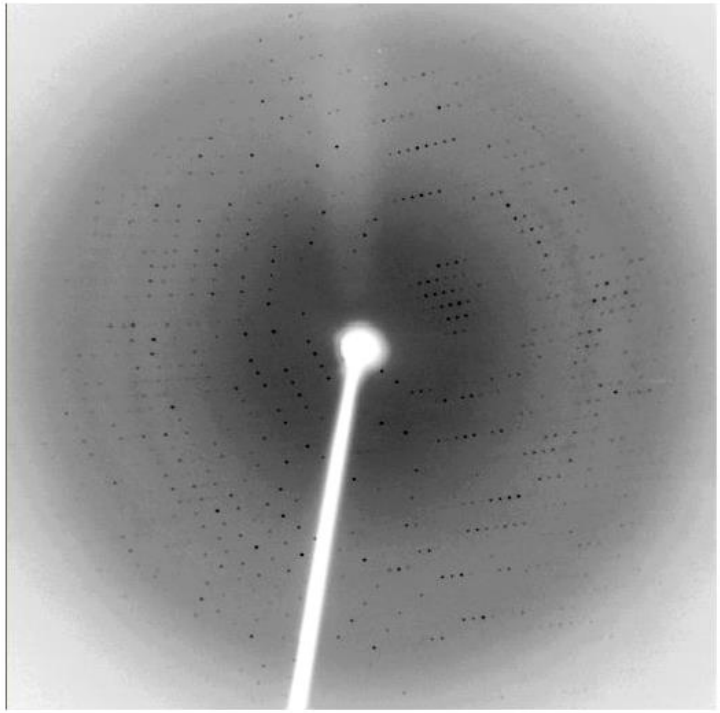
APS Chicago

European Synchrotron Radiation Facility, Grenoble

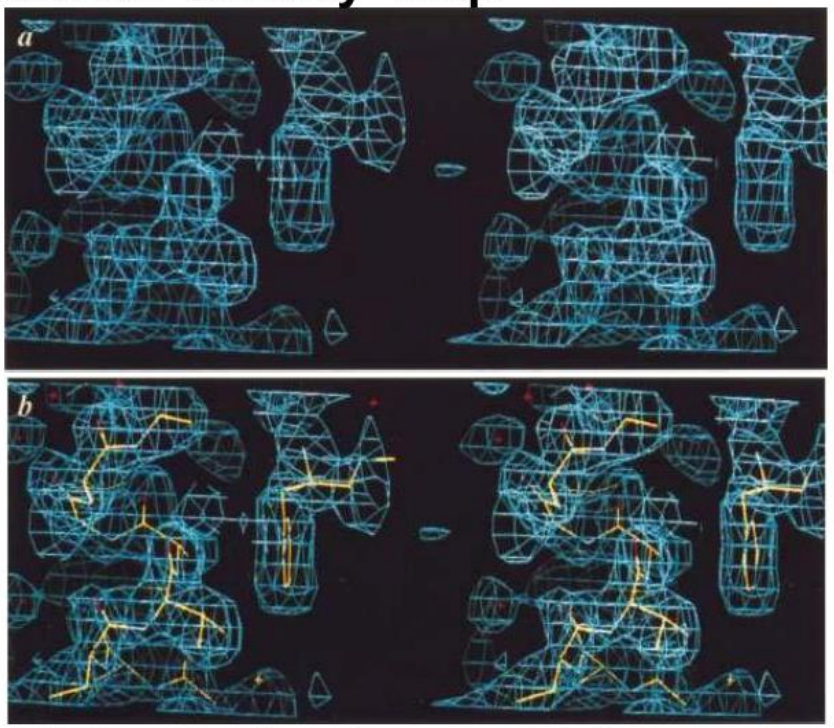
X-ray sources: X-ray tubes, rotating anodes and synchrotrons.
Synchrotrons produce the brightest X-rays (~70 worldwide)

X-ray crystallography

Image of diffraction



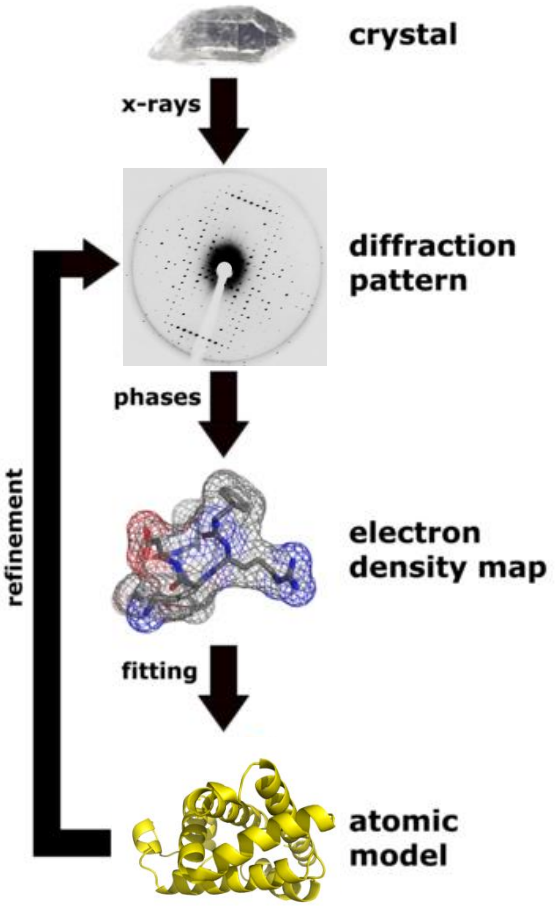
Electron density map



Building a structure model

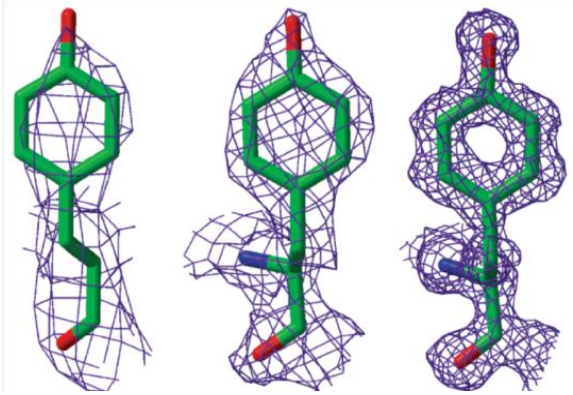
X-ray crystallography

- Crystallization
  - Hanging drop, sitting drop, microbatch
- Data collection
  - Diffractometers, synchrotrons
- Analysis of diffraction data
  - Solving phase problem
    - Molecular replacement
    - Isomorphous replacement
    - Anomalous scattering
- Iterative model building



Parameters of an X-ray structure

- Resolution
  - Measure of the level of detail present in the diffraction pattern

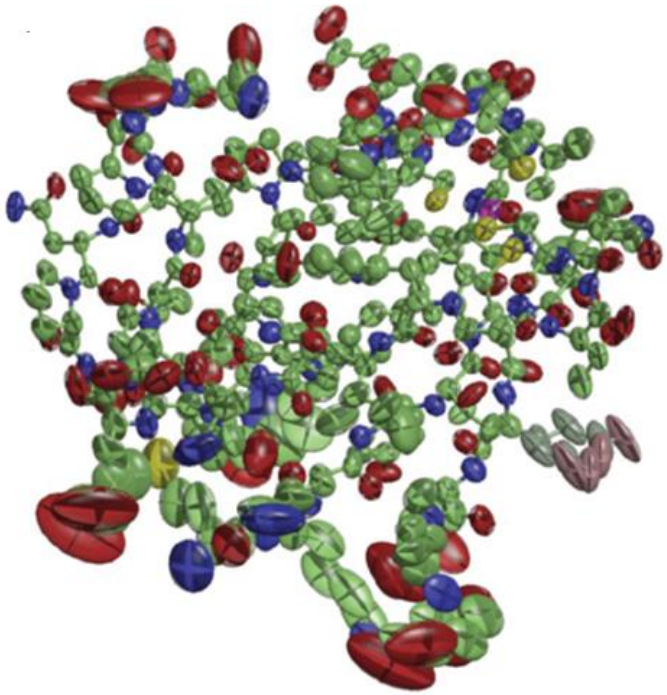


3 Å (bad) **2 Å** (acceptable) **1 Å** (exceptional)

- R-factor (residual factor; R-value)
  - Measure of a model quality – i.e. the agreement between the crystallographic model and the diffraction data
  - Varies from 0 (ideal) to 0.63 (random structure), typically about 0.2

Parameters of an X-ray structure

- B-factors (thermal factors)
  - Measure of how much an atom oscillates or vibrates around the position specified in the model
  - Considered a measure of flexibility



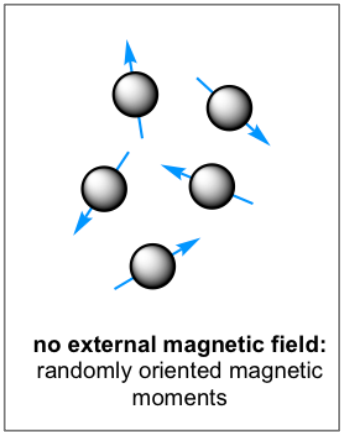
X-ray crystallography

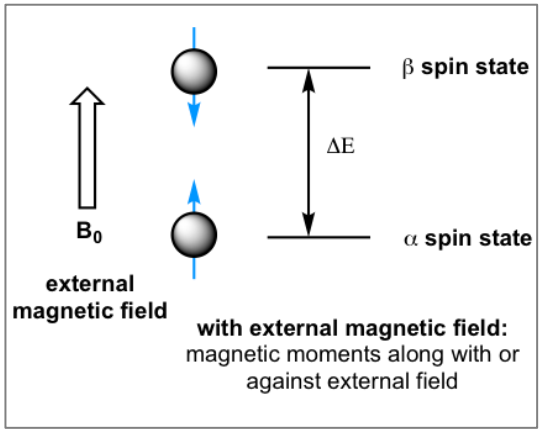
- Advantages
  - No limitations in size
  - Possibility to obtain an atomic resolution
- Disadvantages
  - Requirement of a crystal
  - Structure in a crystalline state (non-native)
  - Static picture of macromolecule
  - Position of hydrogen atoms (usually) are not detected

NMR spectroscopy

- Nuclear magnetic resonance (NMR)
  - Detects energy transitions in the magnetic moments of nuclei with non-zero nuclear spins
  - Common isotopes:

^{1}H, ^{13}C, ^{15}N, ^{31}P, ^{35}Cl





900 MHz NMR spectrometer

NMR spectroscopy

Sample preparation
0.5 ml/~1 mmol L^{-1} concentration/^{13}C,^{15}N-labelling

NMR spectroscopy
homonuclear/heteronuclear/2D/3D/4D

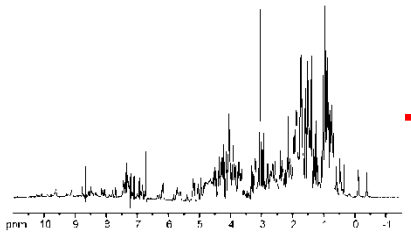
Sequence-specific resonance assignments

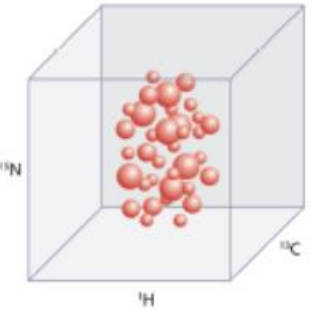
Collection of conformational constraints

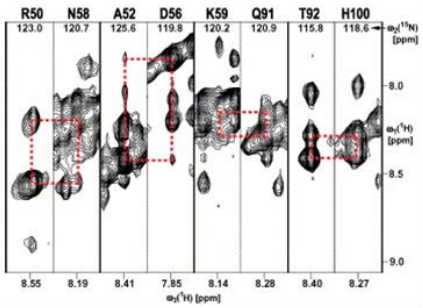
Regular secondary structures

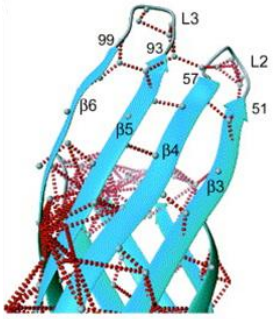
3D structure
distance geometry

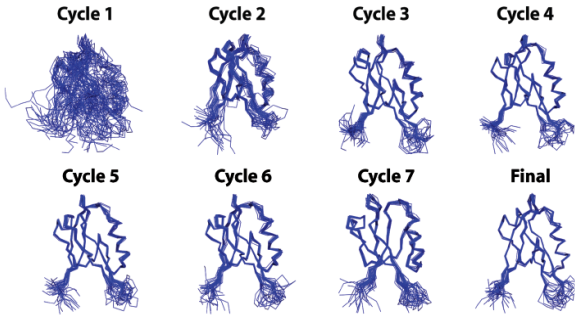
Structure refinement
energy minimization











Parameters of an NMR structure

- RMSD
  - **Root-mean-squared deviation** of atomic positions across the ensemble of solutions
  - Reveals the *mean* differences between individual conformations
  - Important parameter to compare different structures of the same molecule

RMSD = 3.59 Å

RMSD = 1.06 Å

RMSD = 0.42 Å

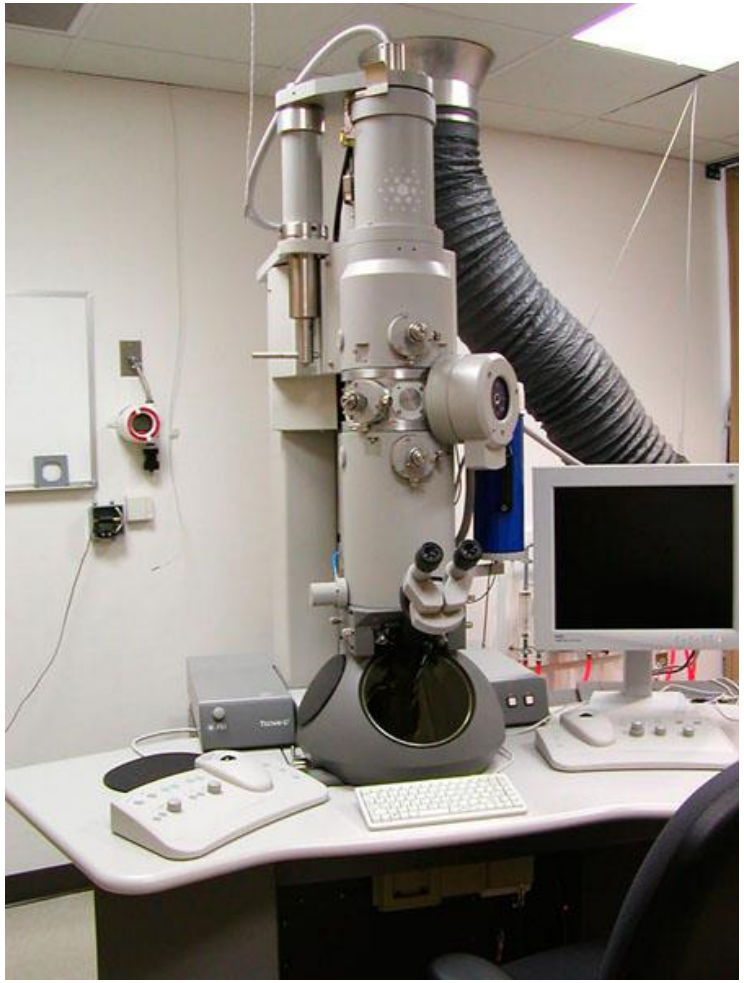
$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

δ = atom displacement
N = total No. atoms

NMR spectroscopy

- Advantages
  - Structure in solution state (native)
  - Possibility to investigate dynamics of macromolecules
  - Position of hydrogen atoms detected
- Disadvantages
  - Size limited to approximately 40 kDa (~ 400 amino acid proteins)
  - Requirement of isotopically labeled sample

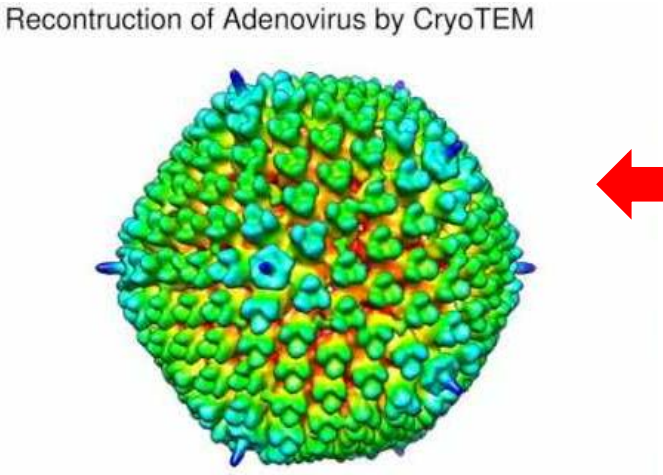
Electron microscopy

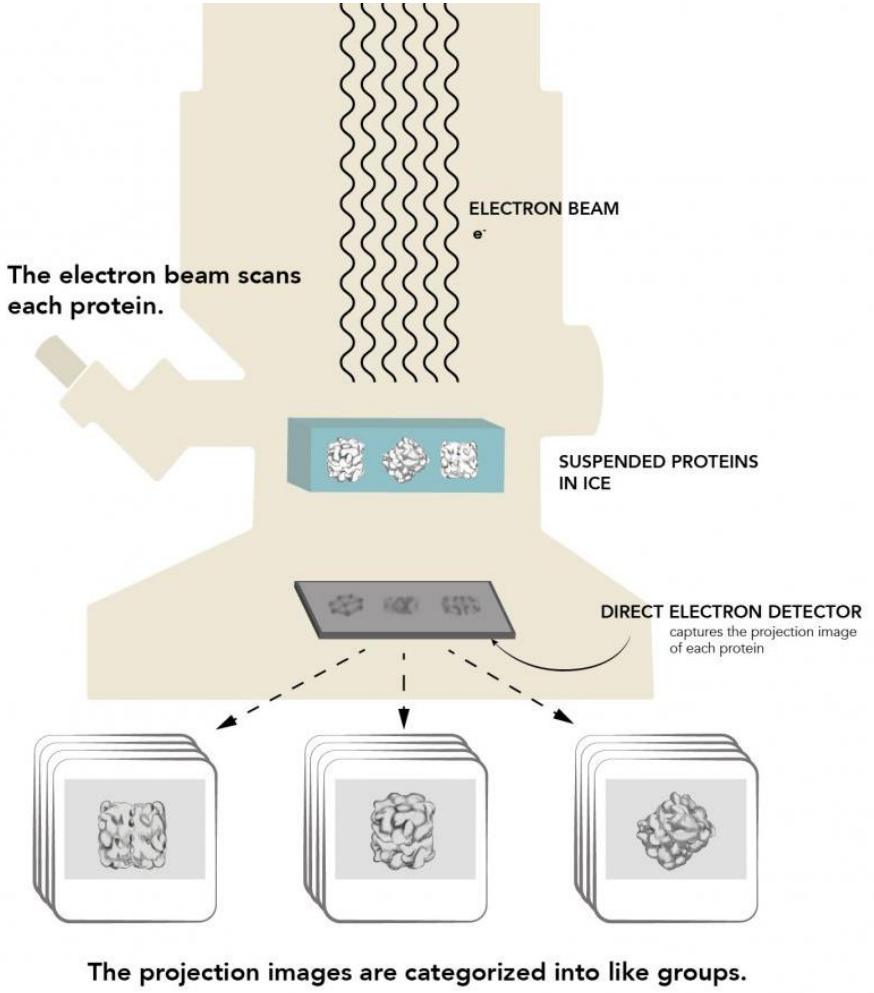


FEI Tecnai T12 Cryotransmission Electron Microscope

Electron microscopy

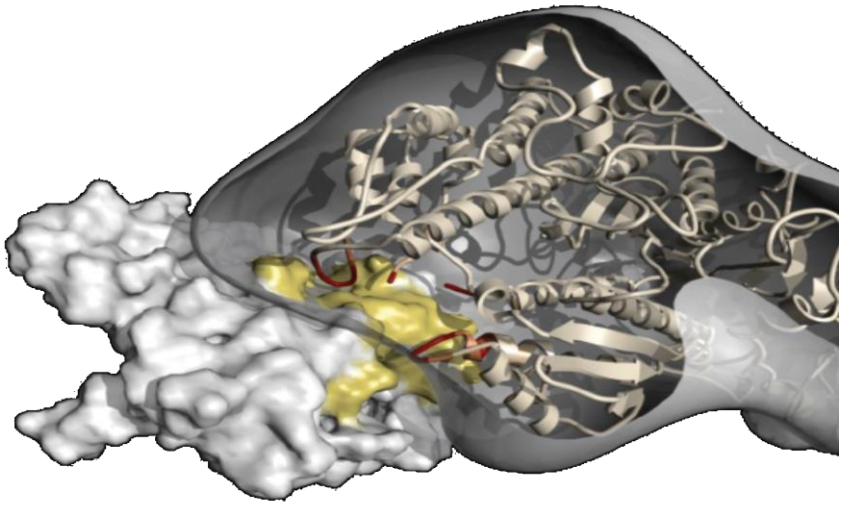
- Wavelength of an electron is much shorter than the wavelength of light
- → so it can reveal much smaller thin
- Samples are flash-frozen in their natural environments (cryo-EM)
- Can generate 3D images of large molecules at nearly atomic resolution





Electron microscopy

- Advantages
  - Applicable to extremely large systems
  - Complements other methods e. g. X-ray, NMR



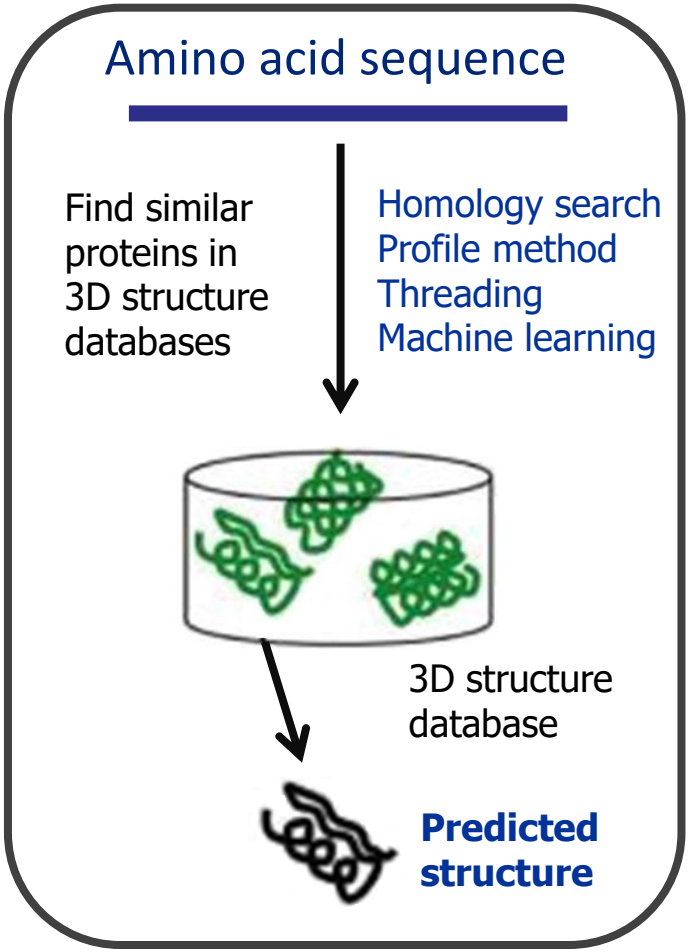
- Disadvantages
  - Lower resolution (2-3 Å at best)

Bioinformatics predictions

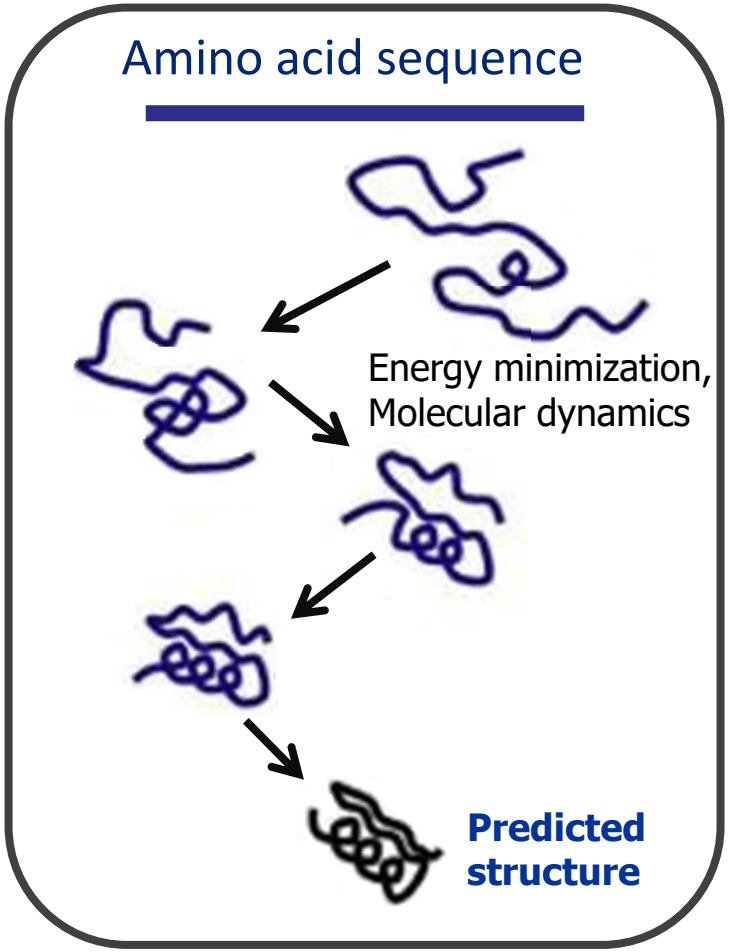
- Homology modeling
- Machine learning
- *Ab initio* prediction

Bioinformatics predictions

Comparative modelling

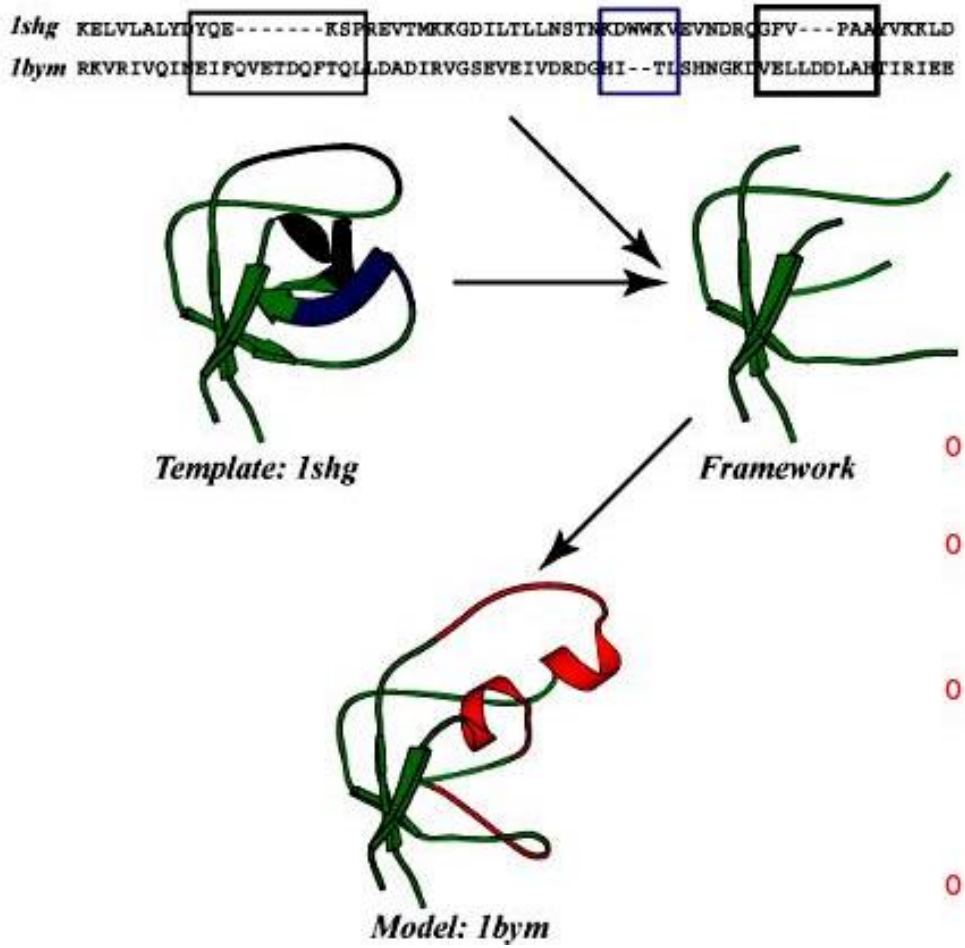


Ab initio predictions



Bioinformatics predictions

- Homology modeling

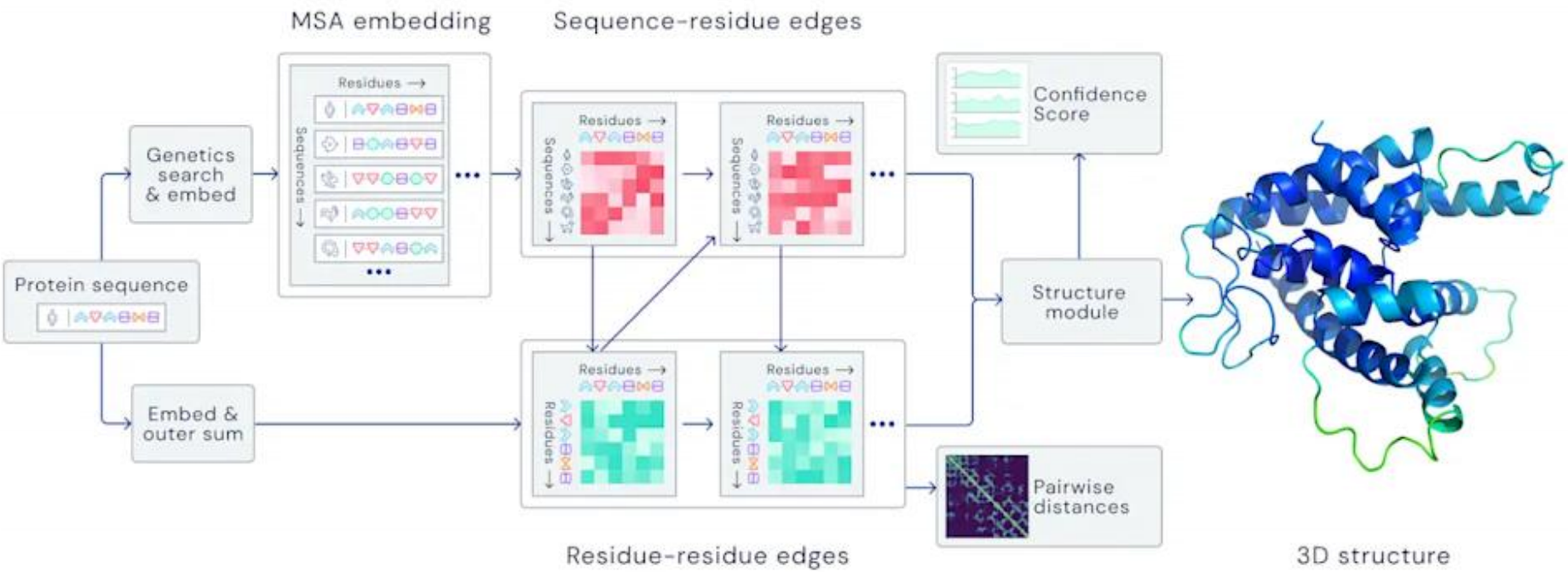


Comparison of sequences in databases:
Multiple sequence alignment (MSA)

- *Find template*
- *Align target sequence with template*
- *Generate model:*
  - *add loops*
  - *add sidechains*
- *Refine model*

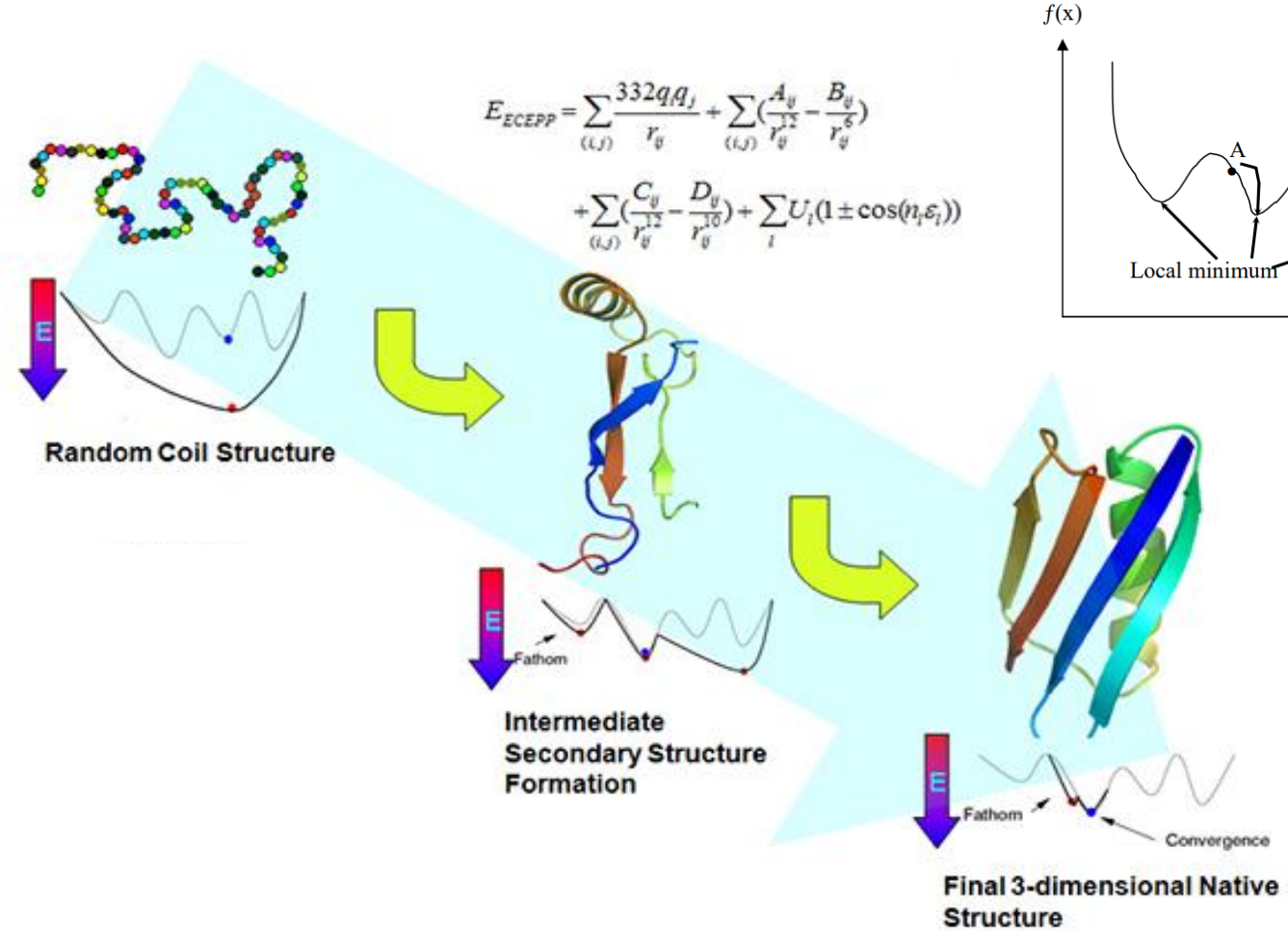
Bioinformatics predictions

- Machine learning
  - Training on sequence and 3D databases
  - Ex.: AlphaFold 2



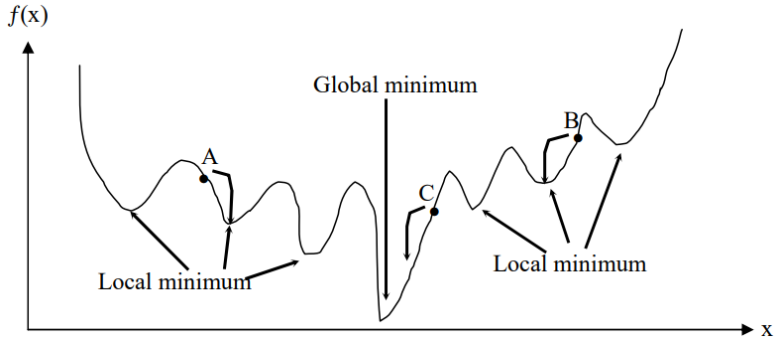
Bioinformatics predictions

- *Ab initio* prediction



$$E_{ECEPP} = \sum_{(i,j)} \frac{332 q_i q_j}{r_{ij}} + \sum_{(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right)$$

$$+ \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + \sum_{i} U_i (1 \pm \cos(n_i \varepsilon_i))$$



Bioinformatics predictions

- Advantages
  - Very fast (except *ab initio*)
  - Low cost
- Disadvantages
  - *Ab initio* is very demanding
  - Theoretical model – experimental validation is needed

References

- Petsko, G. A. & Ringe, D. (2004). **Protein Structure and Function**, New Science Press, London.
- Gu, J. & Bourne, P. E. (2009). **Structural Bioinformatics, 2nd Edition**, Wiley-Blackwell, Hoboken.
- Liljas, A. *et al*. (2009). **Textbook Of Structural Biology**, World Scientific Publishing Company, Singapore.
- Schwede, T. & Peitsch, M. C. (2008). **Computational Structural Biology: Methods and Applications**, World Scientific Publishing Company, Singapore.
- O'Donoghue, S. *et al*. (2010) Visualization of macromolecular structures. *Nature Methods* **7**: S42–S55.
- Zhou, H-X. & Pang, X. (2018) Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical Reviews*. **118**: 1691–1741