Introduction to the structure of macromolecules

Course information

❑ 10 lectures ≈ 20 h

    1. Introduction to the structure of macromolecules

    2. Structure of biomolecules

    3. Bioinformatics databases and structure prediction

    4. Models of structures

    5. Stability and dynamics of macromolecules

    6. Analysis of protein structures

    7. Protein-ligand complexes

    8. Macromolecular complexes and interactions

    9. Engineering of protein structures

    10. Applications of structural biology and bioinformatics

Course information

- 3 lecturers

  - Sérgio Marques, PhD

    → Lectures 1-2, 5-10

  - Joan Planas, PhD

    → Lectures 3, 4

  - Mgr. David Bednář, PhD

    → Lecture 5

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

  - 12 Jan. 2021; 13:00 - B11/305

 - 25 Jan 2021; 13:00 - B11/305

  - 10 Feb. 2021; 09:00 - B11/305

- ❑ Slides with essential information have the sign:

AI in Biology, Chemistry, and Bioengineering - **Bi9680En**

EN

- Semester: autumn

- Lectures: 2 hours/week

- Lecturer: Dr. Stanislav Mazurenko

- Outline:

  - Modern bio-challenges: drug design, DNA interpretation, protein engineering

  - Types of AI algorithms and workflow for designing predictors

  - Clustering algorithms, random forests, artificial neural networks

  - Features, databases, and predictors used in applications
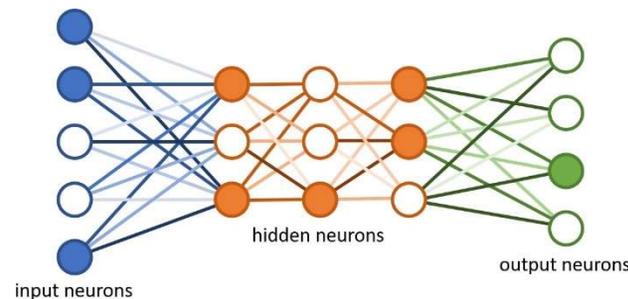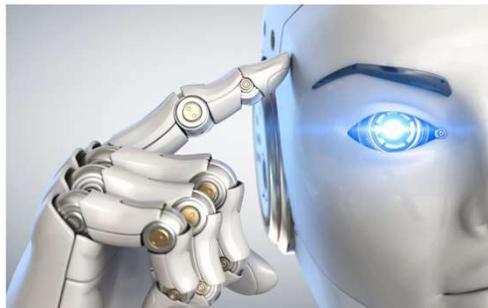
Bioinformatics - **Bi5000**

CZ

- Semester: autumn

- Lectures: 2 hours/week

- Lecturer: Prof. Jiří Damborský and Dr. Roman Pantůček

- Outline:

    - Searching in bioinformatics databases

    - Analysis of nucleotide and protein sequences

    - Identification of genes

    - Analysis and prediction of protein structure
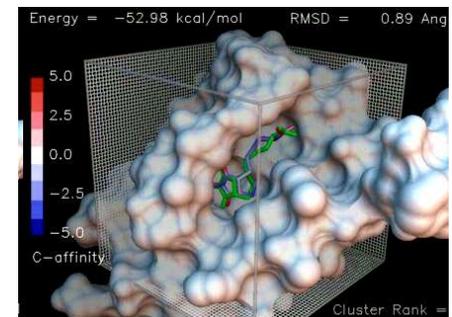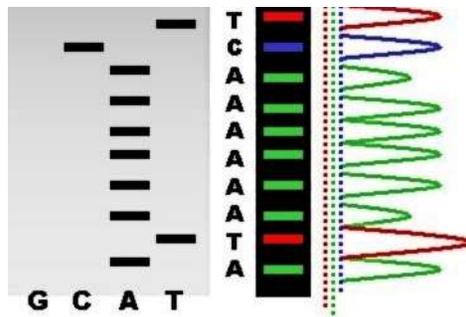
# Molecular biotechnology - **Bi7430**

CZ

- Semester: autumn

- Lectures: 2 hours/week; exercises: 2 hours/week

- Lecturer: Doc. Prokop, Dr. Dvořák, Dr. Bidmanová

- Outline:

    - Protein and metabolic engineering

    - Molecular diagnostics and modern vaccines

    - Cell and gene therapy and regenerative medicine

   - Molecular biotechnology in industry and agriculture
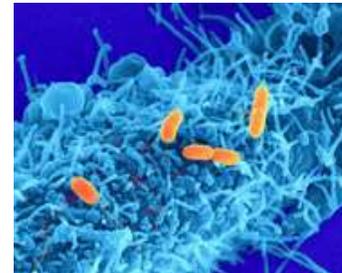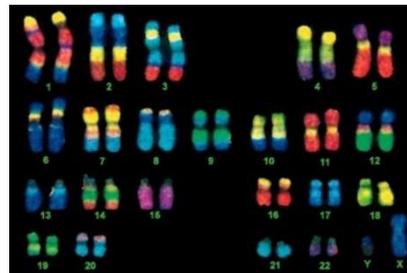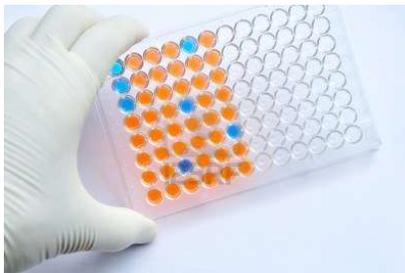
Synthetic biology - **S2015**

- Semester: autumn

- Lectures: 2 hours/week

- Lecturer: Dr. Martin Marek and Dr. Karel 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

CZ
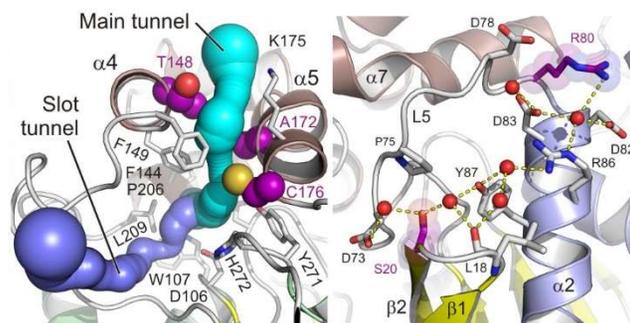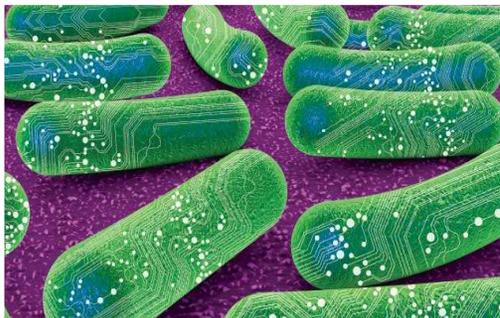
Protein engineering - **Bi7410**

- Semester: spring

- Lectures: 1 hour/week

- Lecturer: Dr. Radka Chaloupková

- Outline:

  - Structure-function relationships in proteins

  - Expression and purification of recombinant proteins

  - Methods of structure and function analysis

  - Rational design, semi-rational design, directed evolution

 - Application of protein engineering

CZ
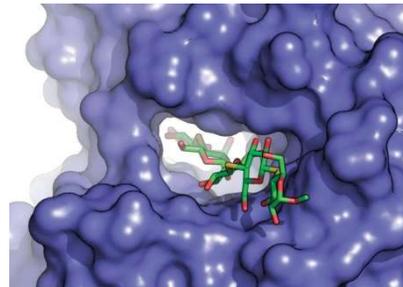
Outline

- ❑ Motivation

- ❑ Structural biology

- ❑ Bioinformatics

- ❑ Visualization of structure

- ❑ Energetics of structures

- ❑ Molecular interactions

- ❑ Determination of structure

Motivation

❑ **Sequence-to-structure-to-function paradigm**
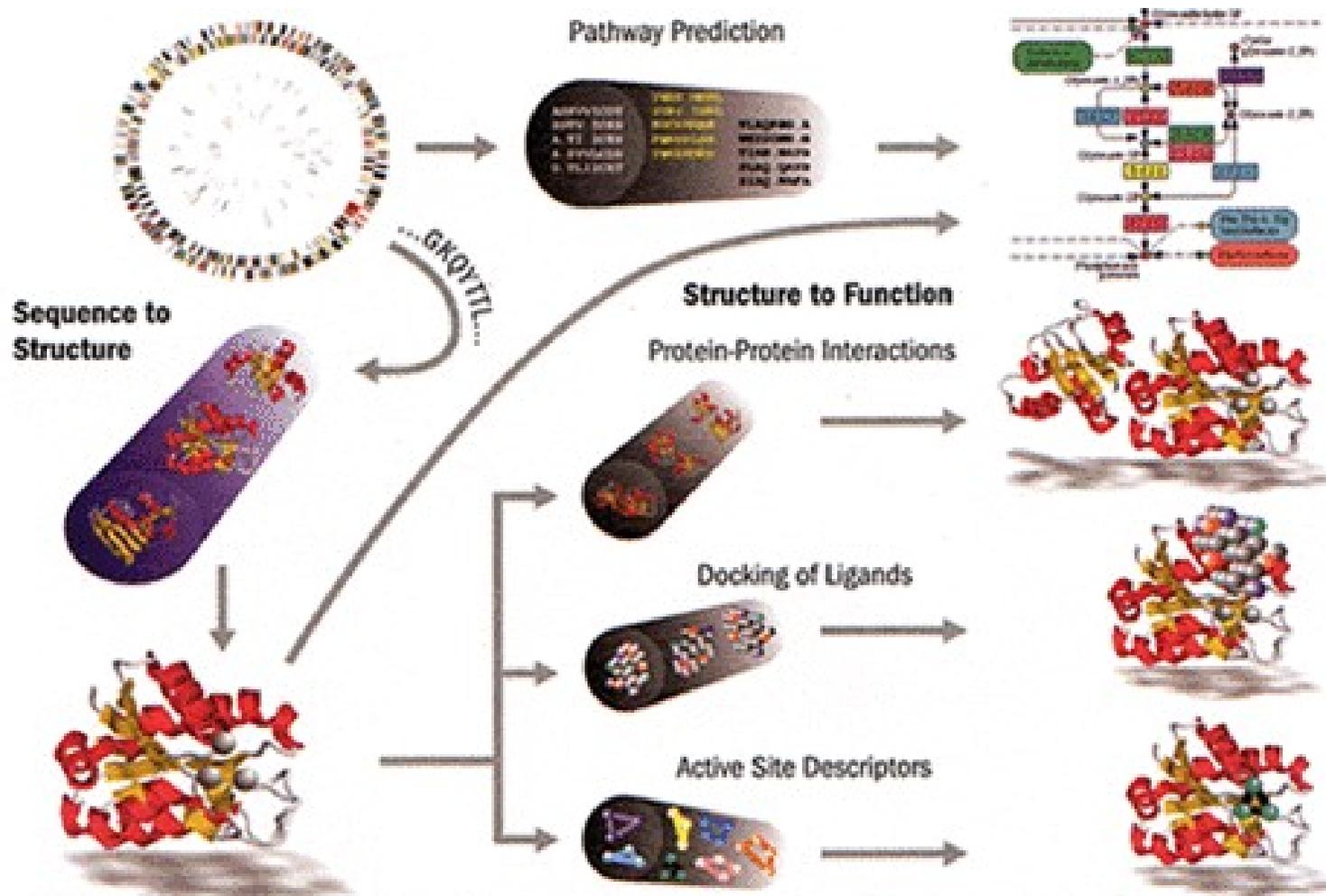
Motivation

❑ 3D structure ⇔ biological function
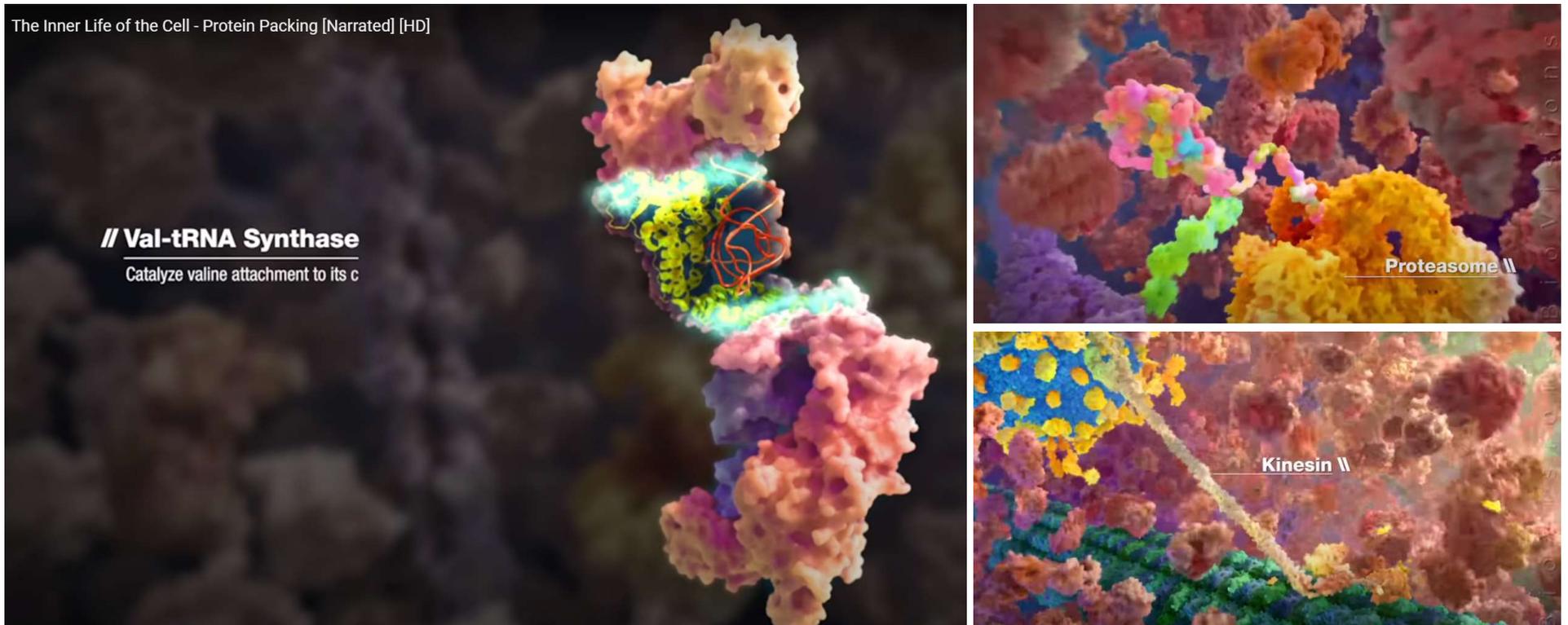
XVIVO & Harvard University – The inner life of the cell

https://www.youtube.com/watch?v=VdmbpAo9JR4

Structural biology

❑ Focused on the three dimensional structures of biomolecules and their mutual interactions to understand their functions in the cell.

❑ Tries to make biological objects visible and understood

  ▪ "Seeing is believing."

 ▪ "In order to understand, we need to see."
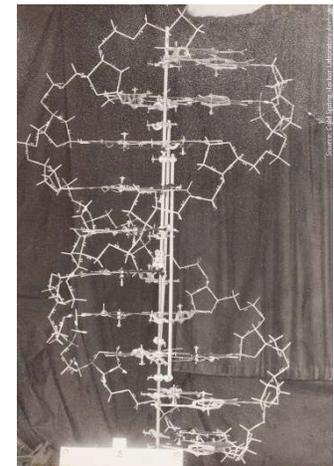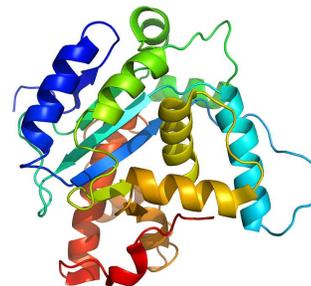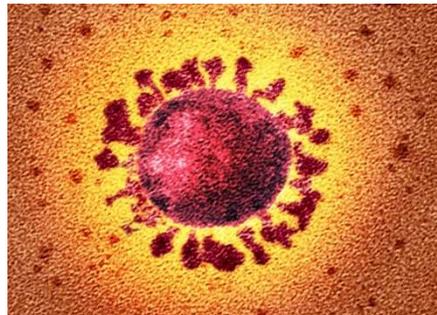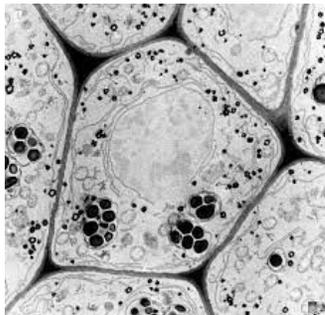
Structural biology

❑ Important milestones

- 1869 – DNA discovery

- 1953 – DNA structure

- 1959 – Myoglobin structure

- 1968 – Hemoglobin structure

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

James Watson, 1964

Structural biology
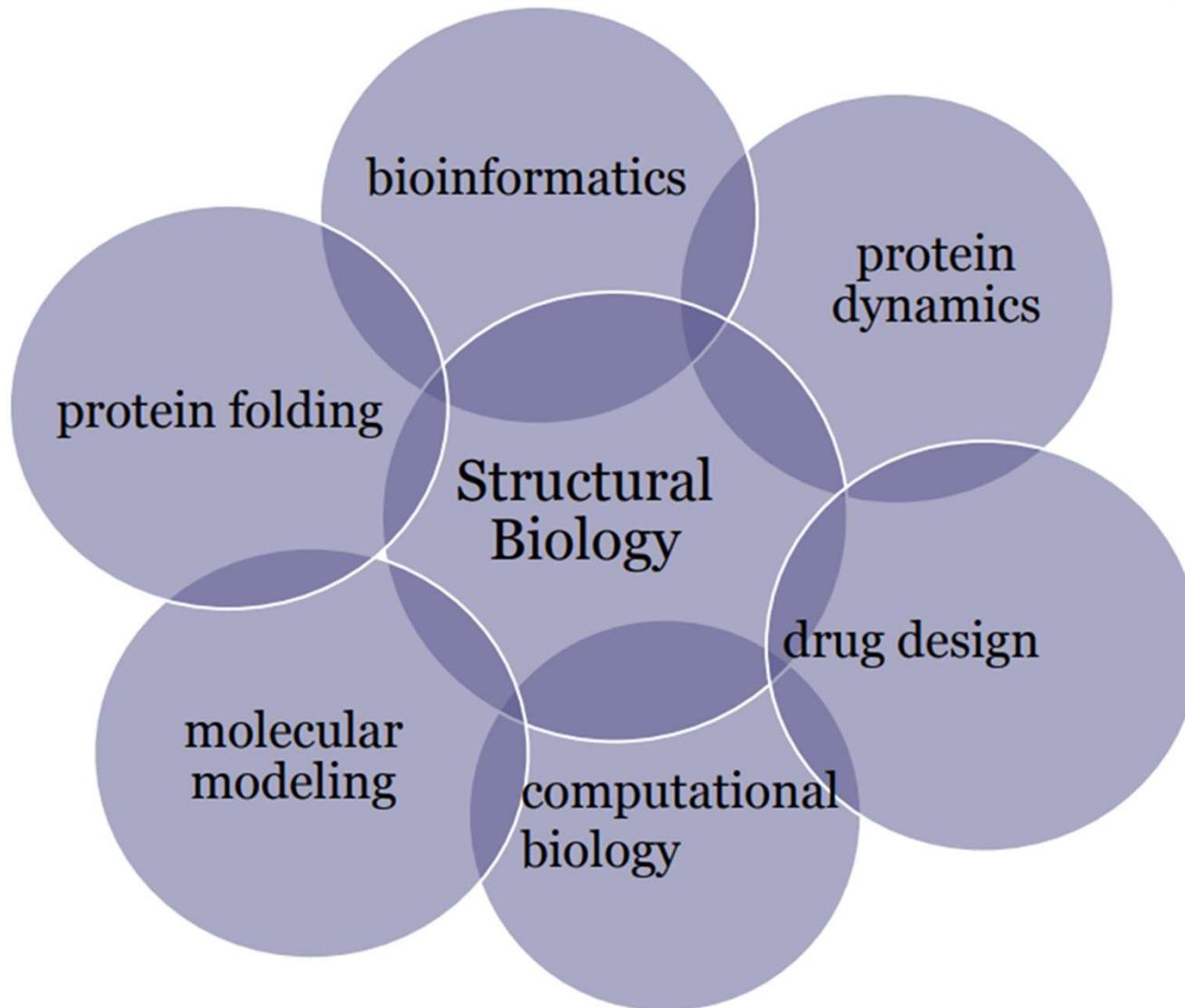
Structural biology

	antibiotic drug	substrate-enzyme	cellular systems	organism phenotype
scale				
objectives	ligand structure similarity identifies promiscuous activity on antibiotics	prediction of ligand off-target binding to protein active or allosteric binding sites	metabolic pathway perturbations as identified by constraint-based modeling	Development of resistance; expression profiles of over expressed genes in presense of drug
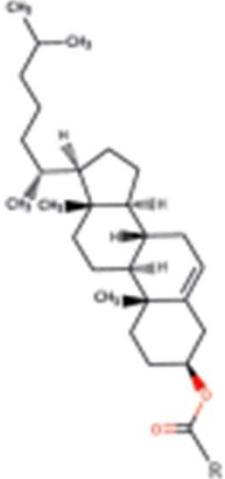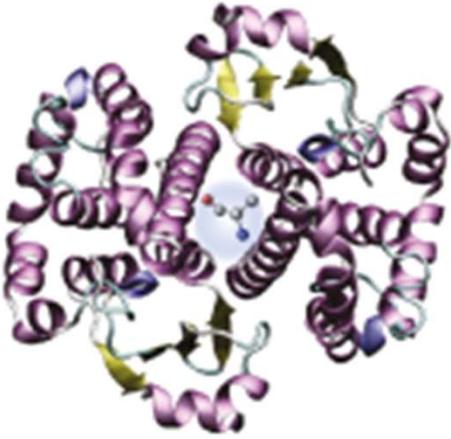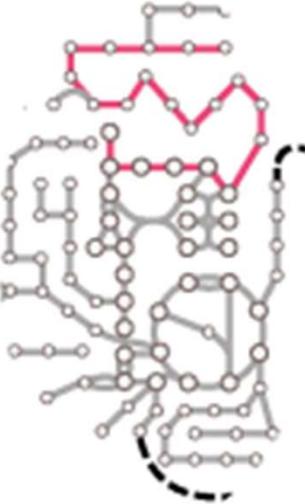
Bioinformatics

❑ "**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."

In Wikipedia

❑ Sequence analysis, genomics, proteomics, systems biology,

structural bioinformatics

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/

  - Caver Analyst – https://www.caver.cz

- Various representation

  - bond-based

  - backbone-based

  - surface-based

- Seeing is believing but …

  - beware of misinterpretations and over-interpretations!

Structure visualization
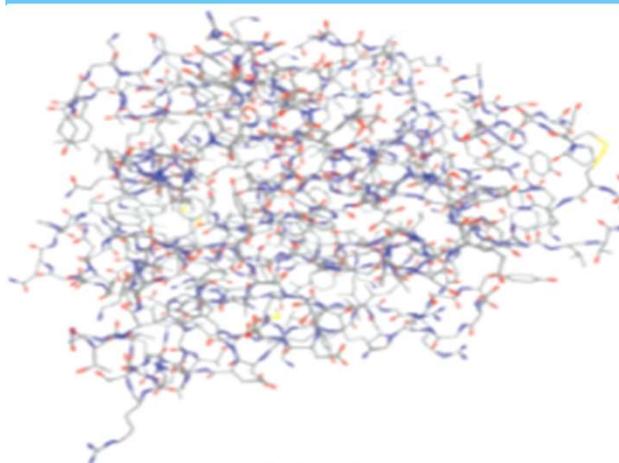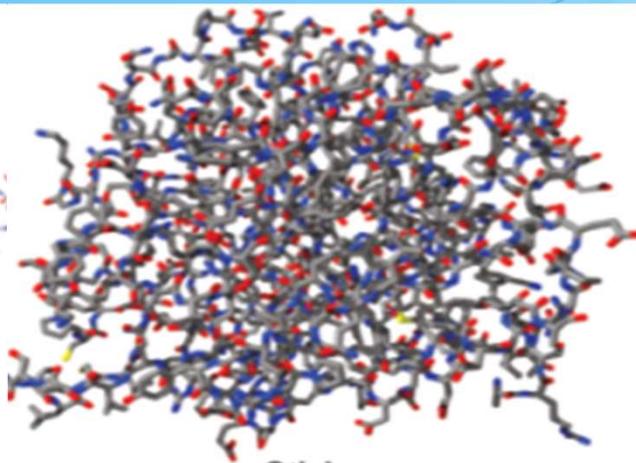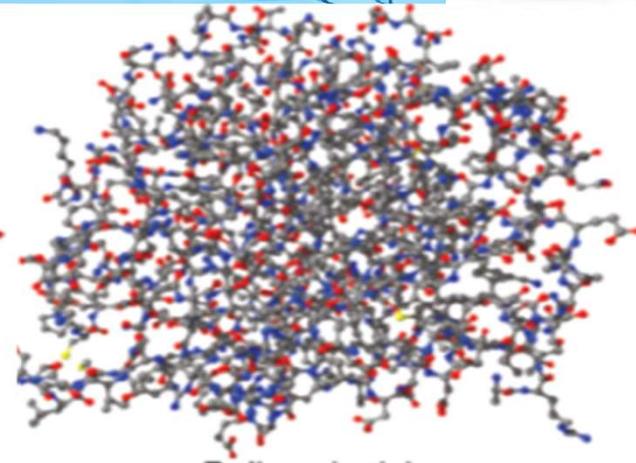
Wireframe   Stick   Ball and stick

❑ Bonds-based representation

- fast, small demands on resources

- suitable to detailed analysis

- incorrect impression about atom packing (empty space) and interatomic distances

Structure visualization

Ribbon

Cartoon
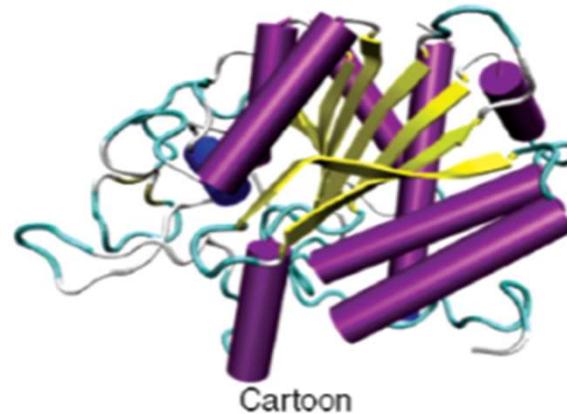
❑ **Backbone-based representation**

- moderately fast & resource-demanding

- suitable to investigation of folds and secondary structure

- good for overall orientation in the structure – shows landmarks

Structure visualization

CPK
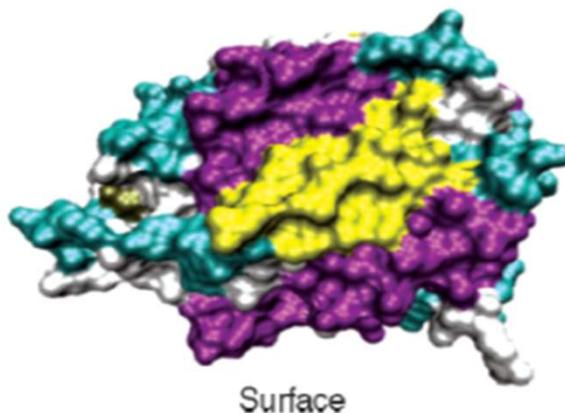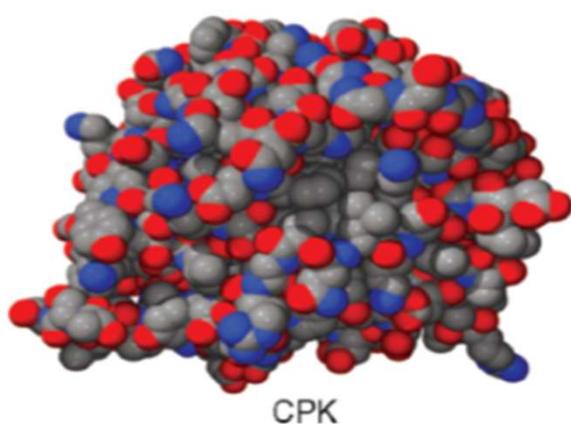
Surface

	hydrogen (H)	white
	carbon (C)	black
	nitrogen (N)	blue
	oxygen (O)	red
	fluorine (F), chlorine (Cl)	green
	bromine (Br)	dark red
	iodine (I)	dark violet
	noble gases (He, Ne, Ar, Kr, Xe)	cyan
	phosphorus (P)	orange
	sulfur (S)	yellow

❑ **Surface-based representation**

- ▪ very slow & very resource-demanding

- ▪ applicable to study of cavities & molecular contacts and shapes

Accessible surface

Molecular surface

vdW surface
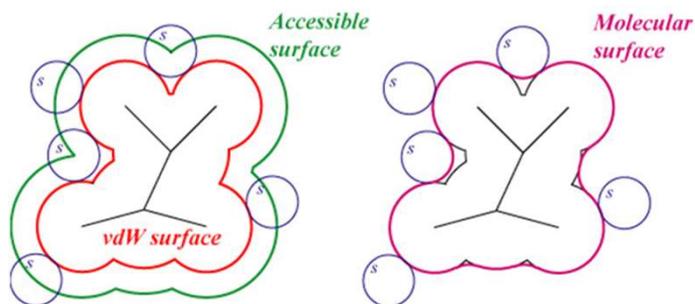
Energetics of structures

- 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

  - negative energy is favorable

  - potential **V/E$_p$** energy – interactions of atoms in a system

  - kinetic **T/E$_k$** energy – movement of atoms

$$U = E_p + E_k$$

Energetics of structures

- Entropy

  - relates to the thermal disorder of the system

  - total entropy **S** > 0

  - higher entropy is more favorable

Energetics of structures

❑ **Free energy**

- Helmholtz **A/F** (const. V), Gibbs **G** (const. P)

- both internal energy **H** (enthalpy) and entropy **S**

- **A = U – TS** and **G = H – TS; $\Delta G = \Delta H - T\Delta S$**

- negative free energy change is favorable ($\Delta G < 0$)
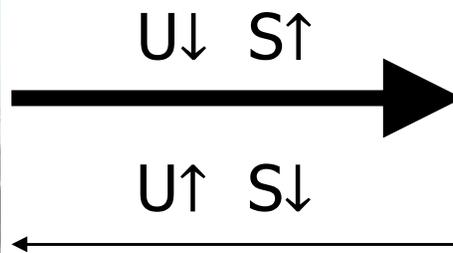
U↓ S↑

U↑ 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 structures occur more frequently
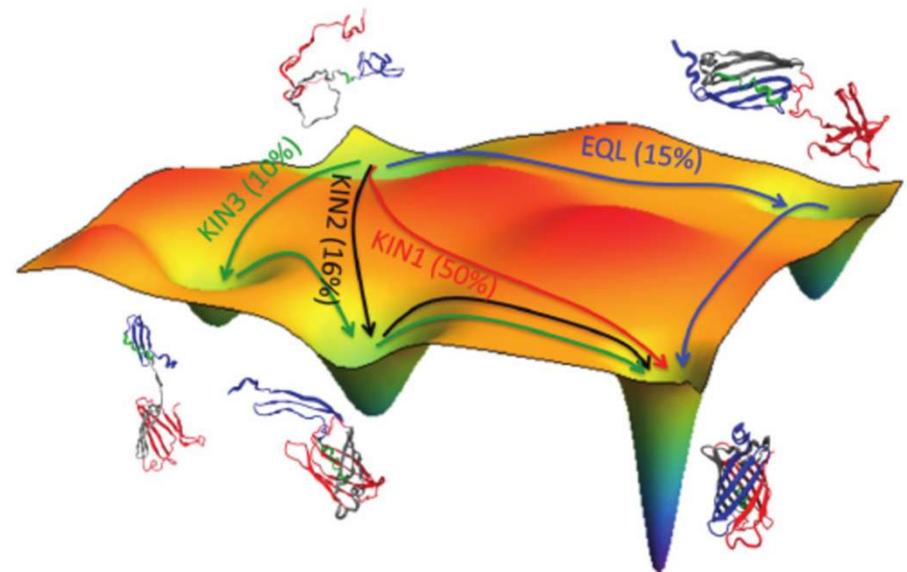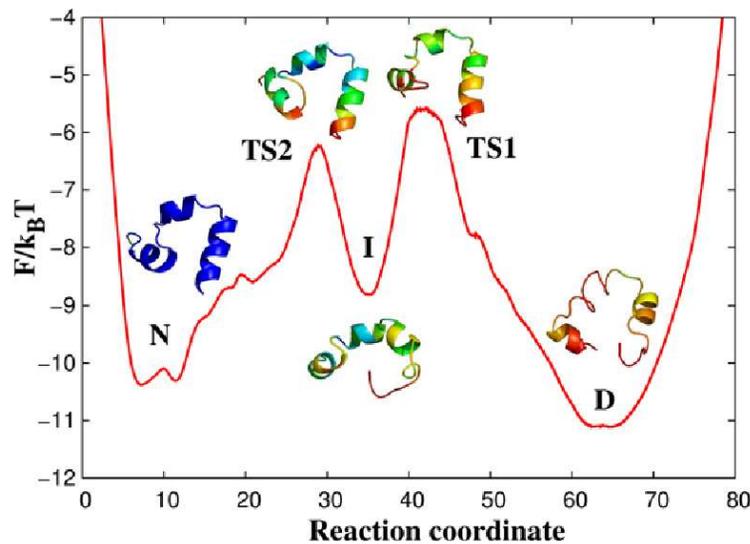
Energy landscape

- 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 structures occur more frequently

- Potential energy surface

  - multidimensional surface

  - minima – stable structures

  - saddle points – transient str

  - maxima – unstable structure
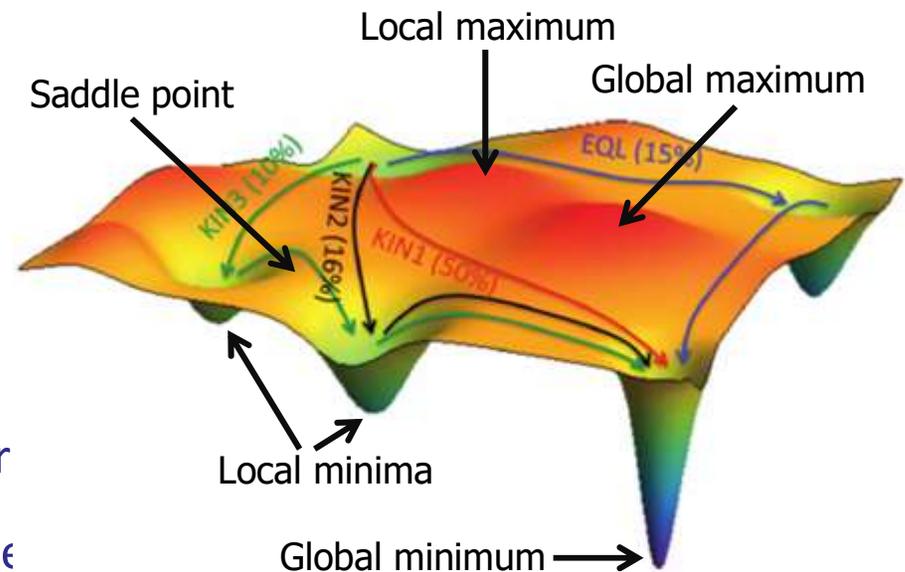
Molecular interactions

□ Covalent interactions (chemical bonds)

- sharing of electrons

- very stable under standard condition

□ Non-covalent interactions

- much weaker than covalent

- electrostatic interactions

- polar interactions

- non-polar interactions

Electrostatic interactions

❑ Charge-charge interactions

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

- Coulomb's law

- long-range interaction – decrease with r^2

- environment dependent

  - permittivity: $\varepsilon = \varepsilon_0 \cdot \varepsilon_r$

  - relative permittivity (ε_r) = dielectric constant

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

Electrostatic interactions

❑ Charge-charge interactions

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

- Coulomb's law

- long-range interaction – decrease with r^2

- environment dependent

  - permittivity: $\varepsilon = \varepsilon_0 \cdot \varepsilon_r$

  - salt concentration – presence of counter ions

  - pH – change in charge of molecule

|	Low pH	High pH	Very high pH
Amines:	NH_3^+	NH_2	NH_2
Carboxylic acids:	OH	O^-	O^-
Alcohols:	OH	OH	O^-
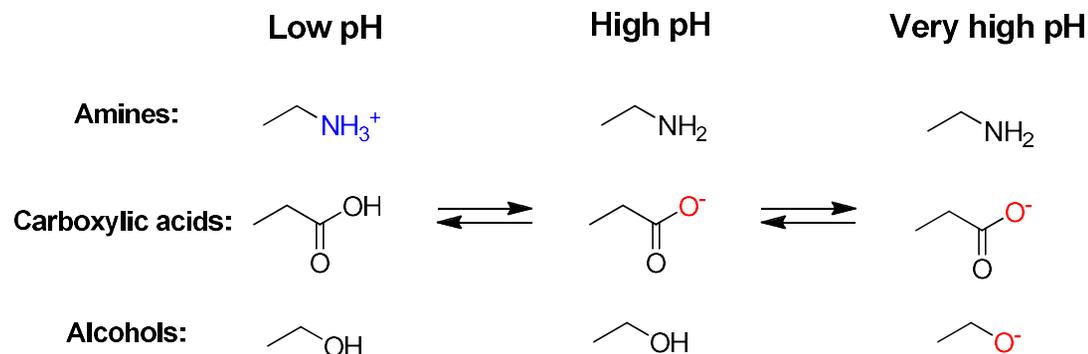
Polar interactions

- ❑ **Hydrogen bonds (H-bonds)**

  - ▪ donor and acceptor atoms sharing hydrogen

  - ▪ H-bond distance: 2.8-3.4 Å

  - ▪ with highly electronegative atoms: F, O, N

acceptor (δ^-)   donor (δ^+)
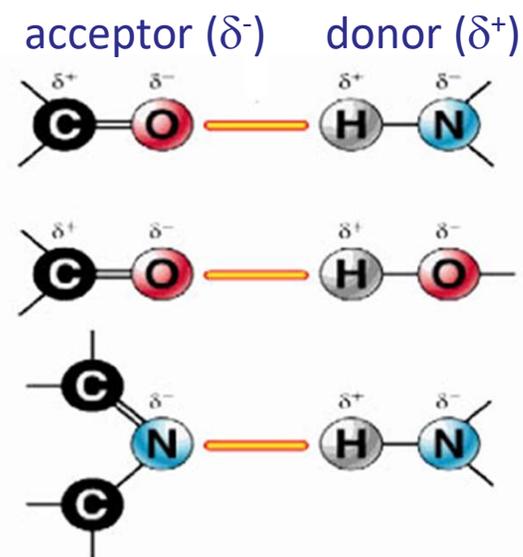
- ❑ **Aromatic (π-π) interactions**

  - ▪ attractive interaction between aromatic rings

  - ▪ cen – a

  sandwich

  parallel displaced

  T-shaped
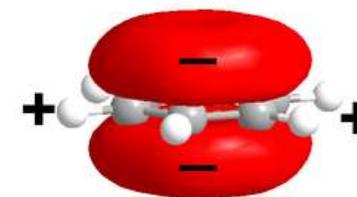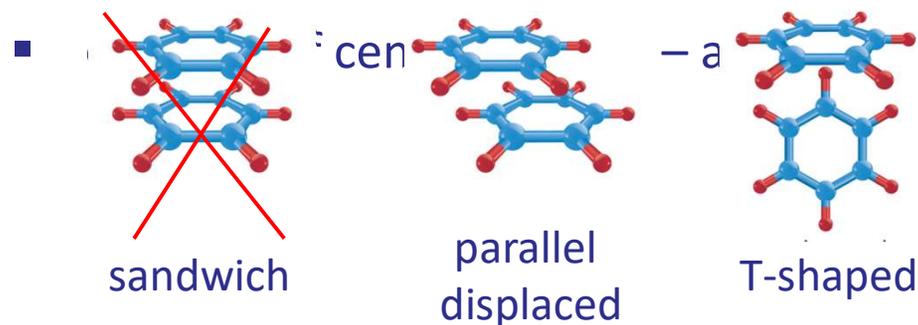
Non-polar interactions

□ **van der Waals (vdW) interactions**

- **between any two atoms**

- **important in non-polar molecules**

- **short-range interactions – up to 5 Å**

$$F_{\text{VdW}}(r) = -\frac{AR_1 R_2}{(R_1 + R_2)6r^2}$$

R_1, R_2 – van der Waals radii
r - distance
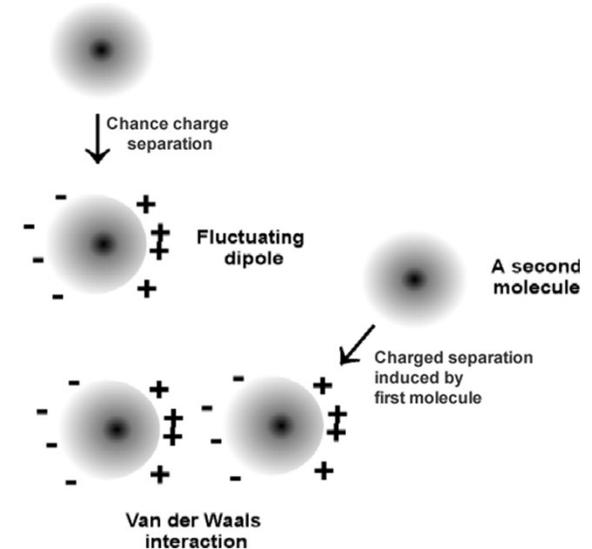
□ **Hydrophobic interactions**

- **entropic origin – water molecules ordered**

  around hydrophobic moiety -> unfavorable

- **hydrophobic packing -> favorable release**

  of some ordered water molecules
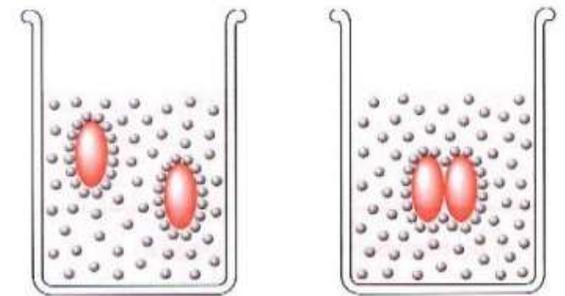
# Protein folding game

❑ FOLD.IT – http://fold.it/portal/

- crowdsourcing online game

- prediction of protein structures
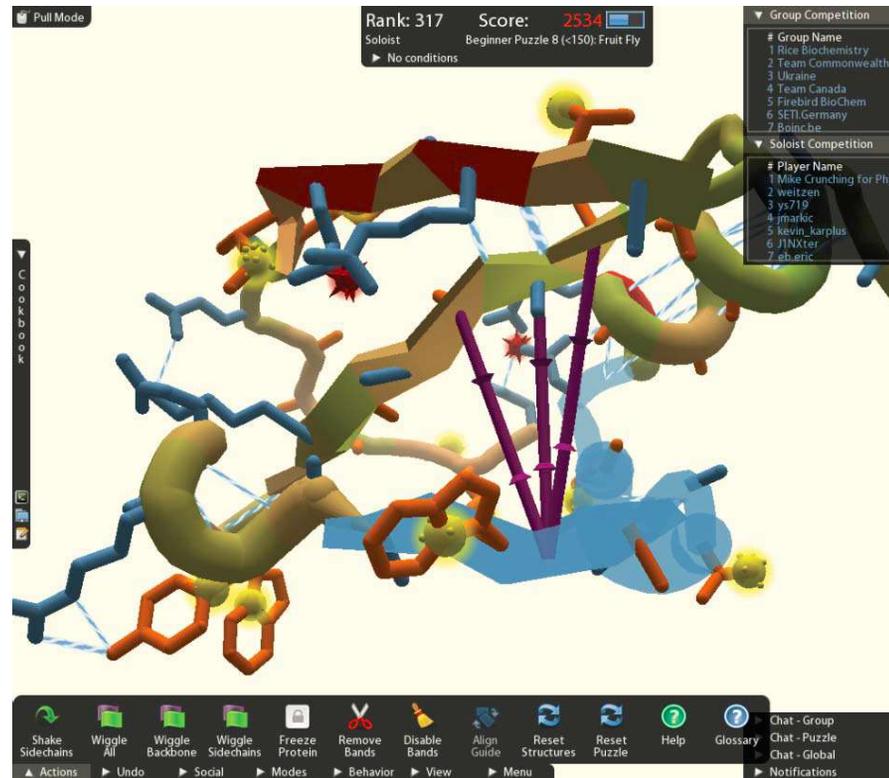
Structure determination

- ❑ **Established methods**

    - ▪ X-ray crystallography

    - ▪ NMR spectroscopy

    - ▪ electron microscopy

   - ▪ bioinformatics predictions – theoretical
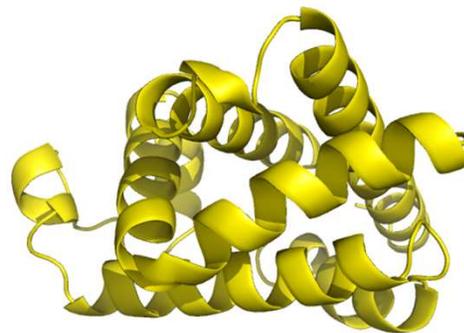
X-ray crystallography

- Crystallization procedures
    - slow (days-weeks)
   - high risk of failure
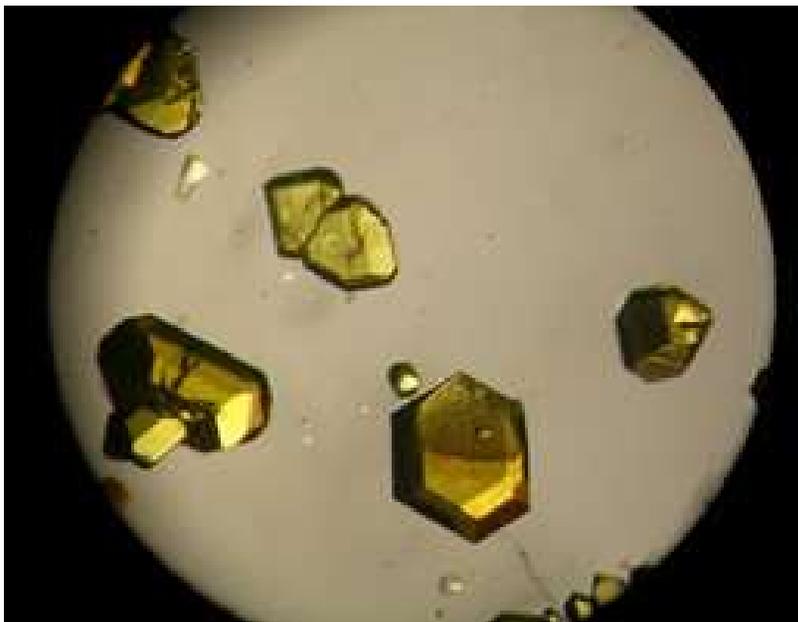
**Some Crystallization Methods:**

Vapor diffusion
Hanging-drop
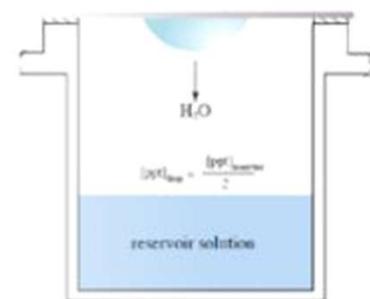
Sitting-drop
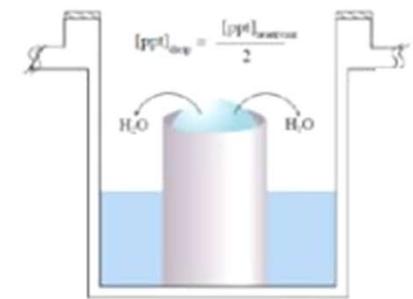
Batch:
micro batch under oil
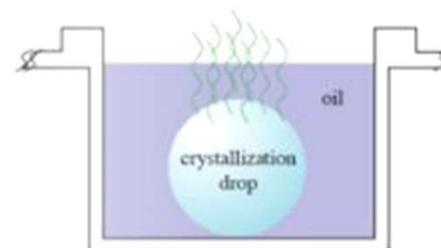
Dialysis
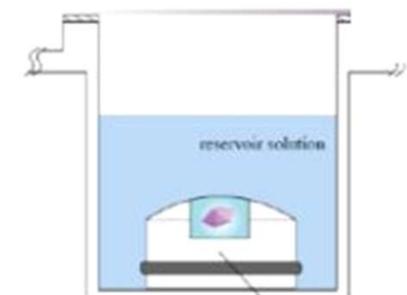
X-ray crystallography

Data Collection
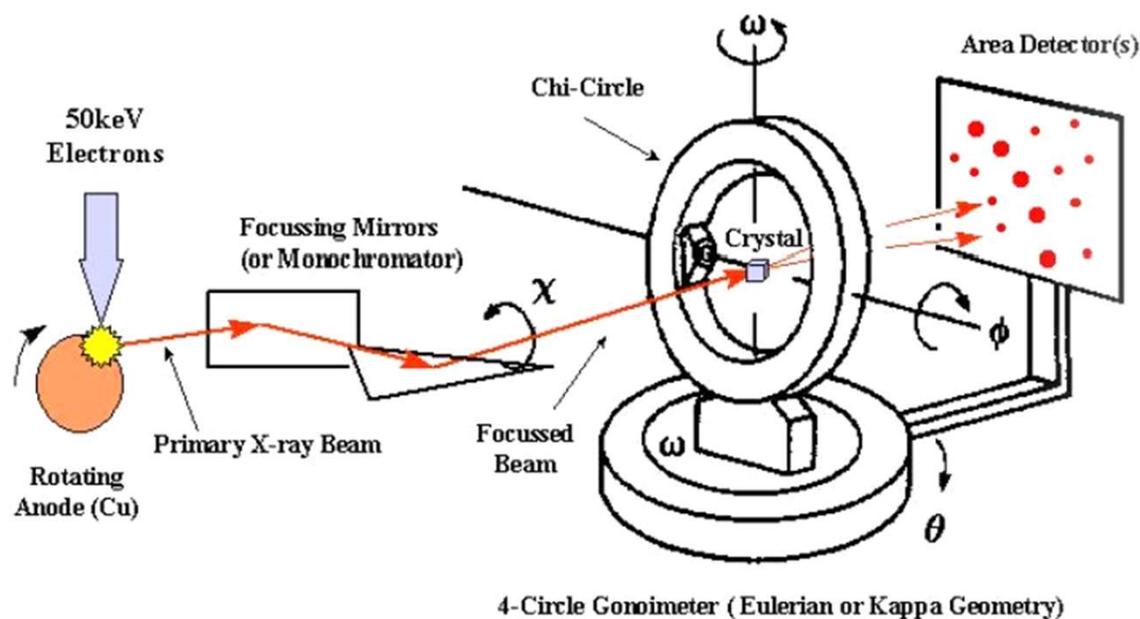
X-ray sources: X-ray tubes, rotating anodes and synchrotrons.
Synchrotrons produce the brightest X-rays

APS Chicago

European Synchrotron Radiation Facility, Grenoble

X-ray crystallography

Image of diffraction
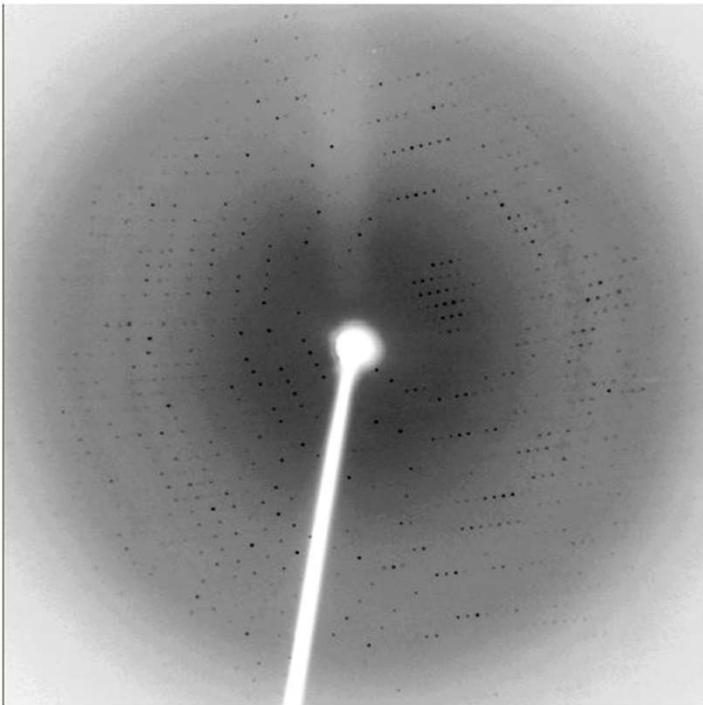
Electron density map
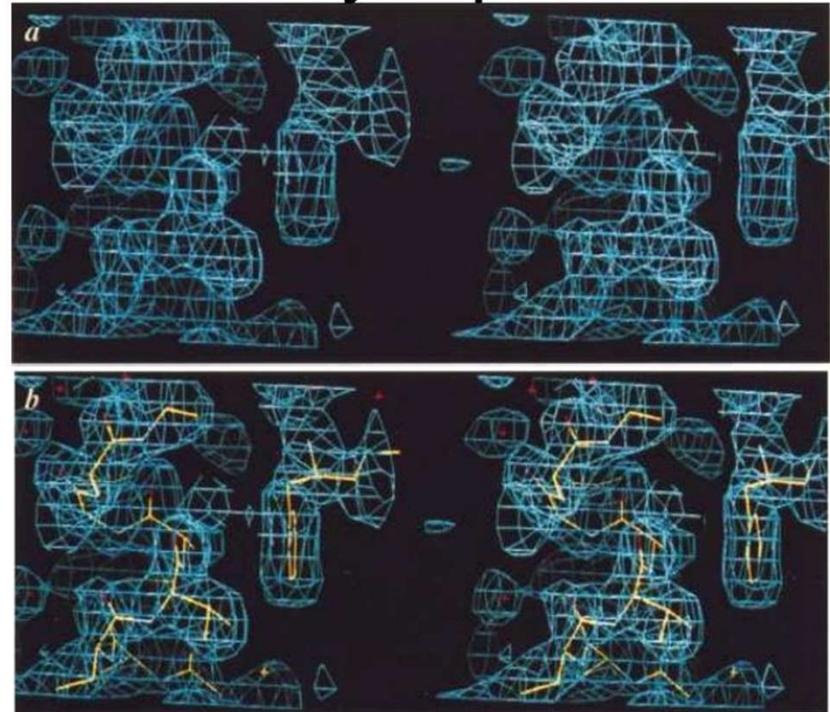
Building a structure model

X-ray crystallography

- ❑ Crystallization
  - ▪ hanging drop, sitting drop, microbatch

- ❑ Data collection
  - ▪ diffractometers, synchrotrons

- ❑ Analysis of diffraction data
  - ▪ solving phase problem
    - ▪ isomorphous replacement
    - ▪ molecular replacement
    - ▪ anomalous scattering

- ❑ Iterative model building
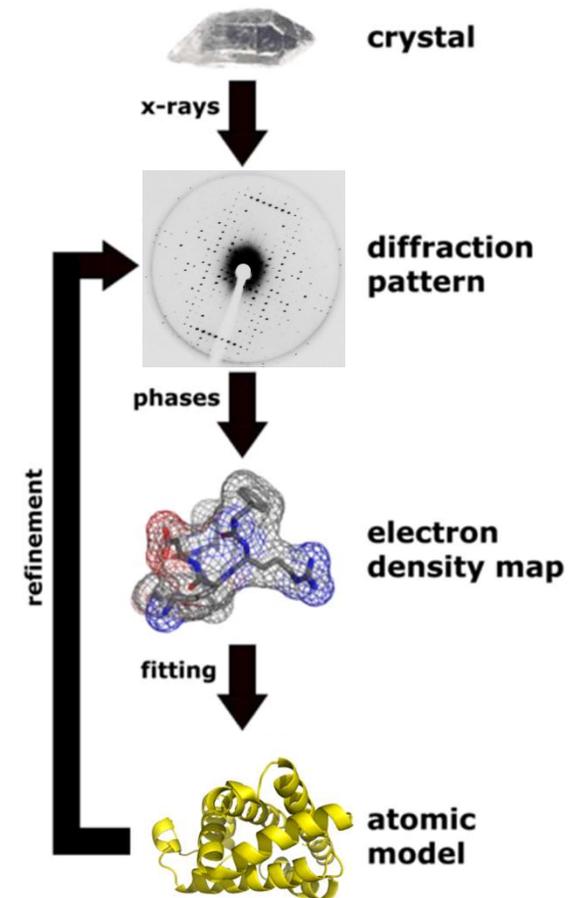
Parameters of an X-ray structure

❑ Resolution

▪ measure of the level of detail present in the diffraction pattern

3 Å   **2 Å**   **1 Å**
(bad)   (acceptable) (exceptional)
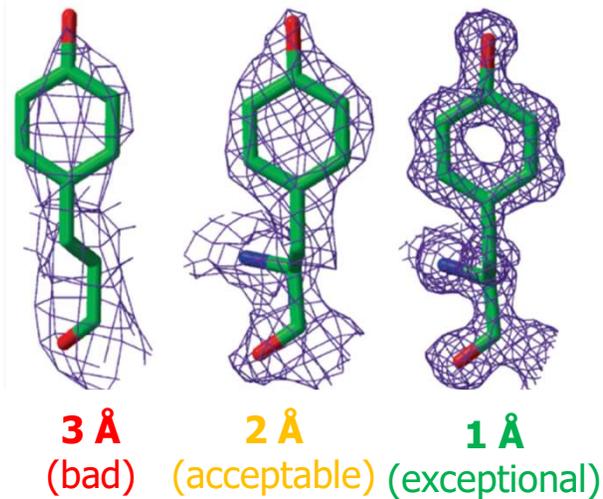
❑ R-factor (R-value)

▪ measure of a model quality - i.e. the agreement between the crystallographic model and the diffraction data

▪ values from 0 (ideal) to 0.63 (random structure), typically about 0.2

Parameters of an X-ray structure

❑ Thermal factors (B-factors)

▪ measure of how much an atom oscillates or vibrates around the

position specified in the model
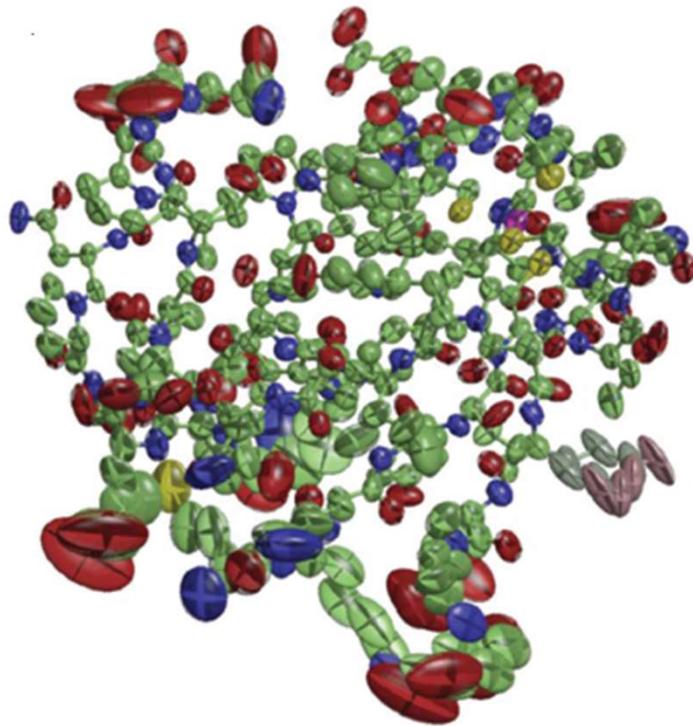
X-ray crystallography

❑ Advantages

   ▪ no limitations in size

   ▪ possibility to obtain an atomic resolution

❑ Disadvantages

   ▪ requirement of a crystal

   ▪ structure in a crystalline (non-native) state

   ▪ static picture of macromolecule

   ▪ position of hydrogen atoms (usually) 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
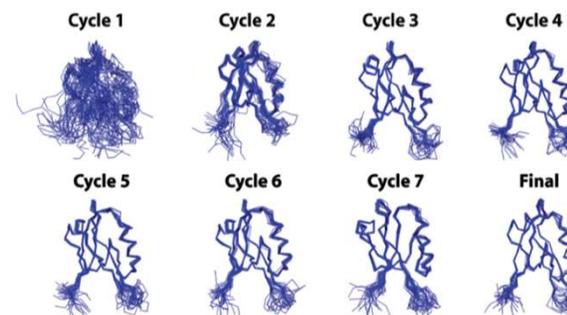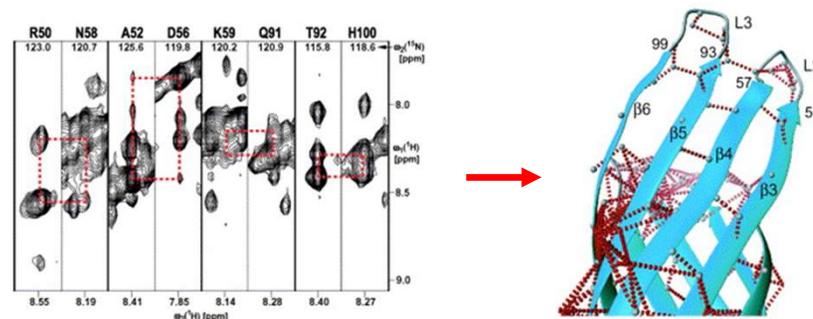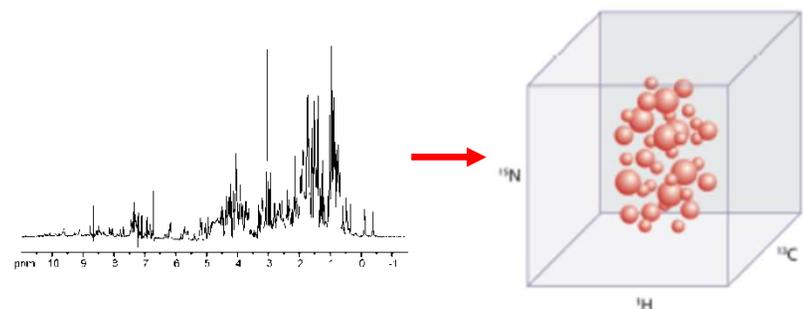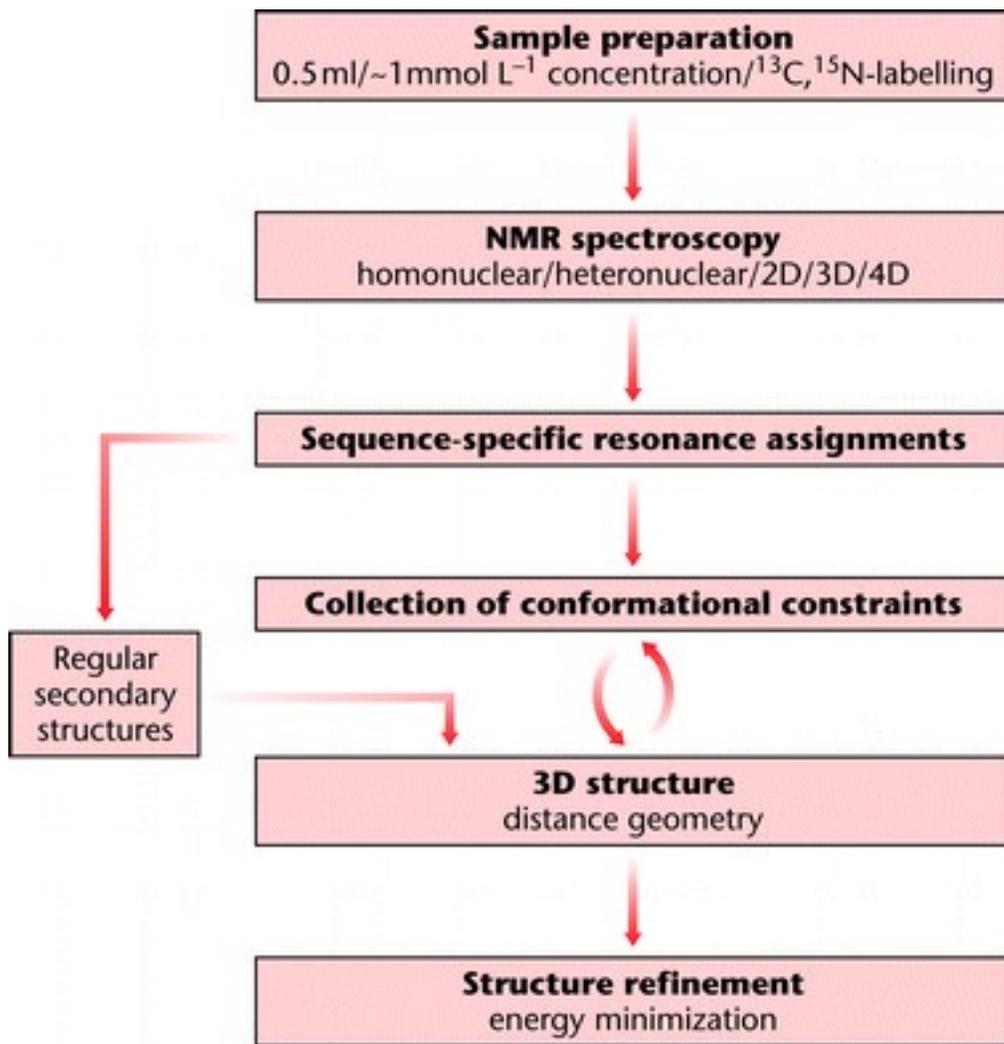
Parameters of an NMR structure

- RMSD

  - root-mean-squared deviation of atomic positions across the ensemble of solutions

  - reveals differences between individual conformations

  - general parameter in many fields of structural biology
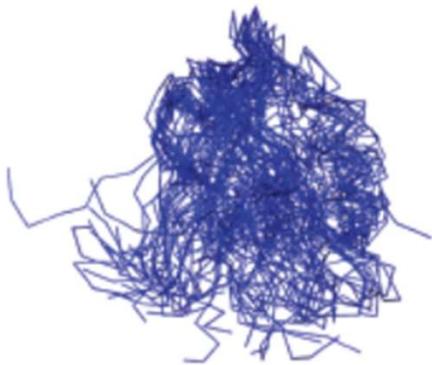
$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2}$$

RMSD = 3.59 Å   RMSD = 1.06 Å   RMSD = 0.42 Å

NMR spectroscopy

❑ Advantages

- structure in solution (native) state

- possibility to investigate dynamics of macromolecule

- position of hydrogen atoms detected

❑ Disadvantages

- size limited to approximately 40 kDa

- requirement of isotopically labeled sample

Electron microscopy
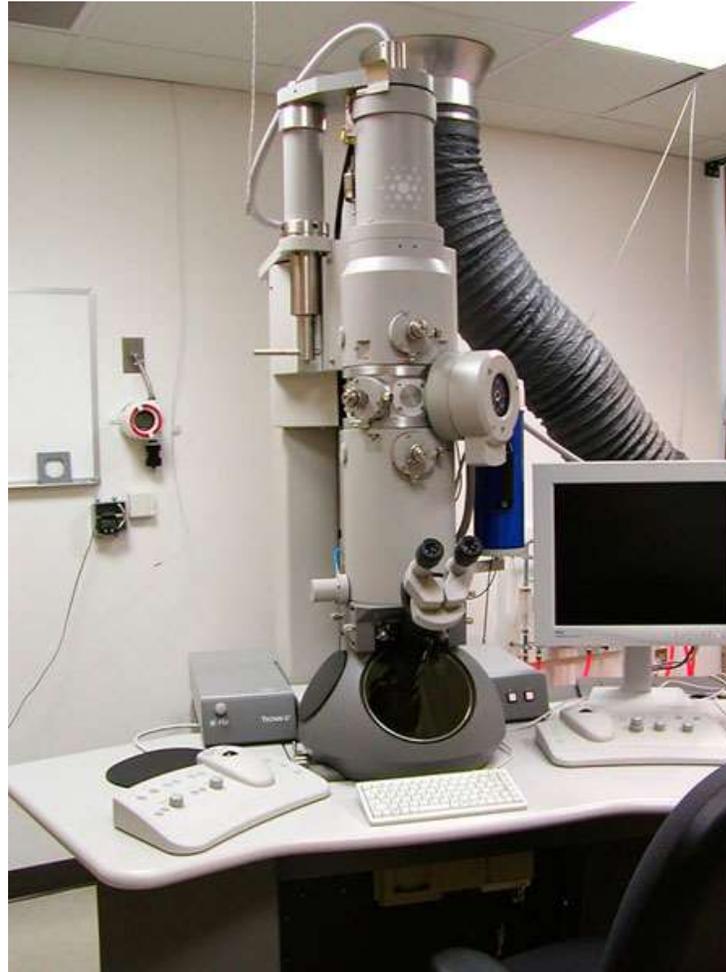
FEI Tecnai T12 Cryotransmission Electron Microscope

Electron microscopy

- wavelength of an electron is much shorter than the wavelength of light

- so it can reveal much smaller things

- samples are flash-frozen in their natural environments (cryo-EM)

- can generate 3D images of large molecules at nearly atomic resolution

Recontruction of Adenovirus by CryoTEM

ELECTRON BEAM
e⁻

The electron beam scans each protein.

SUSPENDED PROTEINS IN ICE

DIRECT ELECTRON DETECTOR
captures the projection image of each protein

The projection images are categorized into like groups.
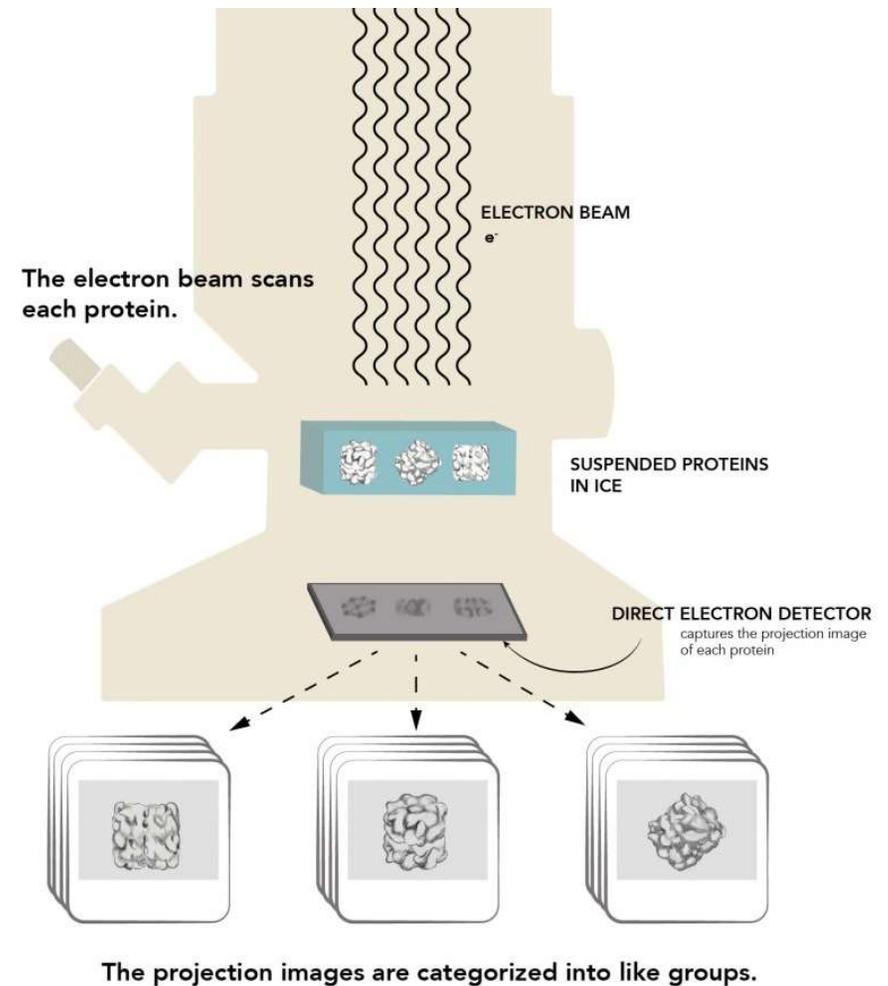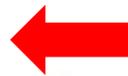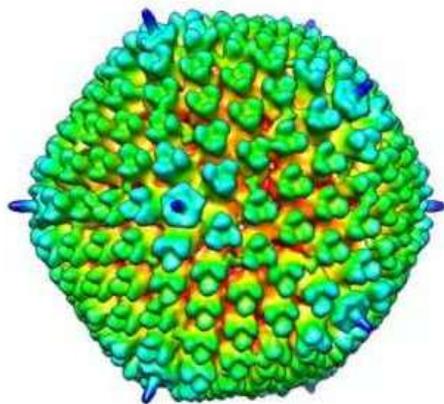
Electron microscopy

❑ Advantages

- applicable to extremely large systems

- complements other methods e. g. X-ray, NMR
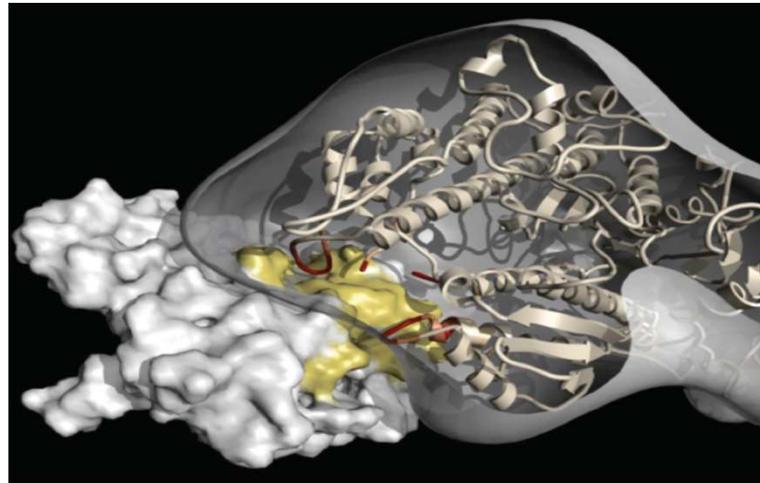
❑ Disadvantages

- lower resolution (2-3 Å at best)

Bioinformatics predictions

- ❑ Homology modeling

- ❑ Machine learning

- ❑ *Ab initio* prediction

# Bioinformatics predictions

Comparative modelling
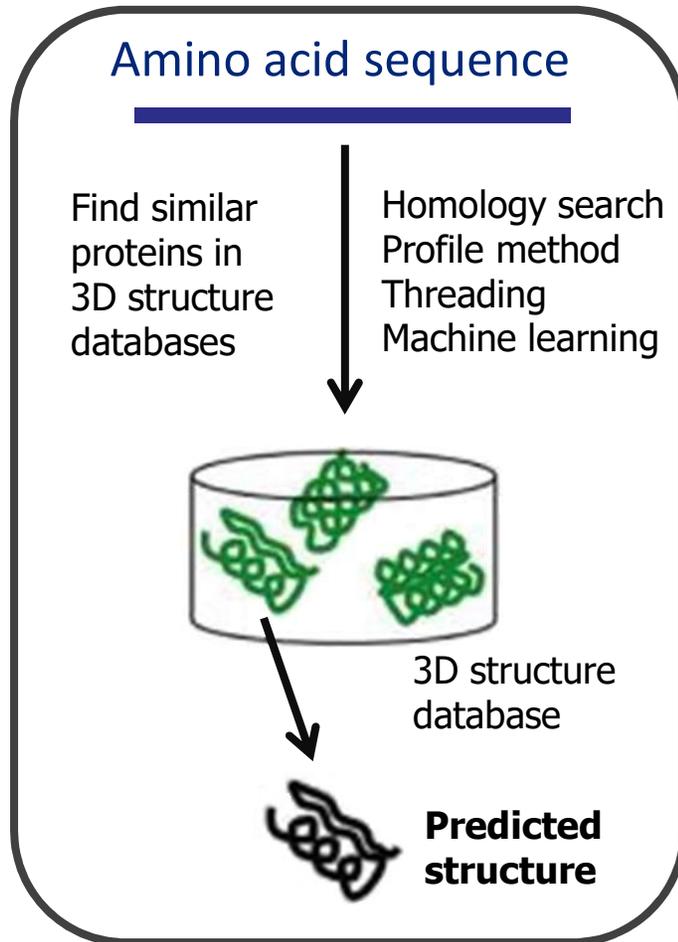
Amino acid sequence
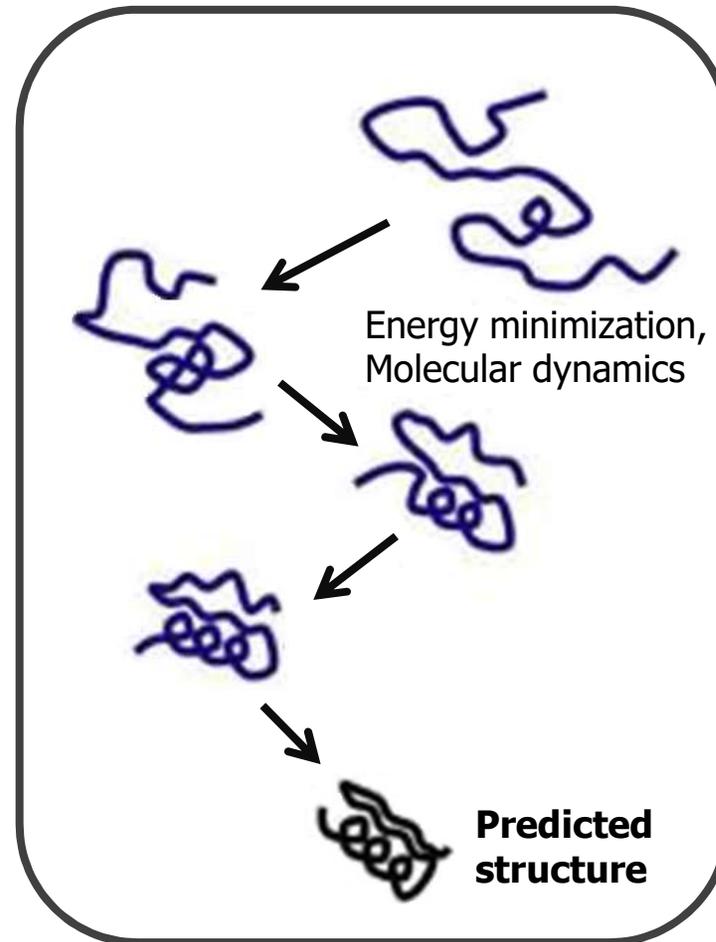
Find similar proteins in 3D structure databases

Homology search
Profile method
Threading
Machine learning

3D structure database

Predicted structure

Ab initio predictions

Energy minimization, Molecular dynamics

Predicted structure

Bioinformatics predictions

❑ Homology modeling

Comparison of sequences in databases
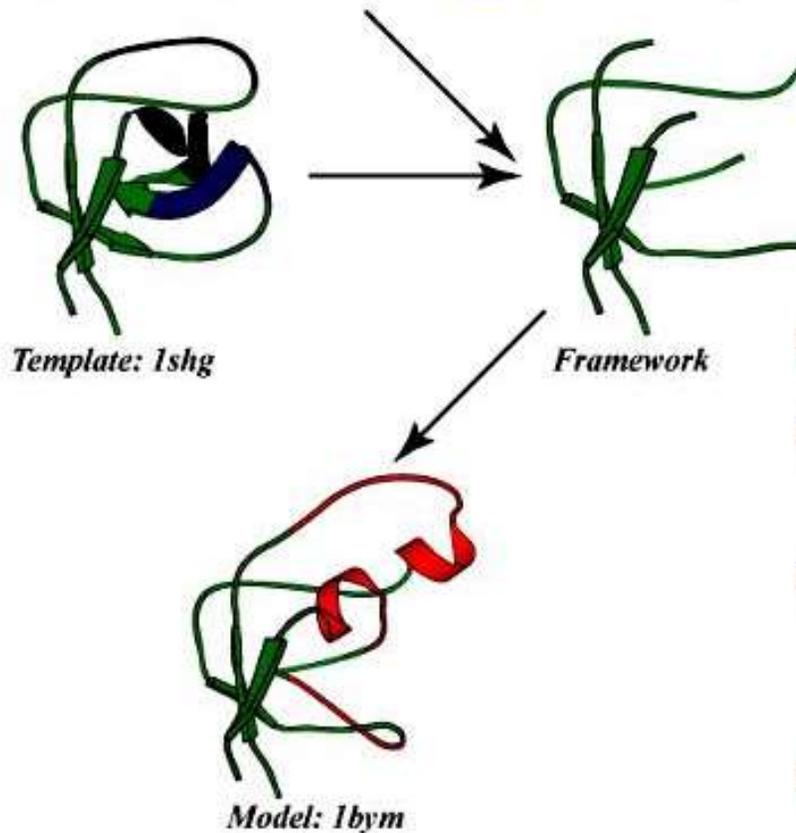
o *Find template*

o *Align target sequence with template*

o *Generate model:*
     - add loops
     - add sidechains

o *Refine model*

Bioinformatics predictions

❑ Machine learning

   ❑ Training on sequence and 3D databases

   ❑ Ex.: AlphaFold 2
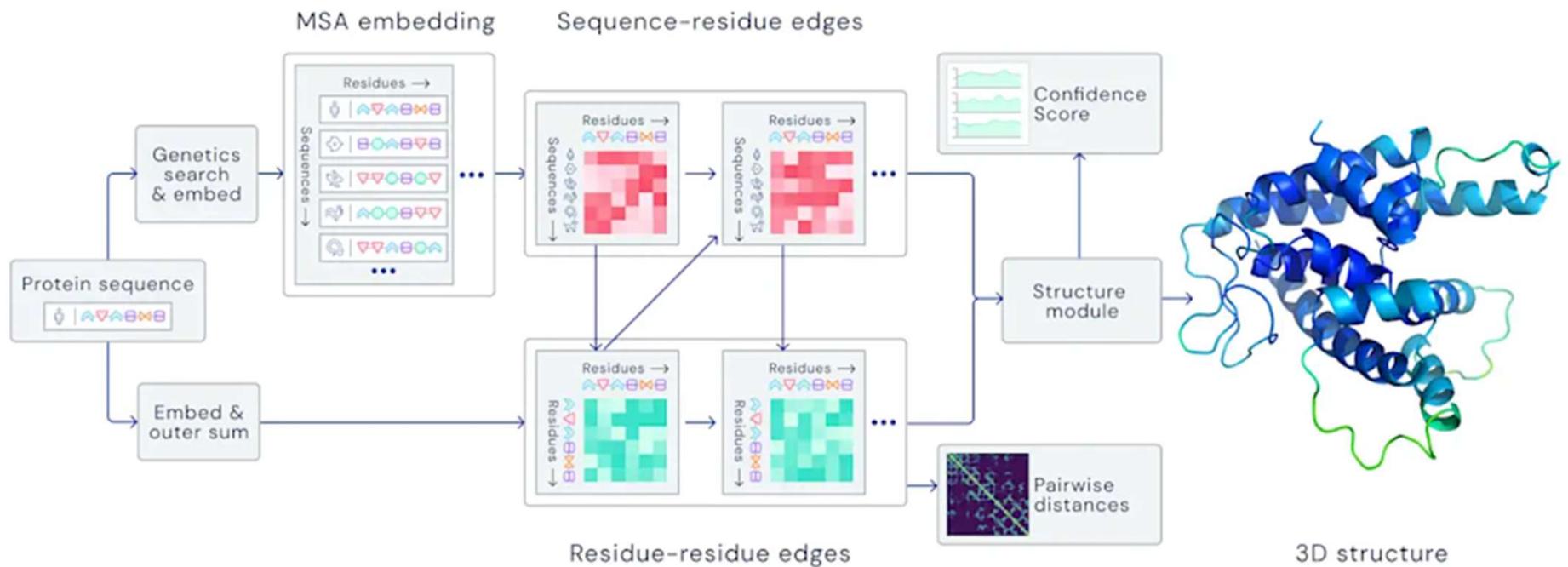
Bioinformatics predictions

❑ *Ab initio* prediction

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

$$+ \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \sum_{l} U_l (1 \pm \cos(n_l \varepsilon_l))$$

Random Coil Structure

Intermediate Secondary Structure Formation

Final 3-dimensional Native Structure
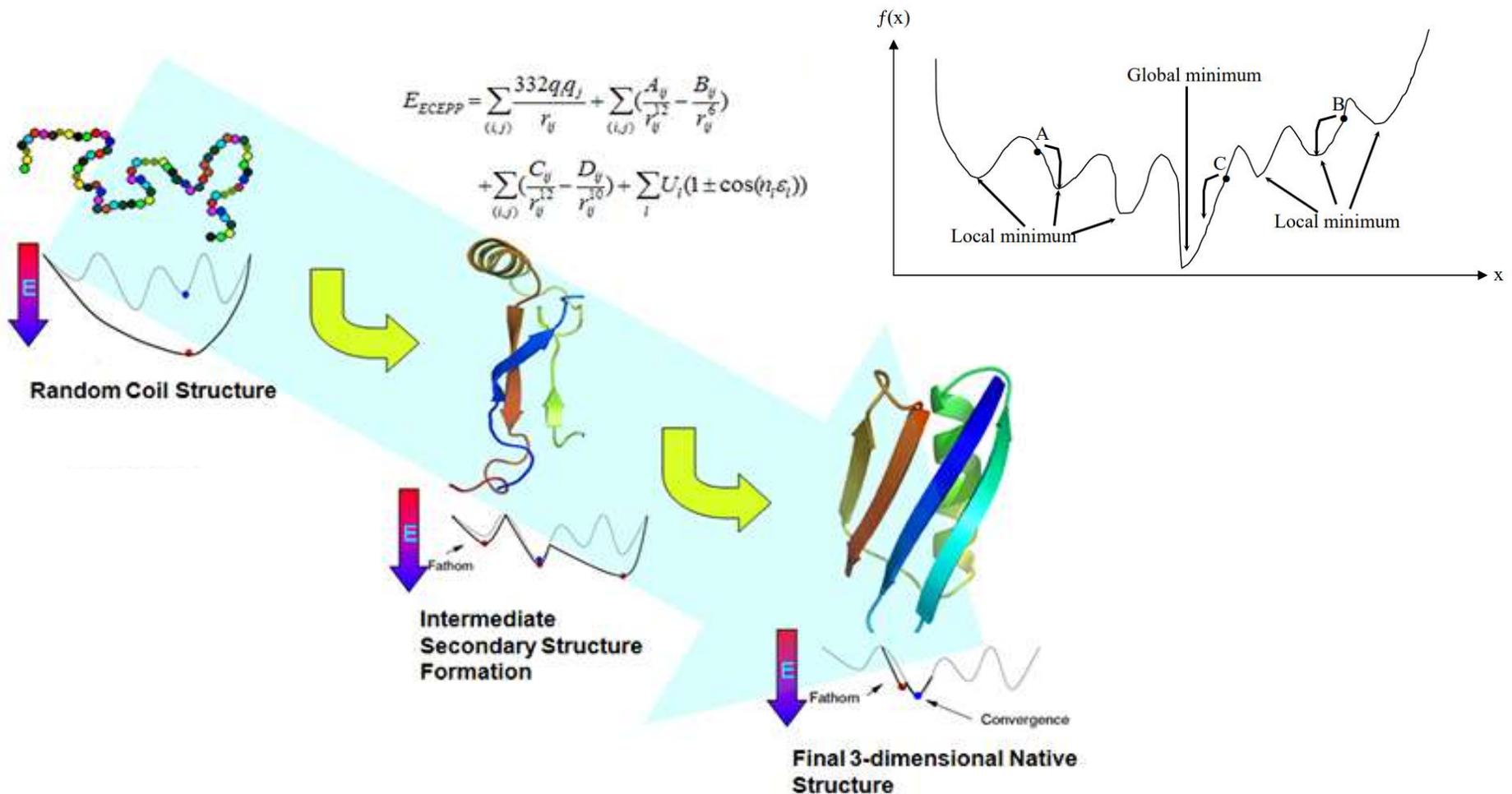
Bioinformatics predictions

- Advantages

  - very fast

  - low cost

- Disadvantages

  - theoretical model – experimental validation is needed

  - *Ab initio* is very demanding

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.

❑ Electrostatic Interactions in Protein Structure, Folding, Binding, and

❑ Zhou, H-X. & Pang, X. (2018) Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical Reviews*. **118**: 1691–1741