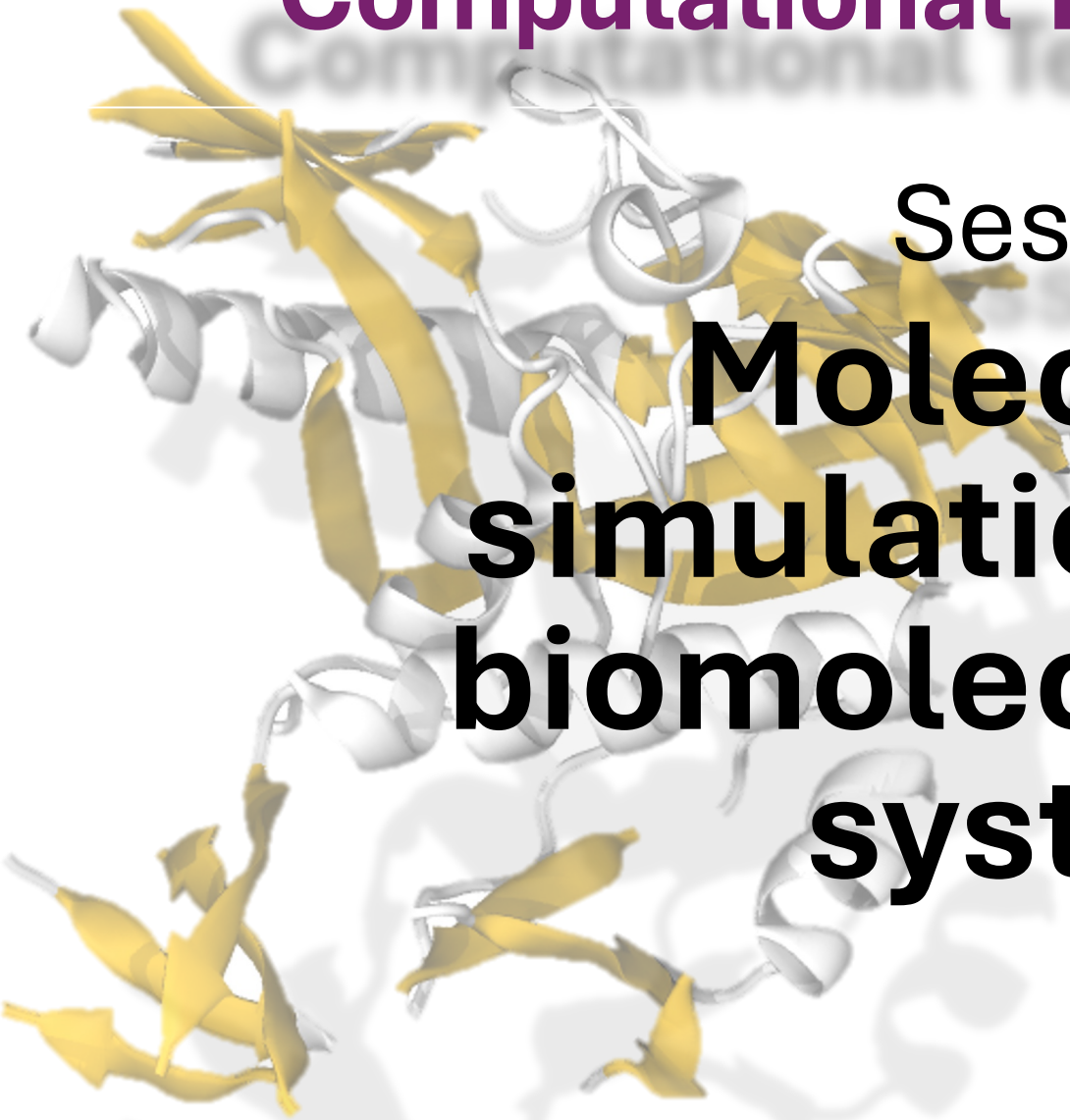
Computational Techniques in Chemistry



Session 3:

Molecular simulation of biomolecular systems

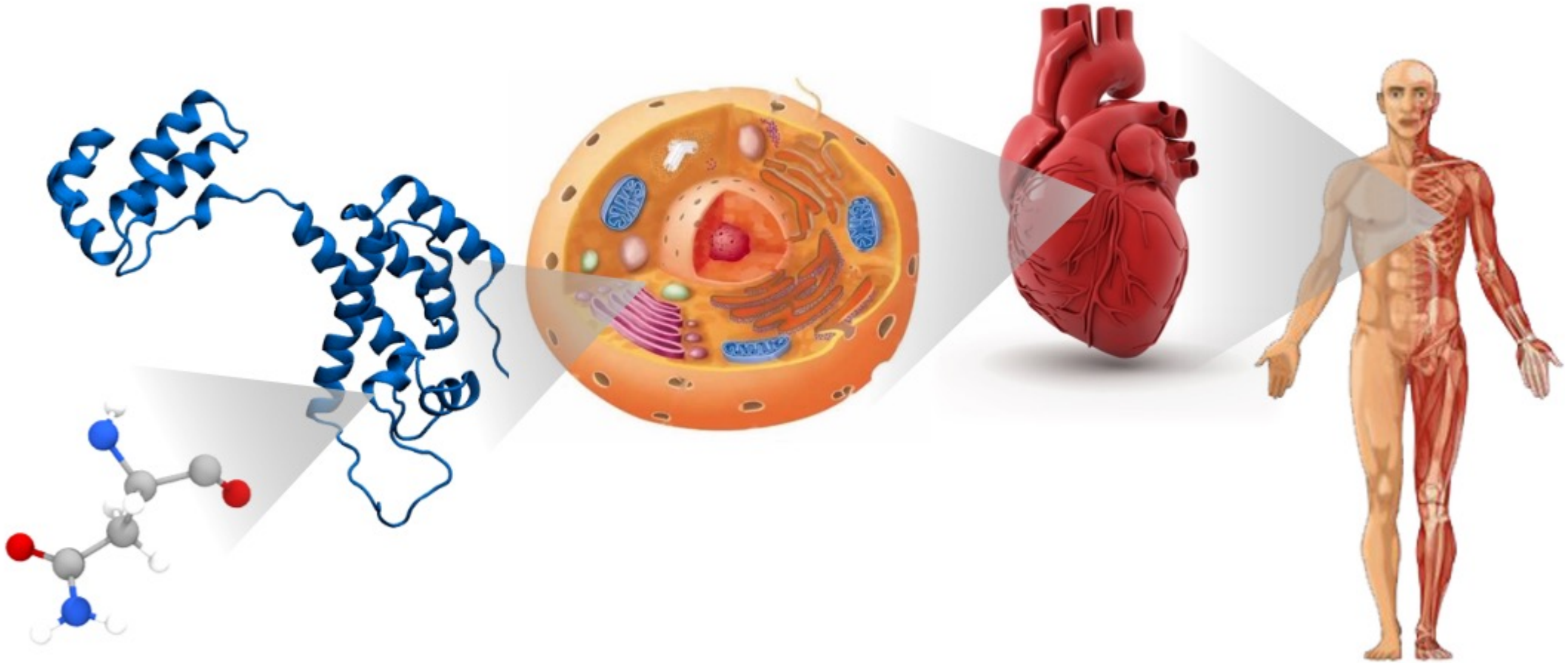
Delivered by **Tim Spankie**

University of Edinburgh

January 2025

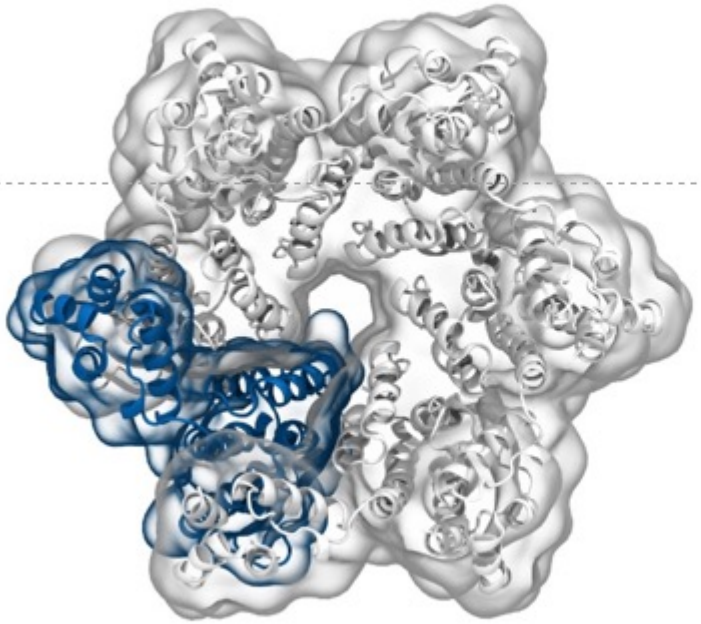
Based on content from a ten-hour series https://github.com/CCPBioSim/BioSim-analysis-workshop by Matteo Degiacomi (matteo.degiacomi@ed.ac.uk) and Antonia Mey (antonia.mey@ed.ac.uk)

The structure determines the (mal)function



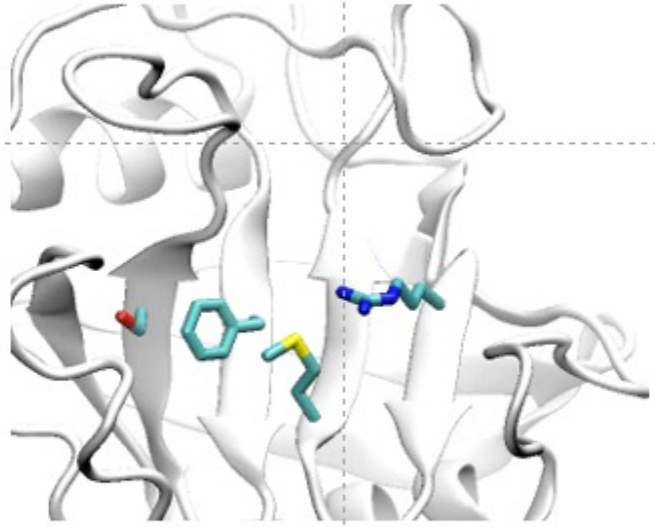
Structure and dynamics determine protein (mal)function

HIV Capsomer



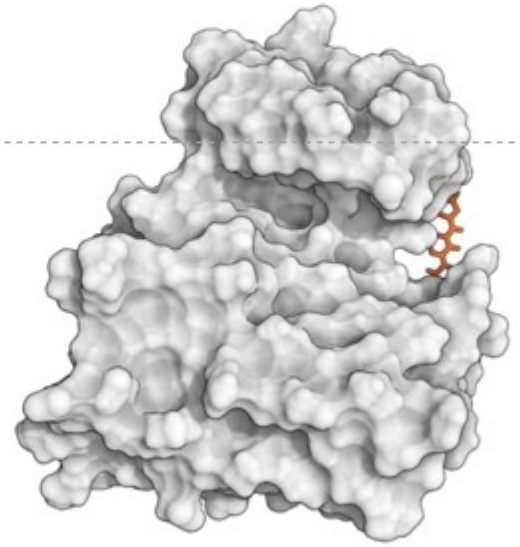
M.T. Degiacomi, ***Structure***, 2019

Cyclophilin



Wapeesittipan, Mey, et al., ***Comms. Chem.***, 2019

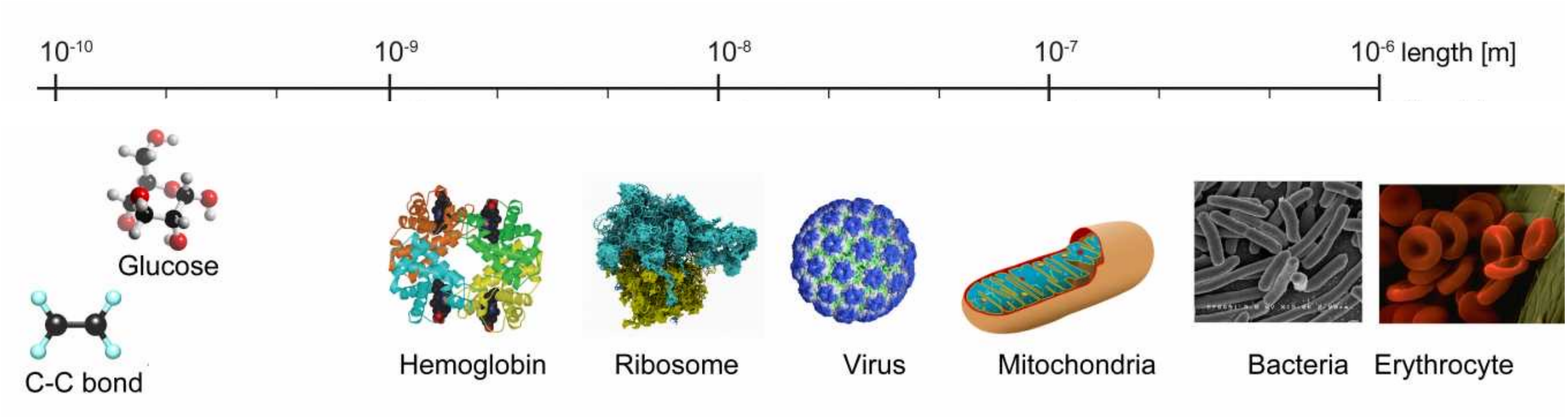
Tyrosine kinase —dasatanib



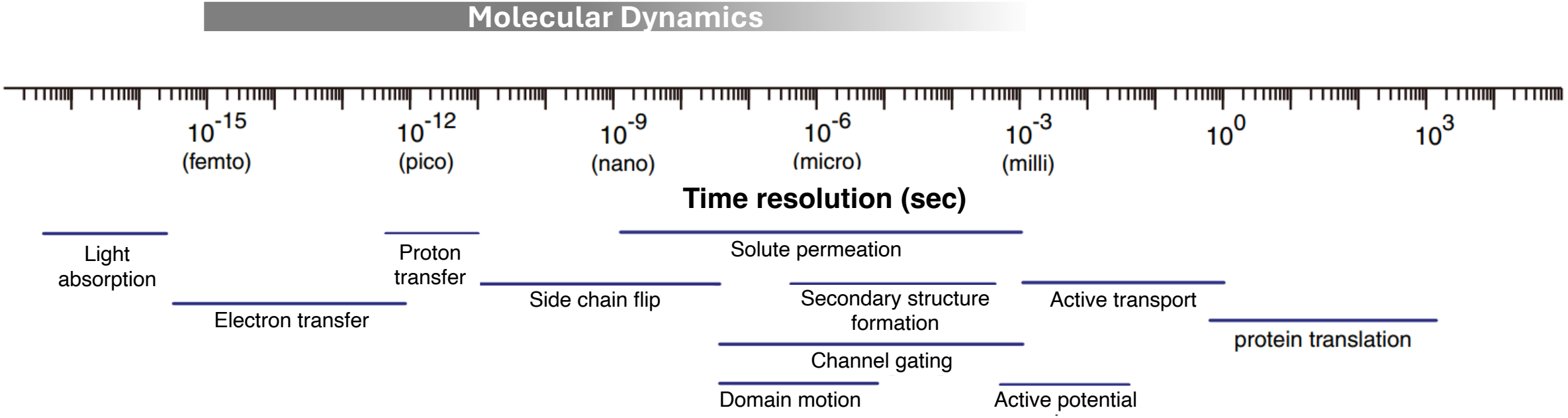
Y Shan et al. ***JACS***, 2011

Sizes in Biochemistry

Molecular Dynamics



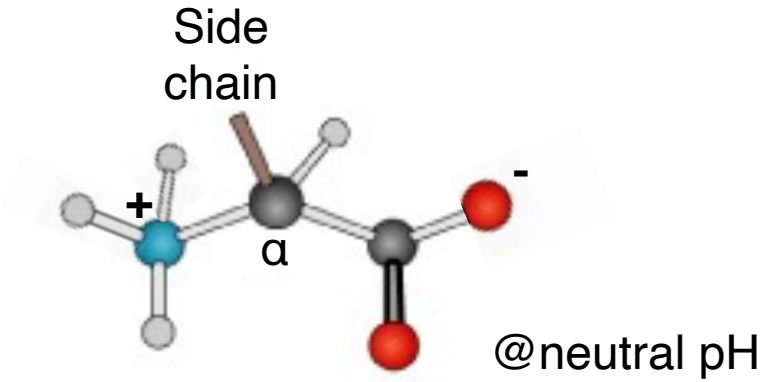
Timescales in biochemistry



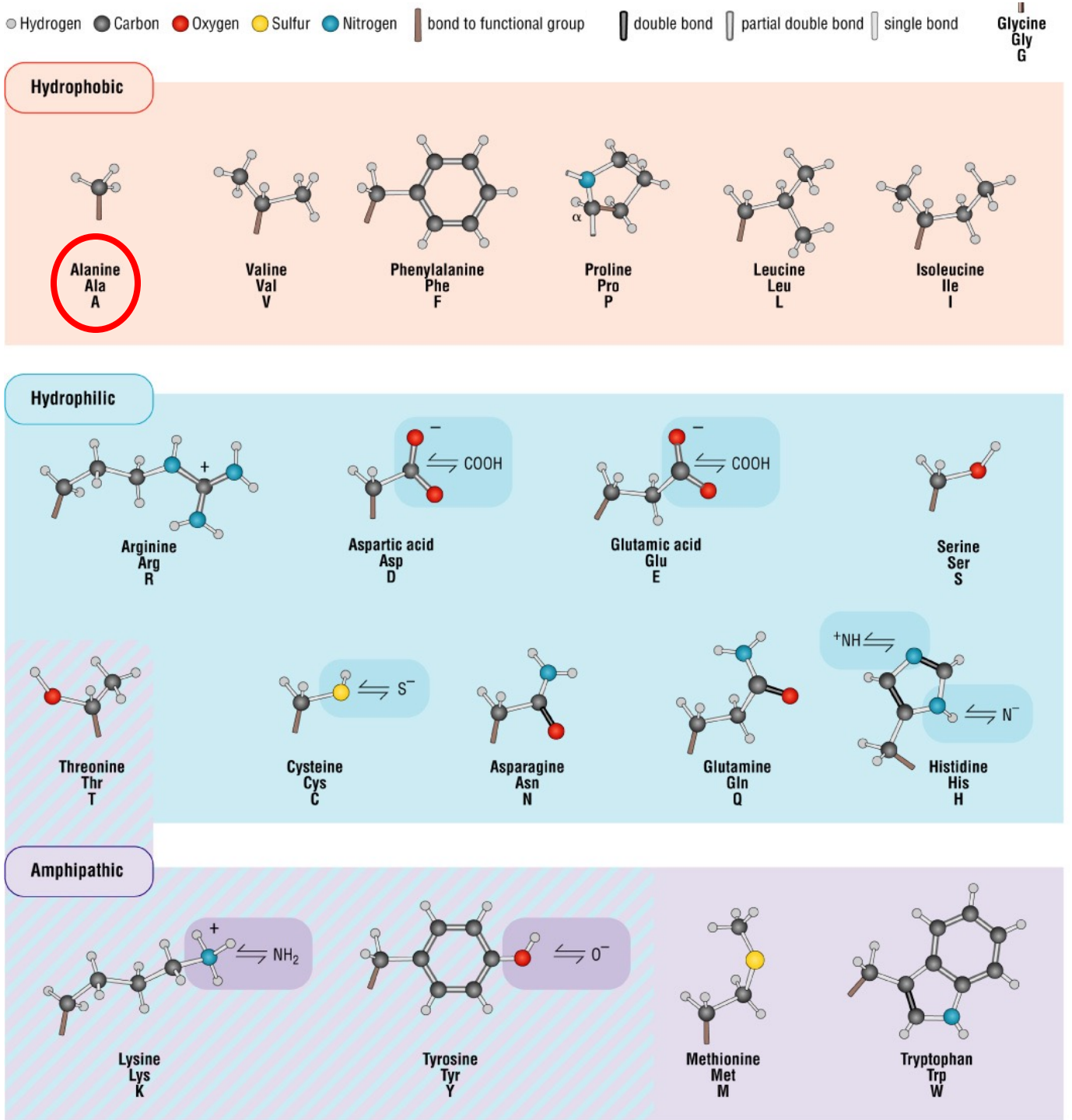
Proteins are amino acids polymers

Amino acids are composed of:

- **Backbone** (conserved)
- **Side chain** (variable)



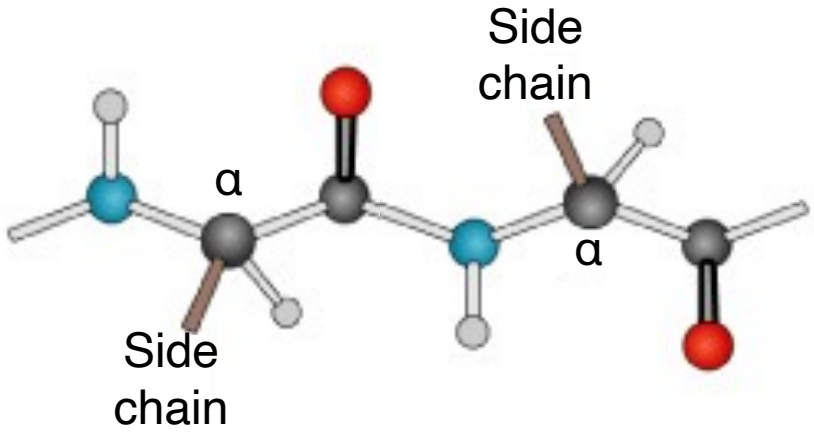
@neutral pH



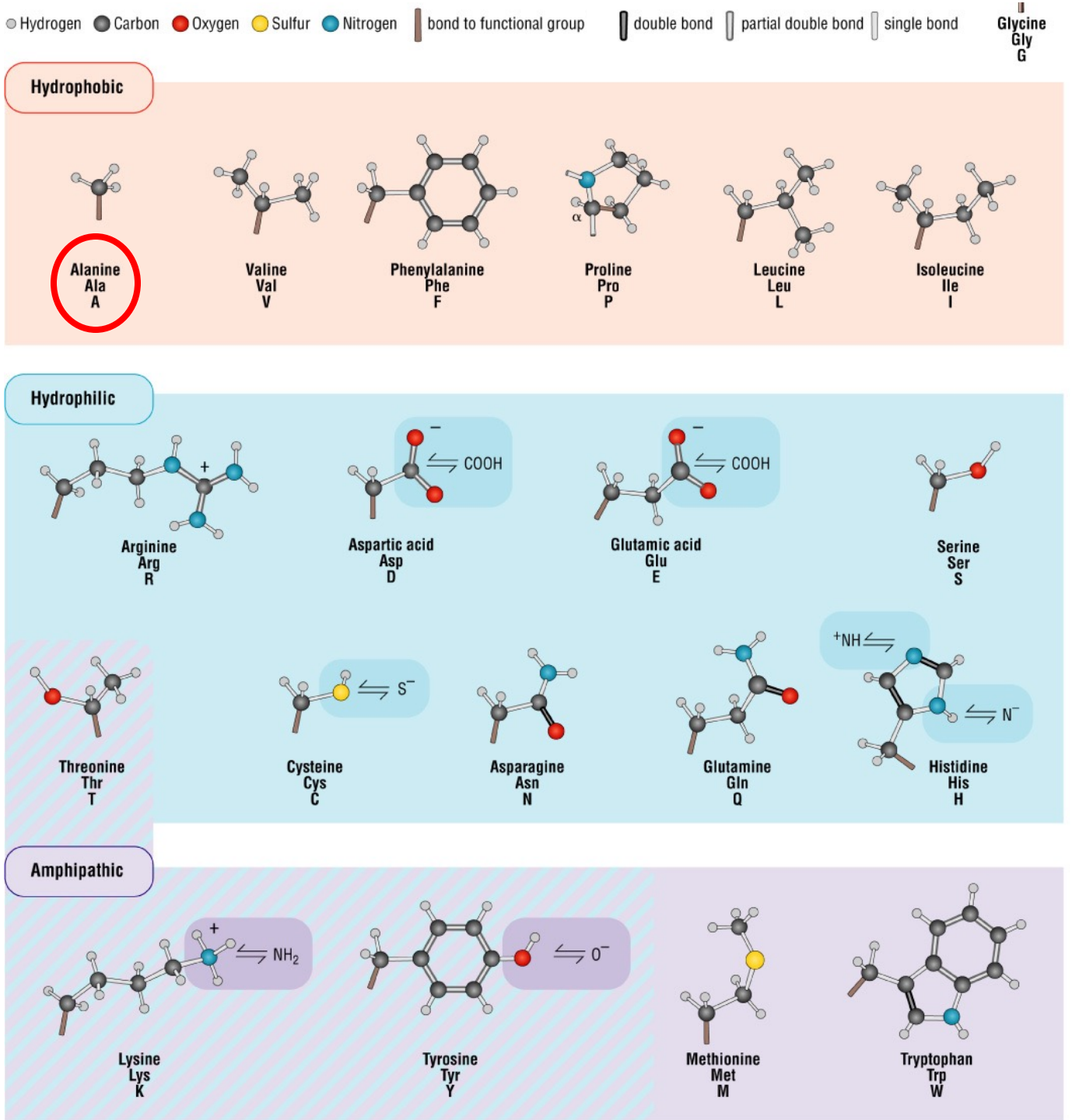
Proteins are amino acids polymers

Amino acids are composed of:

- **Backbone** (conserved)
- **Side chain** (variable)

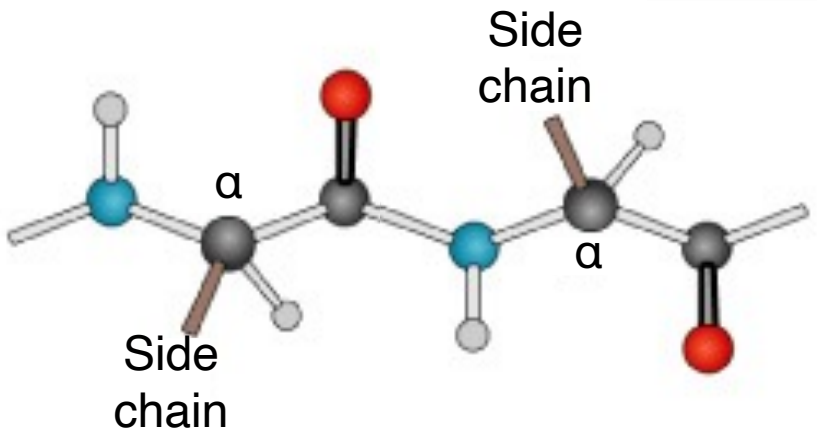


Amino acids polymerize forming a *peptidic bond* (condensation)

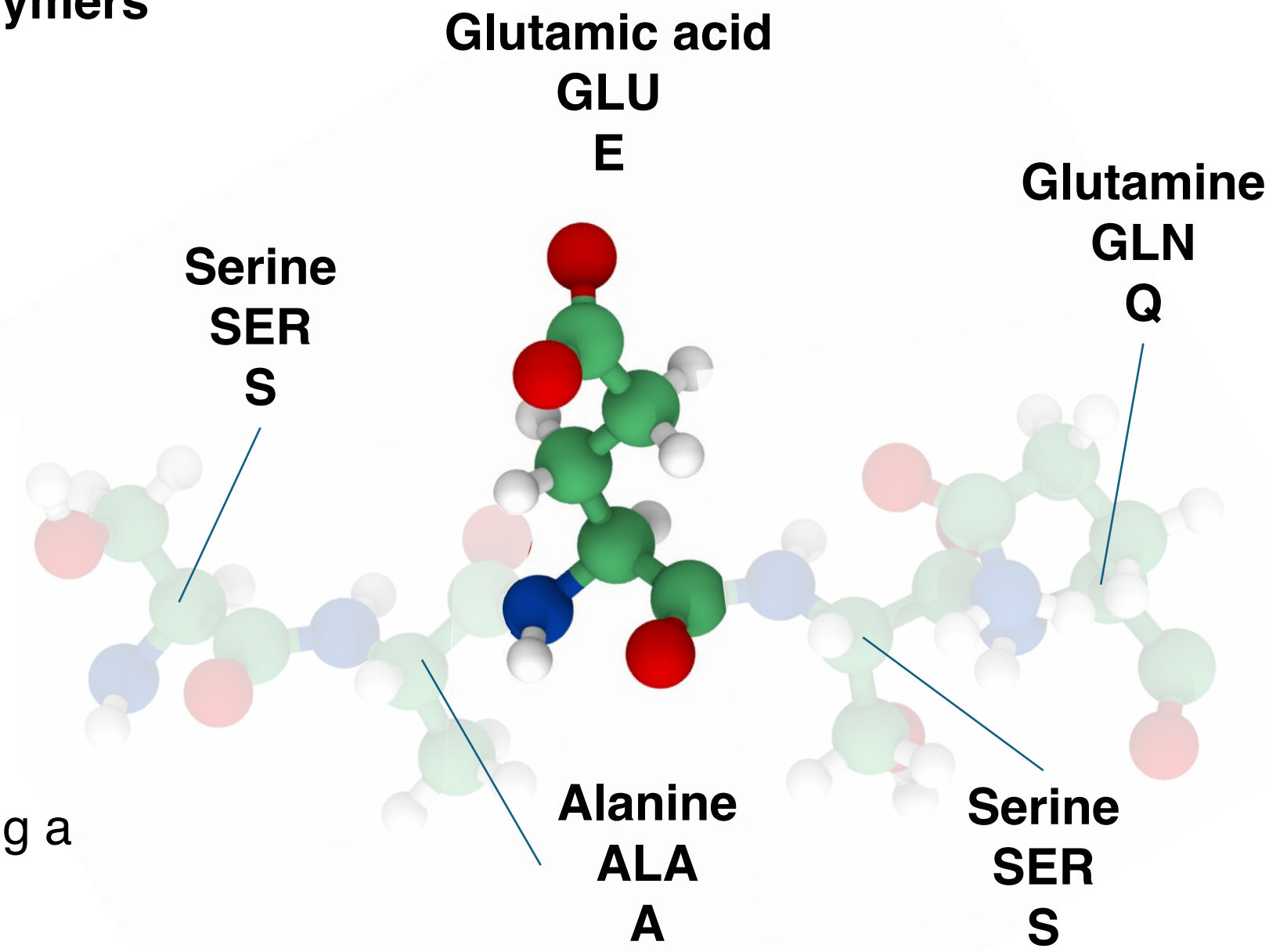


Proteins are amino acids polymers

Amino acids are composed of:
- **Backbone** (conserved)
- **Side chain** (variable)

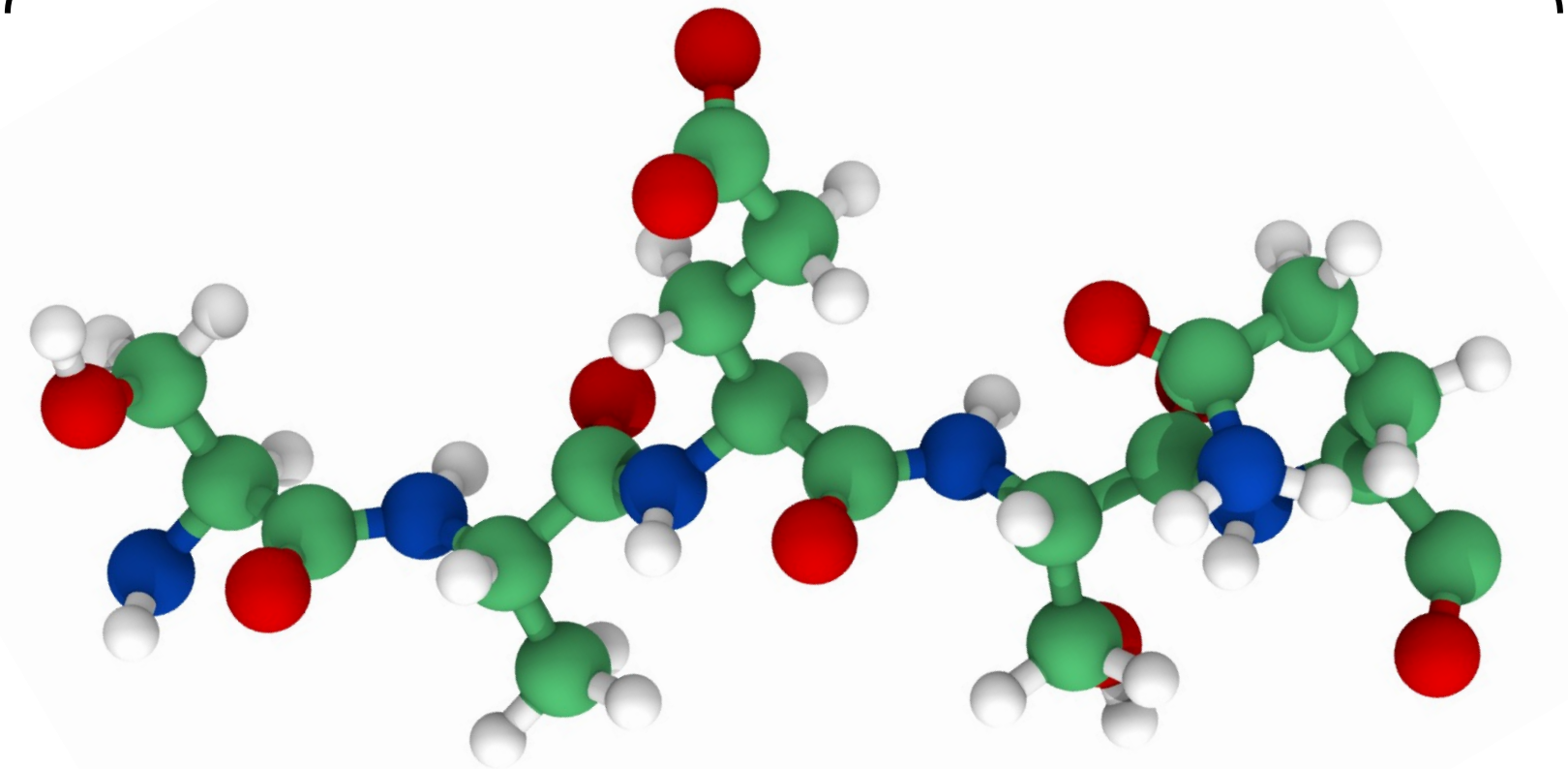


Amino acids polymerize forming a *peptidic bond* (condensation)



Protein Primary **Structure: Sequence**

SAESQ

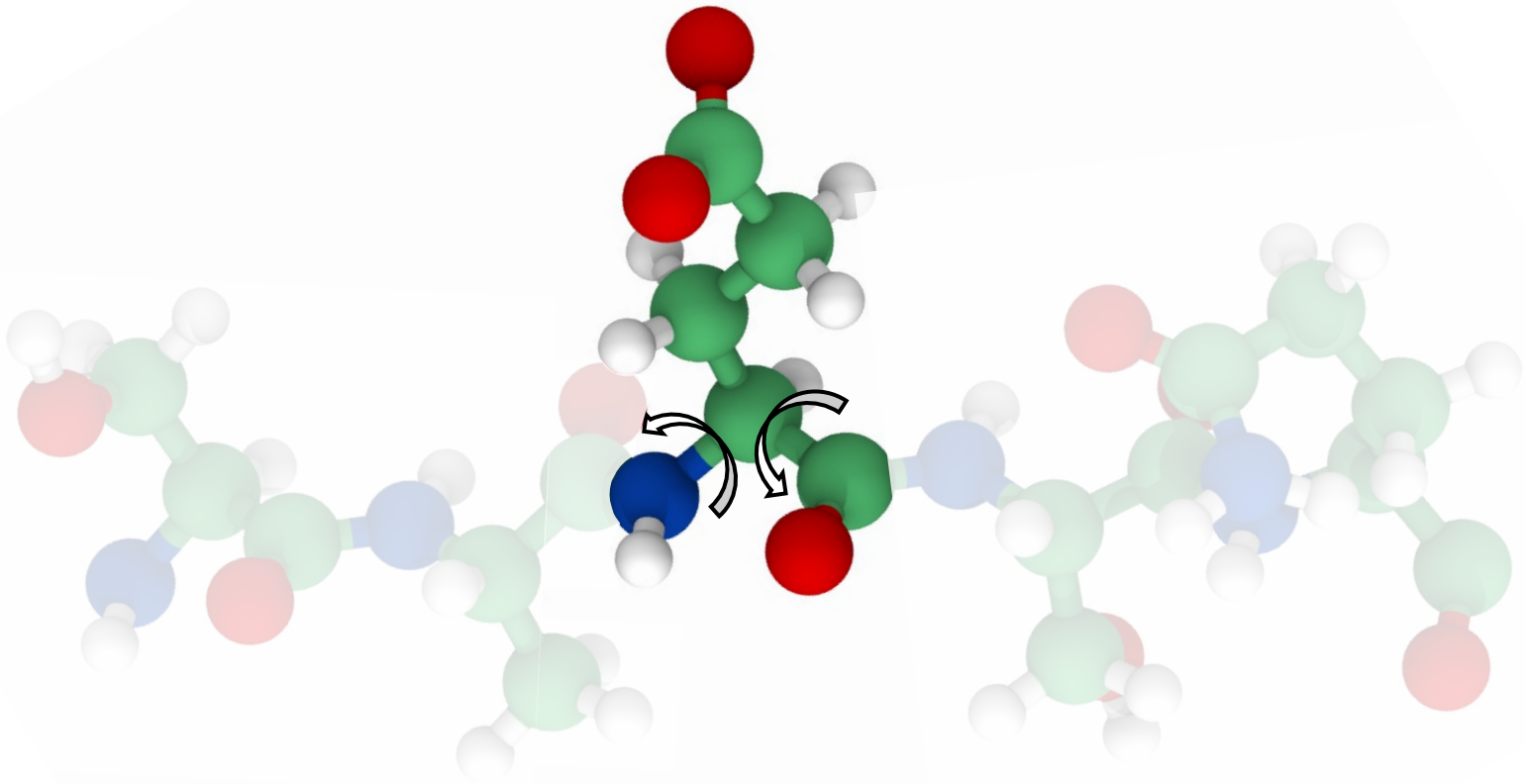


Proteins with similar sequence, likely:

- have similar functions in an organism
- are evolutionarily related

Protein Secondary Structu

The amino acid chain path is determined by backbone torsional angles



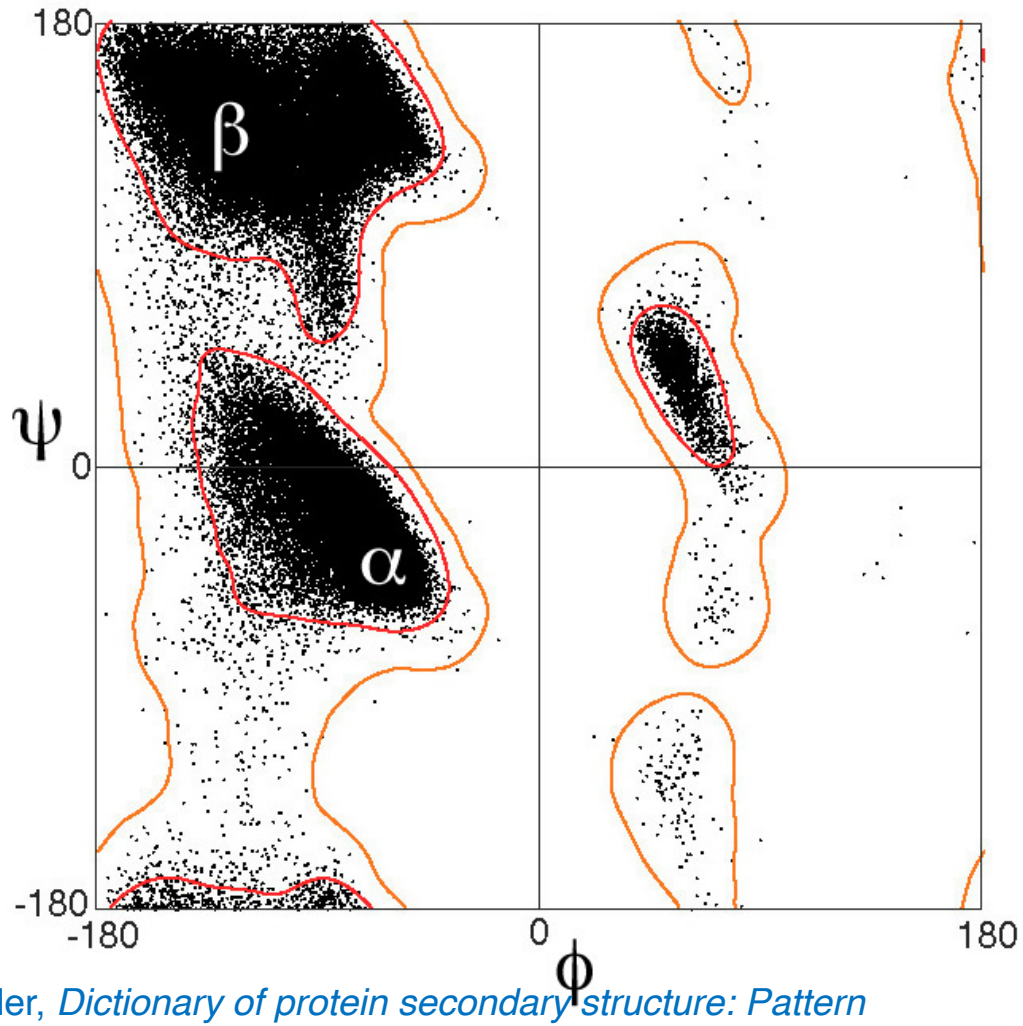
Protein Secondary Structure

The amino acid chain path is determined by backbone torsional angles

Ramachandran plot: scatter plot of amino acids backbone torsional angles

Dictionary of Secondary Structure in Proteins (DSSP) classification defines 7 secondary structure elements (regions in the plot):

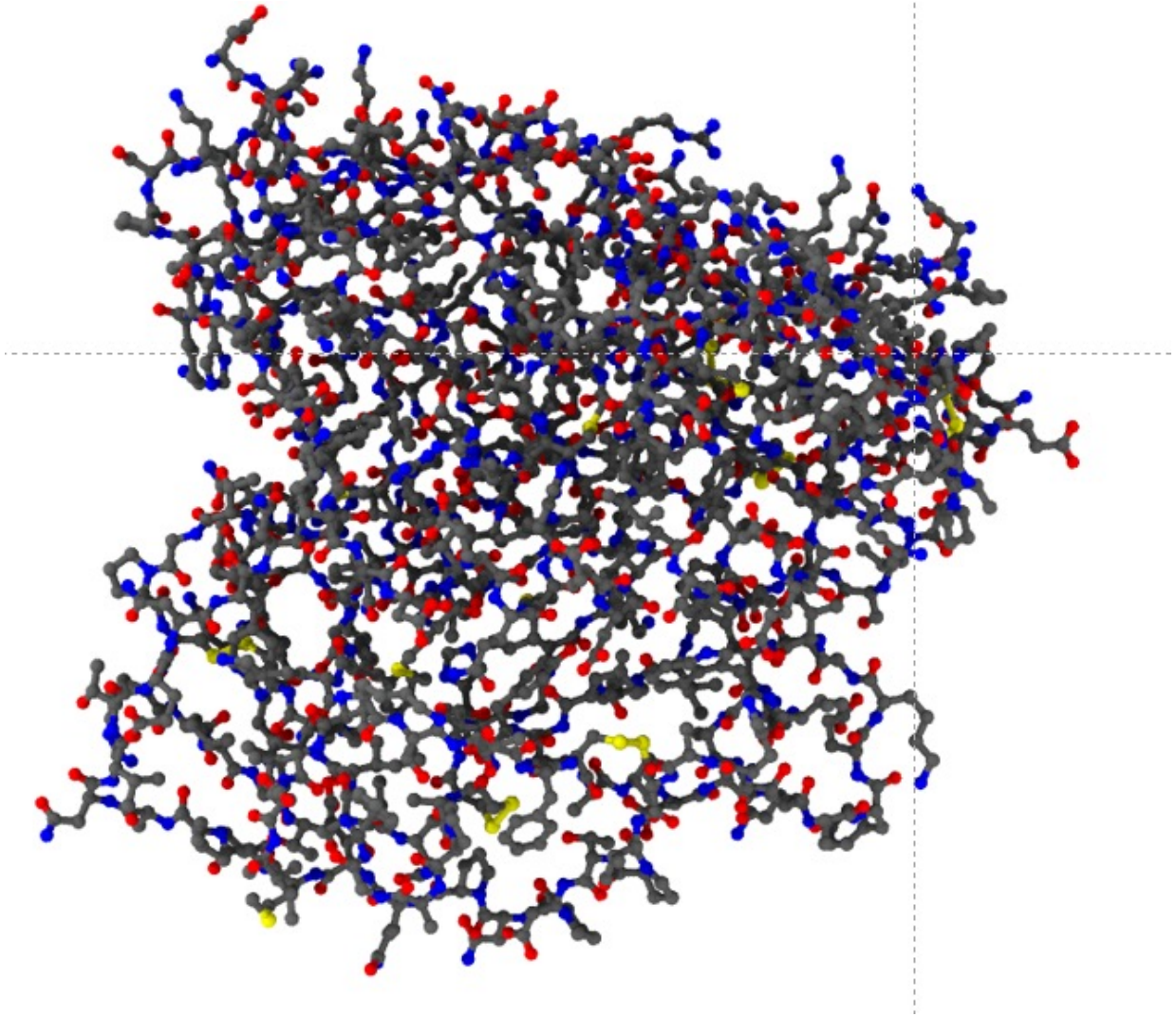
- H = α-helix
- G = 3_{10} helix
- I = π-helix
- B = residue in isolated β-bridge
- E = extended strand
- T = hydrogen bonded turn
- S = bend



W. Kabsch and C. Sander, *Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features,* ***Biopolymers***, 1983

Protein tertiary structure: folding

Protein («polypeptide»): 10 to >1000 amino acids



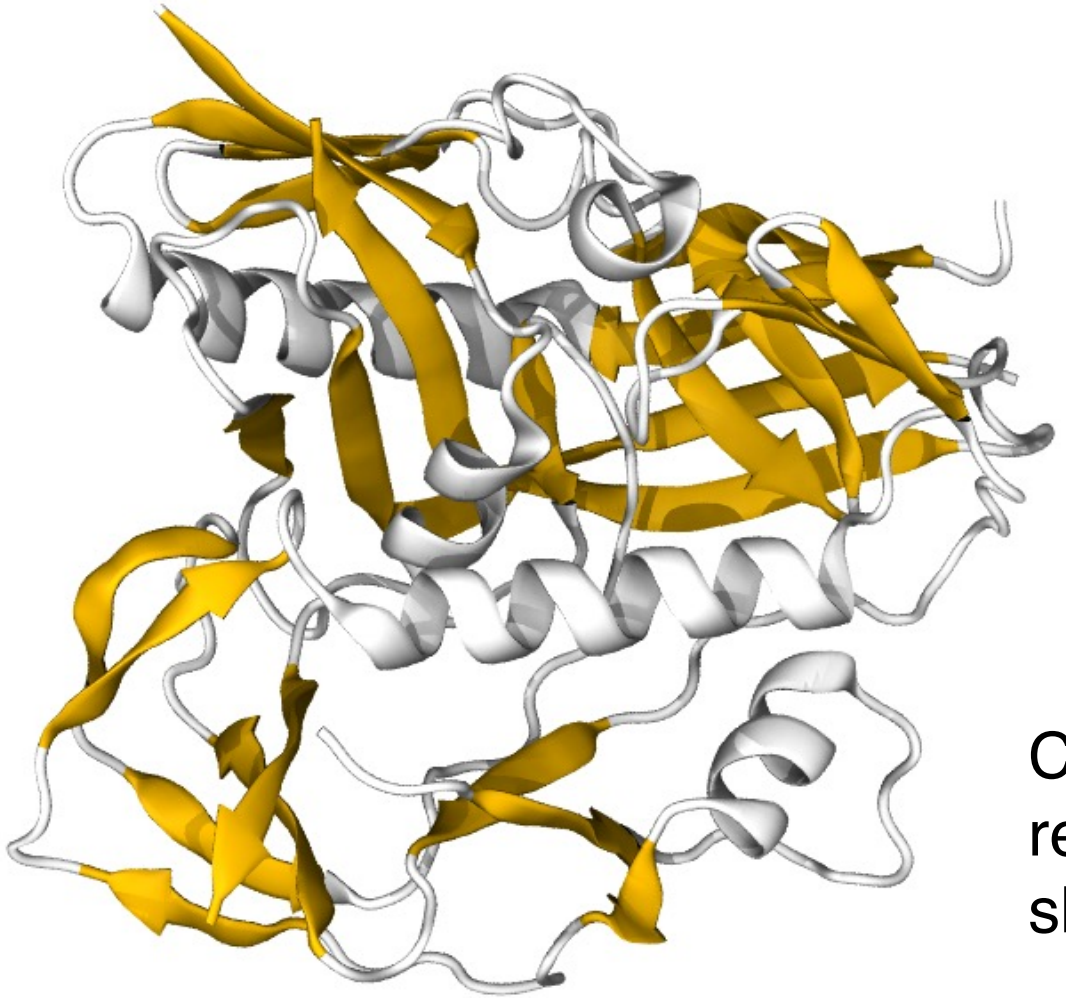
Protein tertiary structure: folding

Protein («polypeptide»): 10 to >1000 amino acids

Cartoon representation: helices

Protein tertiary structure: folding

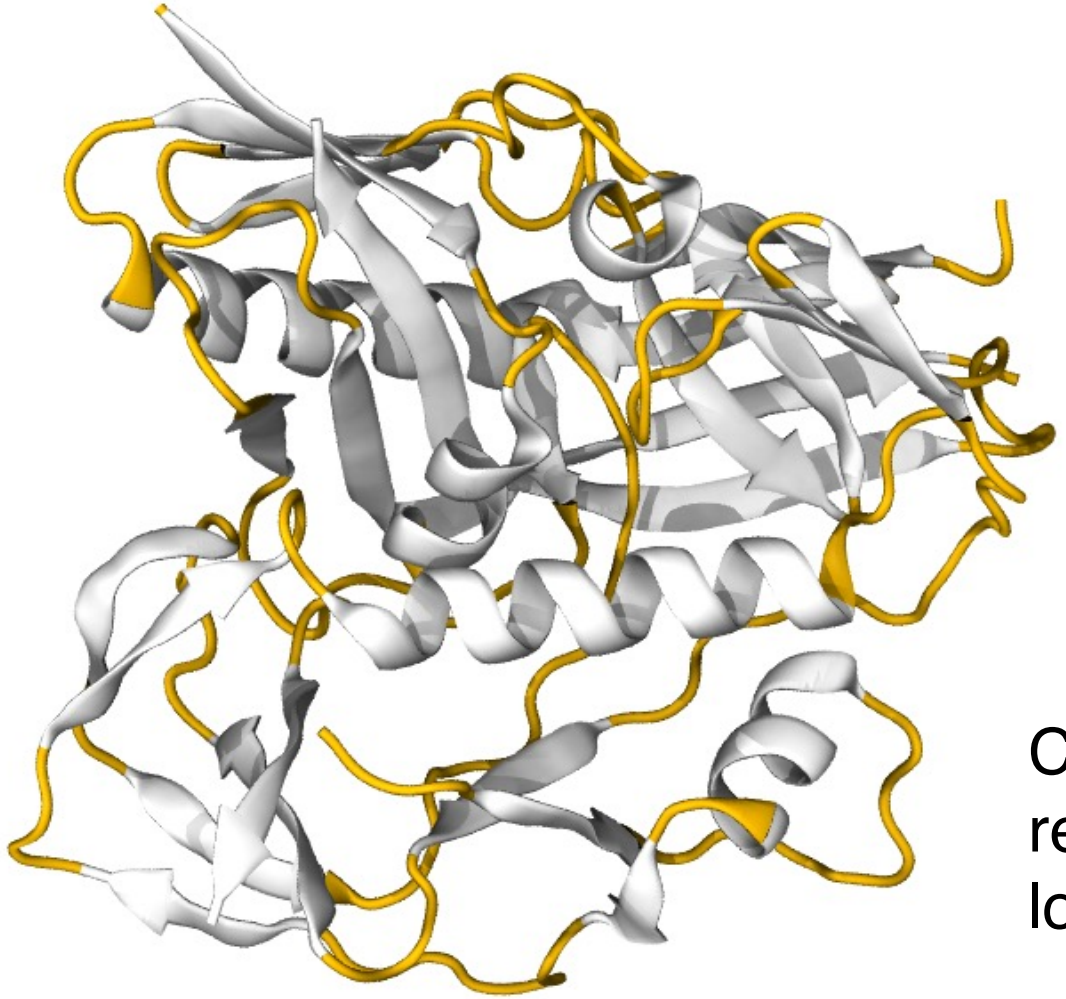
Protein («polypeptide»): 10 to >1000 amino acids



Cartoon representation: sheets

Protein tertiary structure: folding

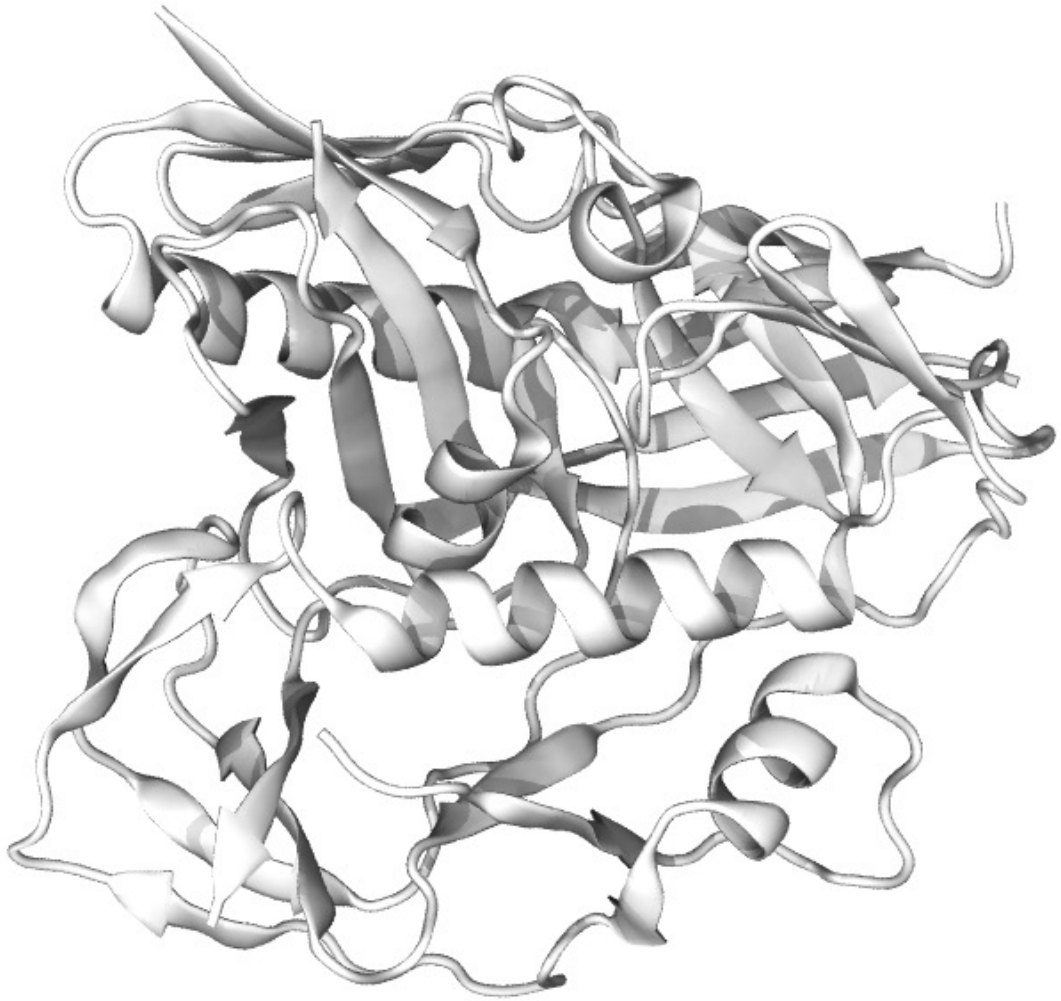
Protein («polypeptide»): 10 to >1000 amino acids



Cartoon representation: loops

Protein tertiary structure: folding

Protein («polypeptide»): 10 to >1000 amino acids



Anfinsen's dogma

The three-dimensional structure of a protein in its native environment is solely determined by its amino acid sequence.

Christian Anfinsen. *Principles that govern the folding of protein chains,* ***Science***, 1973

Hydrogen bonds

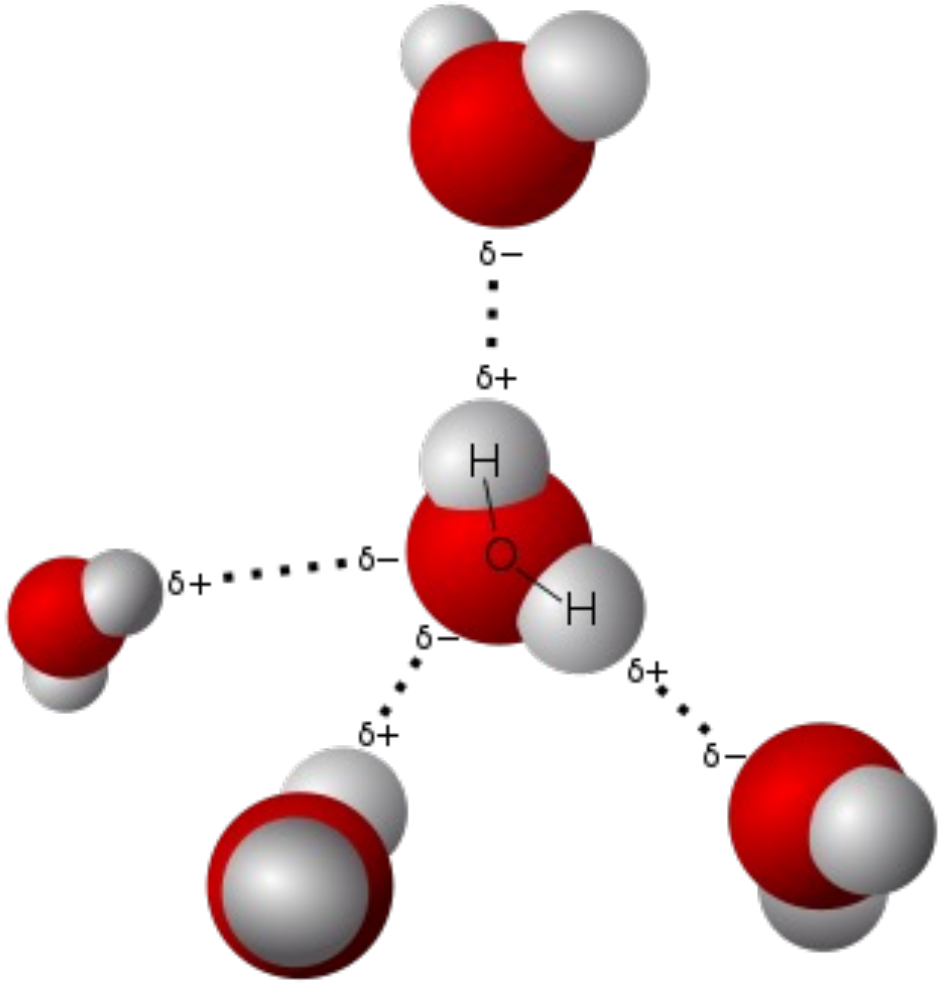
Electrostatic interaction. Structure:

- donor-acceptor distance typically 1.6-2 Å
- donor-acceptor-hydrogen angle must be small

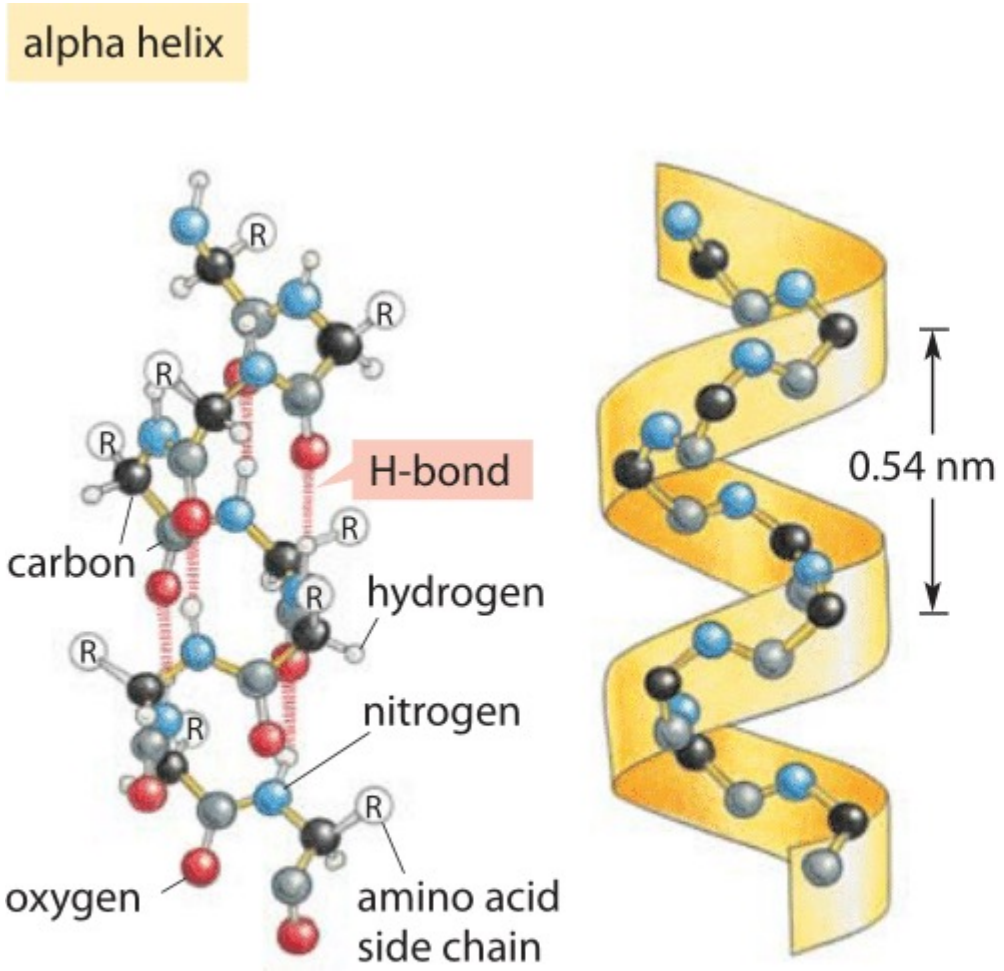
Hydrogen bond energy in biomolecular systems typically 5-25 kJ/mol, e.g.:

- O–H···:O , 21 kJ/mol (5.0 kcal/mol)
- N–H···:O , 8 kJ/mol (1.9 kcal/mol)

amino acids' backbone and polar side chains can be donor/acceptor



Hydrogen bonding on protein backbone

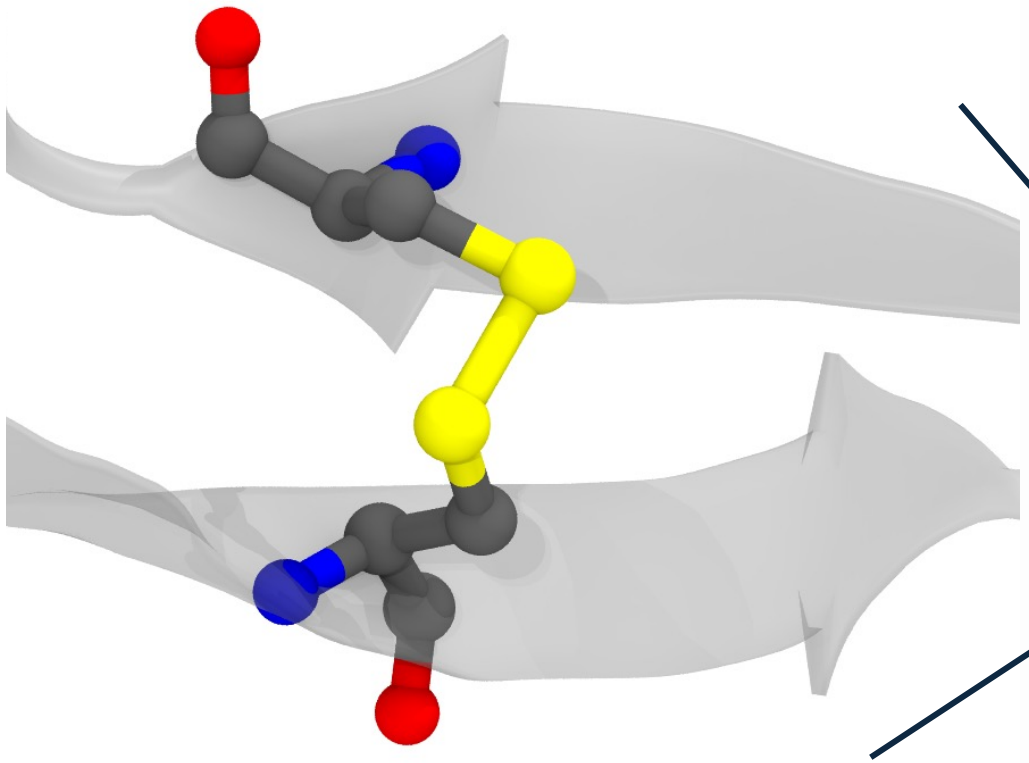


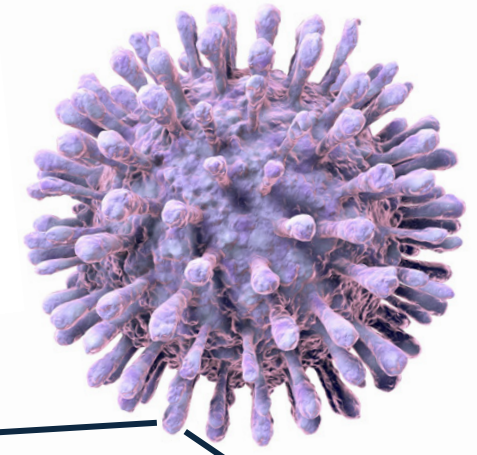
book.bionumbers.org

Disulfide bridges

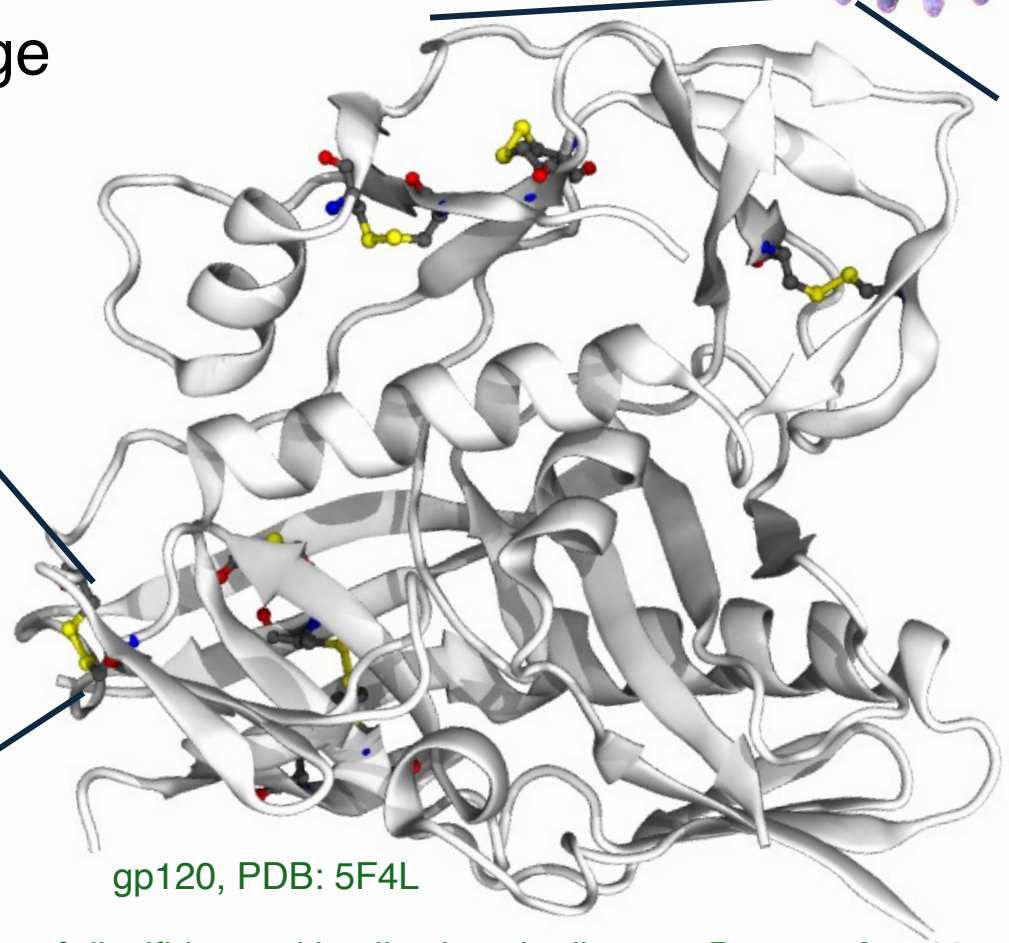
Under *oxidising* conditions, cysteine side chains can form a *covalent* bond.

- Protein more resistent to denaturation
- Changes in conditions ➡ structural change





HIV-1 virus



gp120, PDB: 5F4L

T.J.Bechtel and E.Weerapana, *From structure to redox: the diverse functional roles of disulfides and implications in disease*, ***Proteomics***, 2017

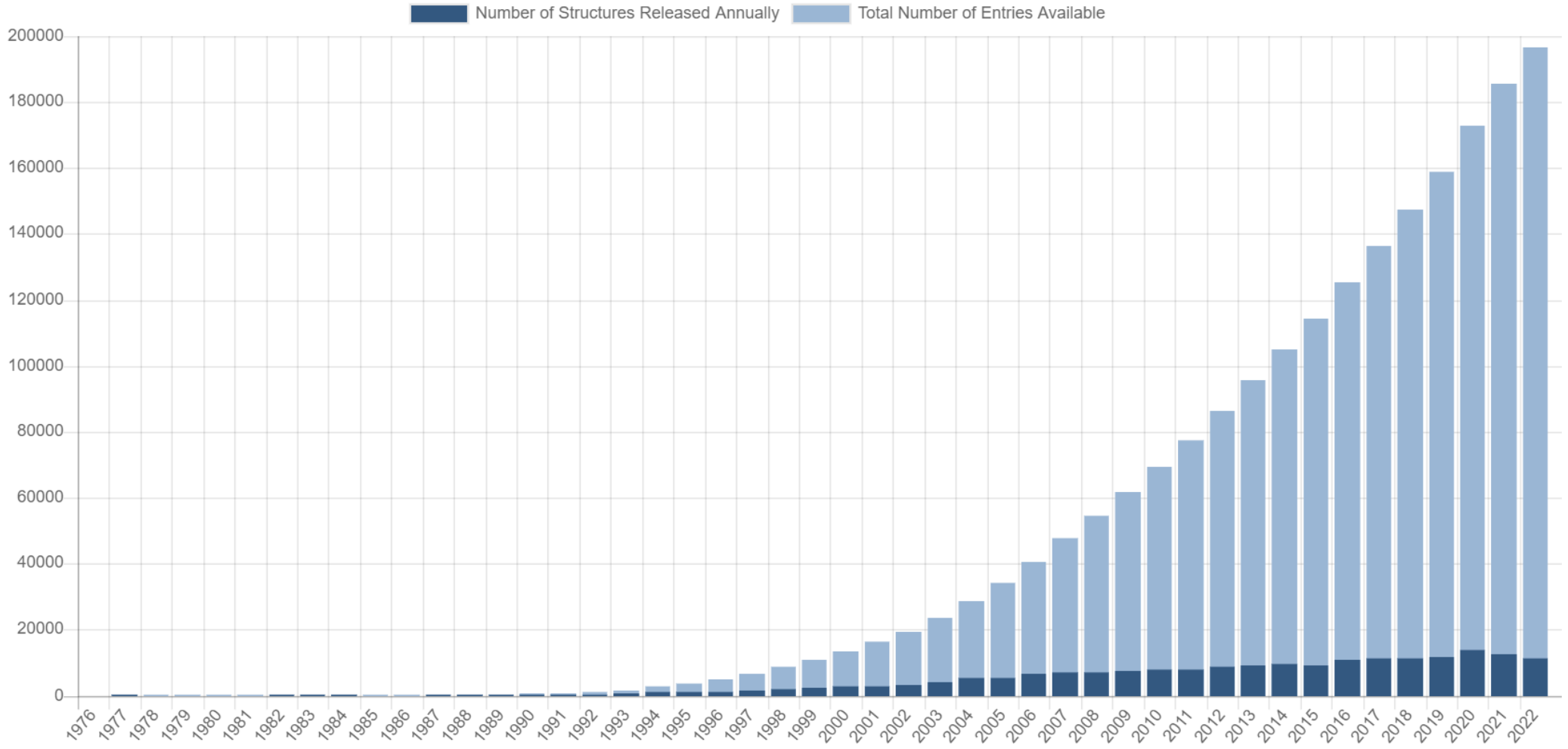
Proteins, illness, and drug design

- **Proteins in diseases** (e.g., COVID-19, salmonella, flu, ...)
  - pathogen's own metabolism/structure
  - pathogen's weapon

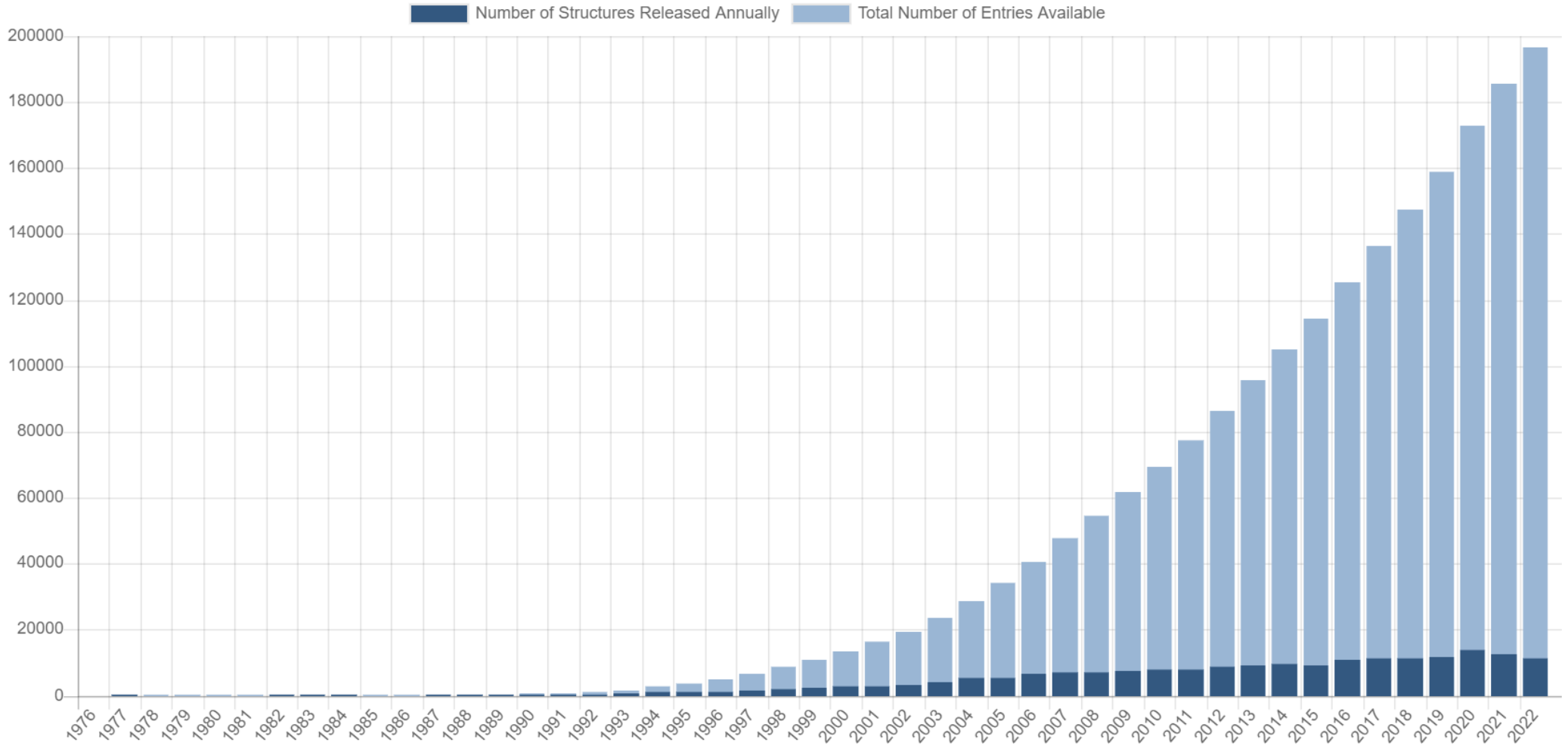
- **Proteins in disorders** (e.g., Cancer, Alzheimer's, ...)
  - own protein misfolds
  - own protein folds, but has different dynamics

- **Drug**
  - Small molecule designed to specifically bind to a protein, so as to affect its function

The Protein Data Bank (PDB)

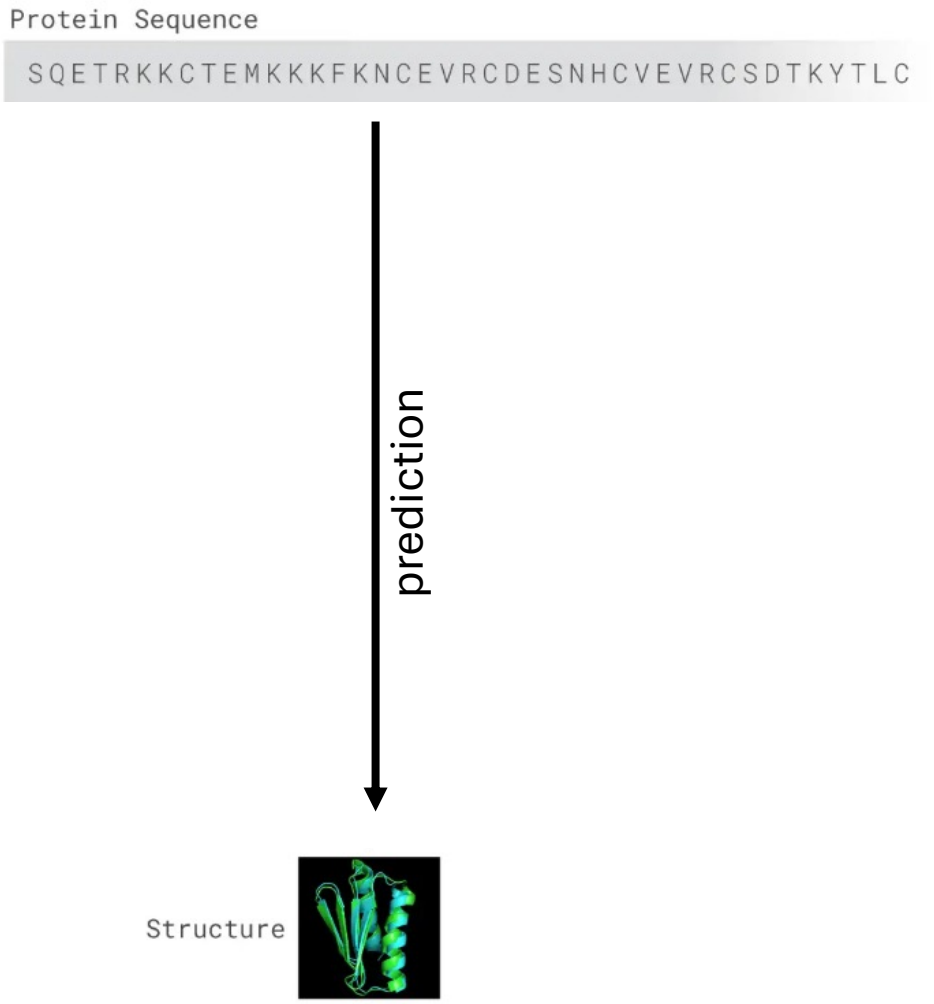


The Protein Data Bank (PDB)



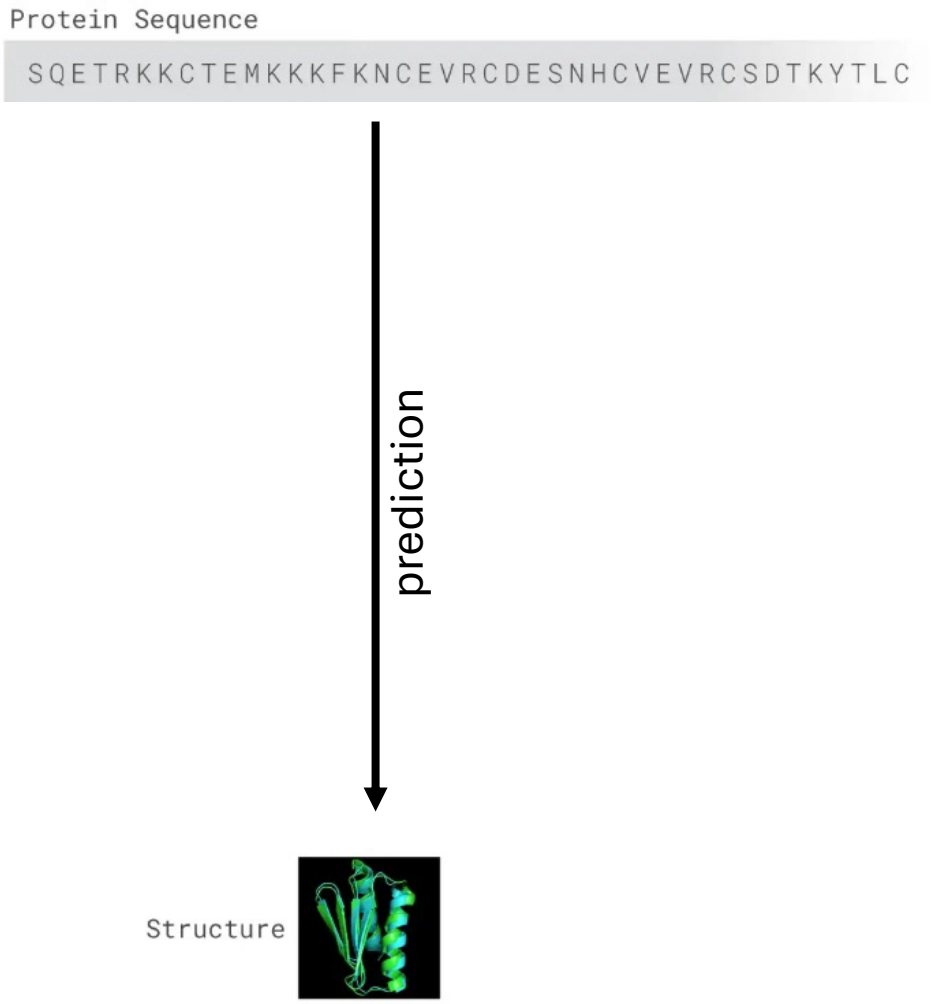
known protein **structures**: ~80’000 (*PDB, 90% identity*)
known protein **sequences**: ~190’000’000 (*UNIPROT*)

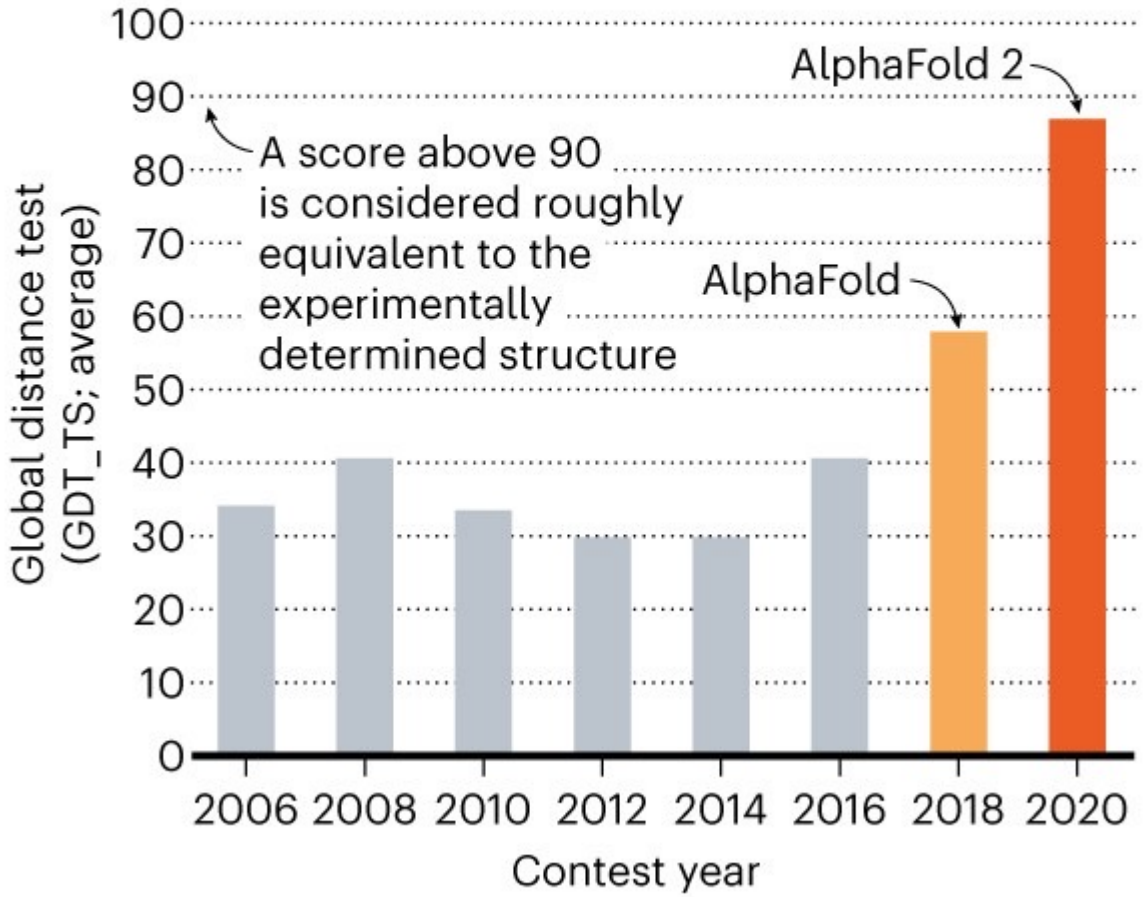
Protein fold prediction



CASP, since 1994 biennial competition on protein fold prediction: predictioncenter.org

Protein fold prediction



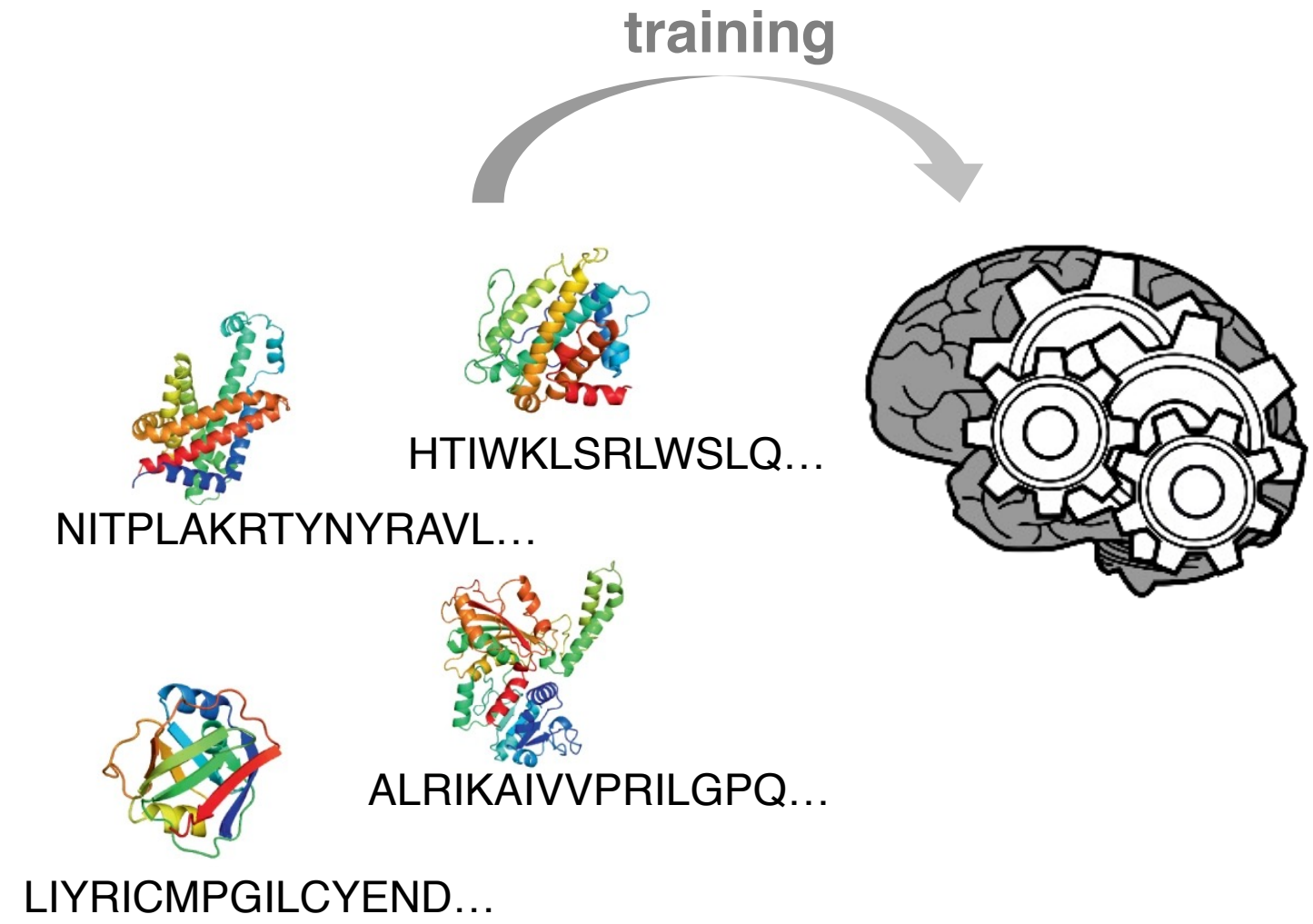


A.W. Senior et al., *Improved protein structure prediction using potentials from deep learning*, ***Nature***, 2020

J. Jumper et al., *Highly accurate structure prediction with AlphaFold*, ***Nature***, 2021

E. Callaway, *'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures,* ***Nature***, 2021

Protein fold prediction: AlphaFold



Protein fold prediction: AlphaFold

prediction

MAPVYEGMAS…

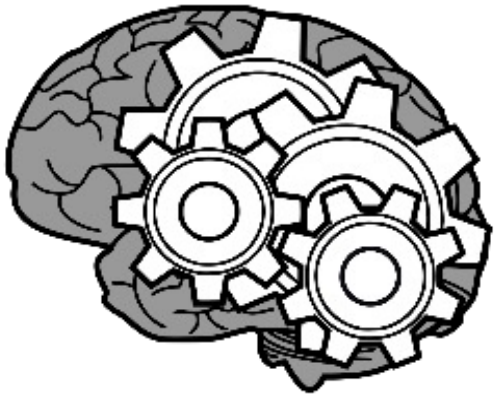
HVQVFSPHTL…

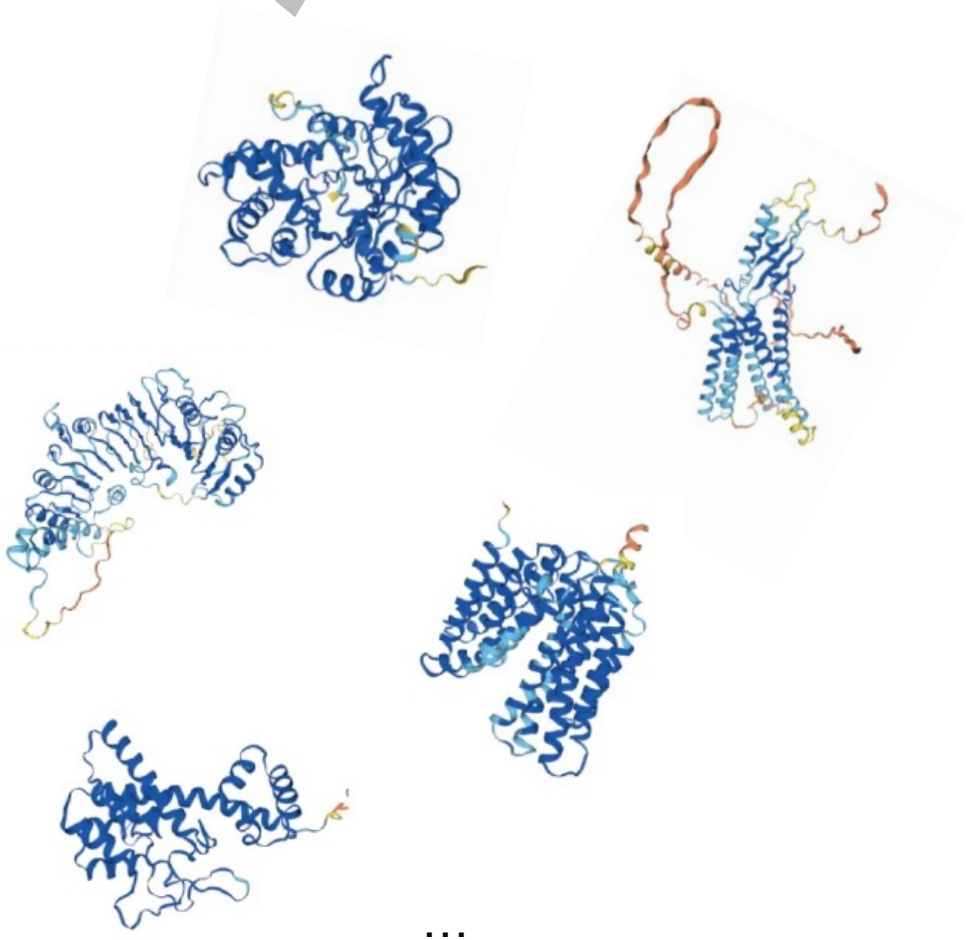
QSSAFCSVKK…

LKIEPSSNWD…

MTGYGSHSKV…

…

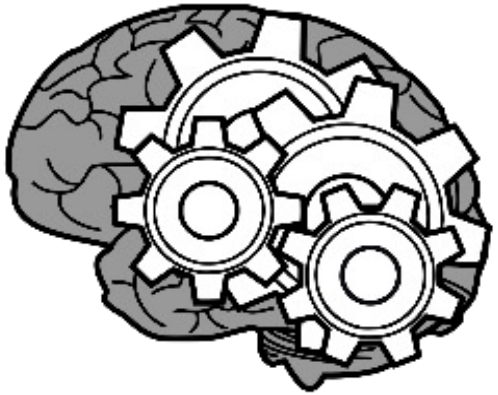




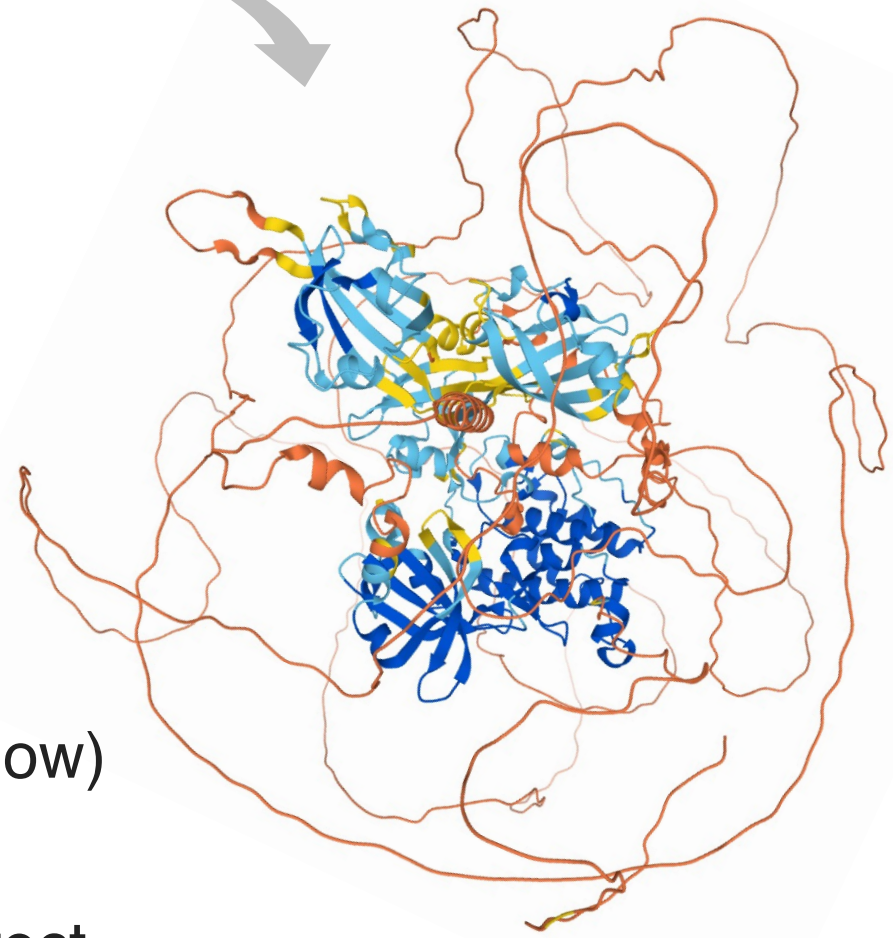
…

Protein fold prediction: AlphaFold

YSQSKNIPLS…



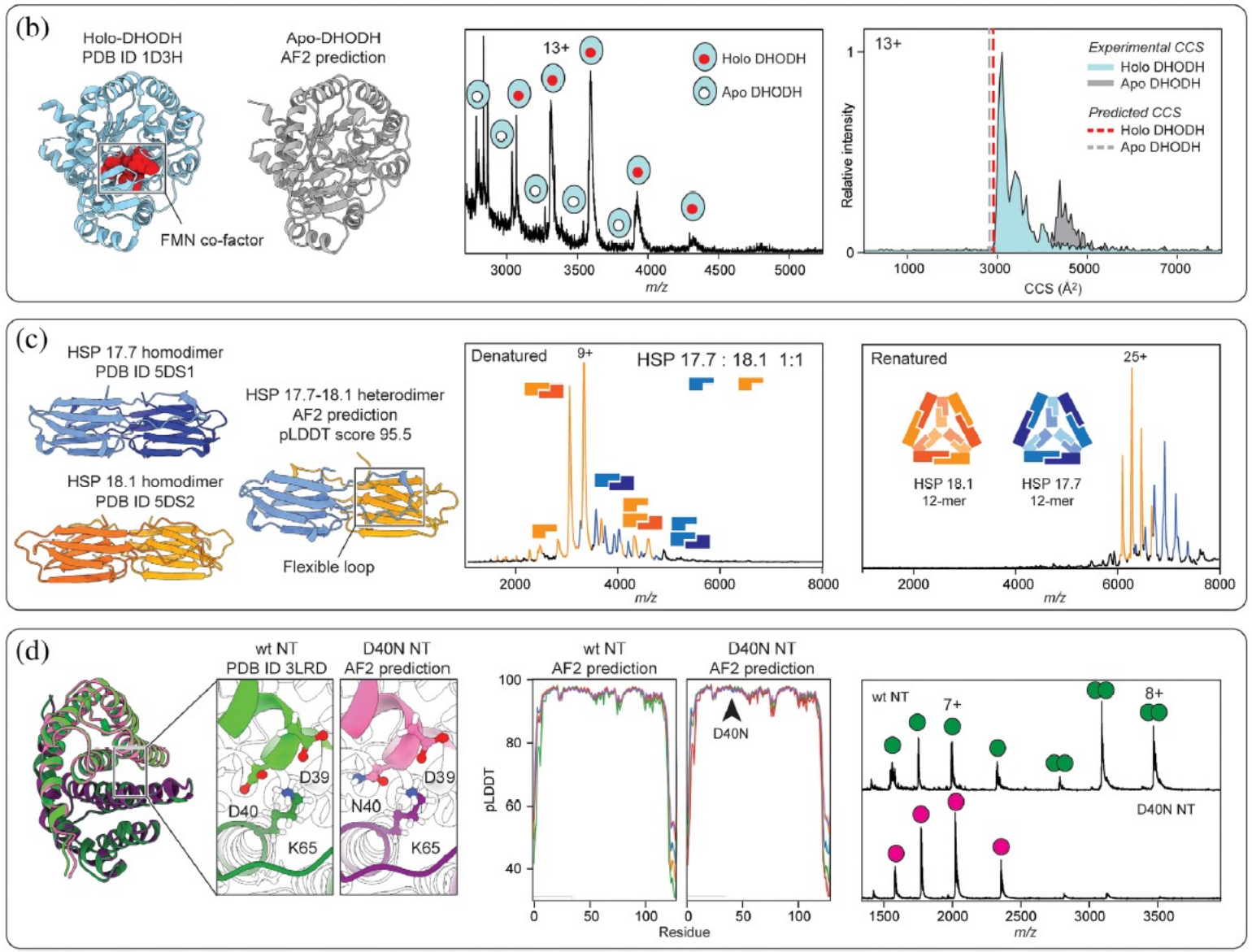
prediction



PLDDT score: confidence score (**100** = high, **<50** = low)

Warning: even high-confidence models can be incorrect.

Protein fold prediction: warning!



Apo protein predicted folded like holo state, but it should be unfolded

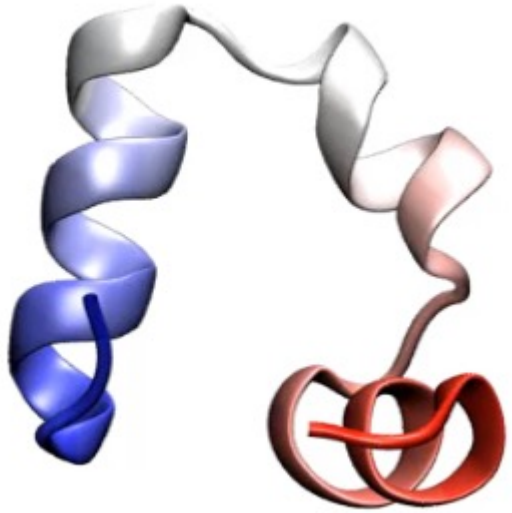
High-confidence hetero multimer predicted, but proteins do not co-assemble

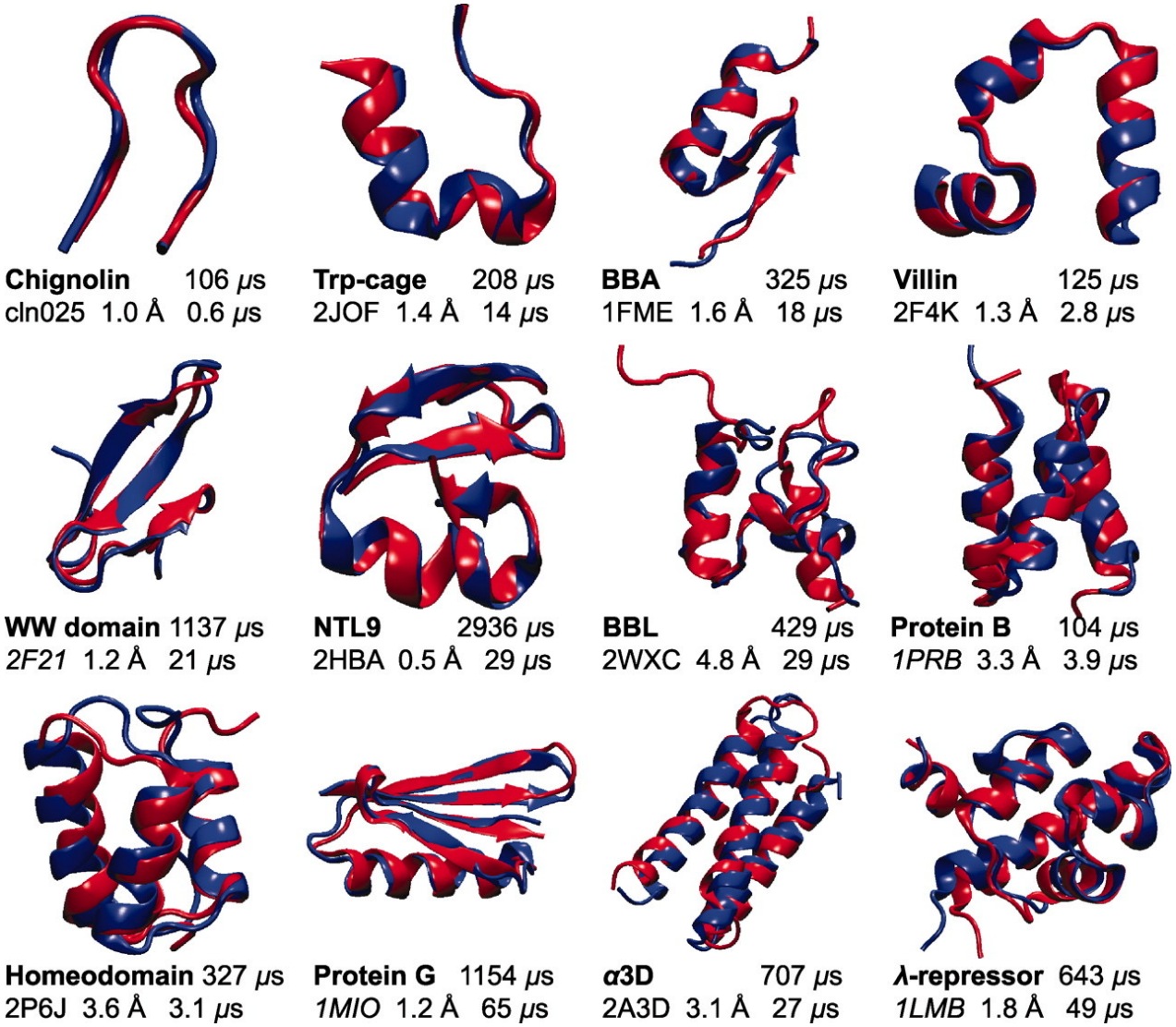
High-confidence homodimer of mutated protein predicted, but mutation abolishes complex formation

T.M.Allison et al., ***Protein Science***, 2022

Watching proteins fold: simulation

- Following experimentally the *folding pathway* of a protein is difficult
- Folding of small fast-folding (<100 μs) proteins can be studied via simulation



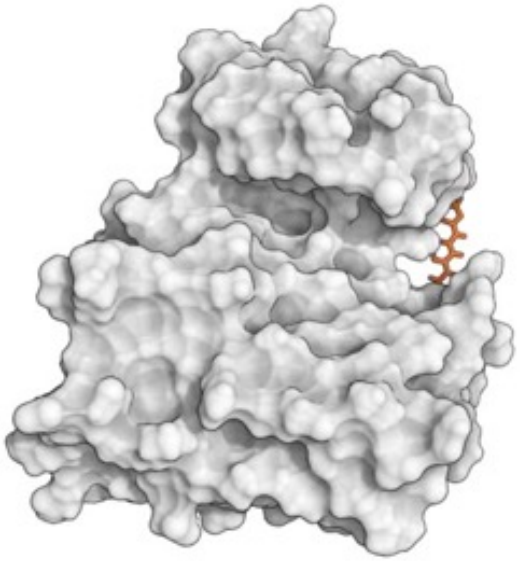


P. Freddolino and K. Schulten. *Common structural transitions in explicit-solvent simulations of villin headpiece folding,* ***Biophysical J.***, 2009

K. Lindorff-Larsen et al. *How fast-folding proteins fold,* ***Science***, 2011

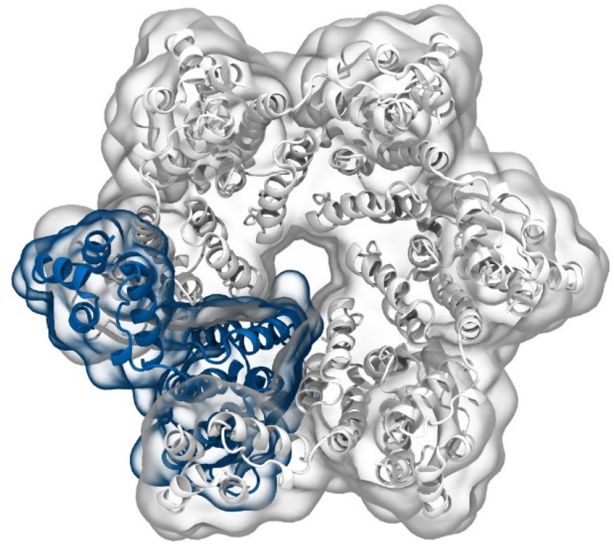
Next: how to prepare a protein structure so that it is ready for molecular modelling?

Simulation of Biomolecules



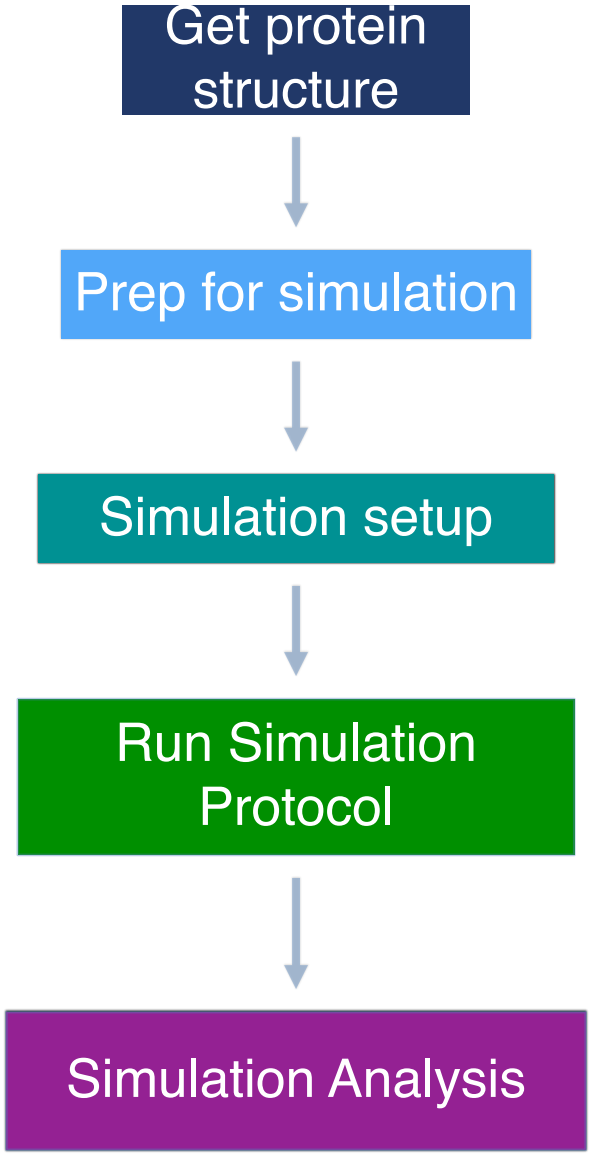
Protein preparation

Delivered by **Tim Spankie**
University of Edinburgh
January 2025



Based on content from a ten-hour series https://github.com/CCPBioSim/BioSim-analysis-workshop
by Matteo Degiacomi (matteo.degiacomi@ed.ac.uk) and Antonia Mey (antonia.mey@ed.ac.uk)

A typical workflow for molecular dynamics



pdb api
pdb2gmx
SCHRÖDINGER Maestro
RDkit
SeeSAR The drug design dashboard
Amber tools
OpenMM tools
[…]

GUI, TCL, bash, Python, Perl … get creative

Gromacs
OpenMM
Amber
NAMD
DLPoly
SOMD
Charmm
BioSimSpace
[…]

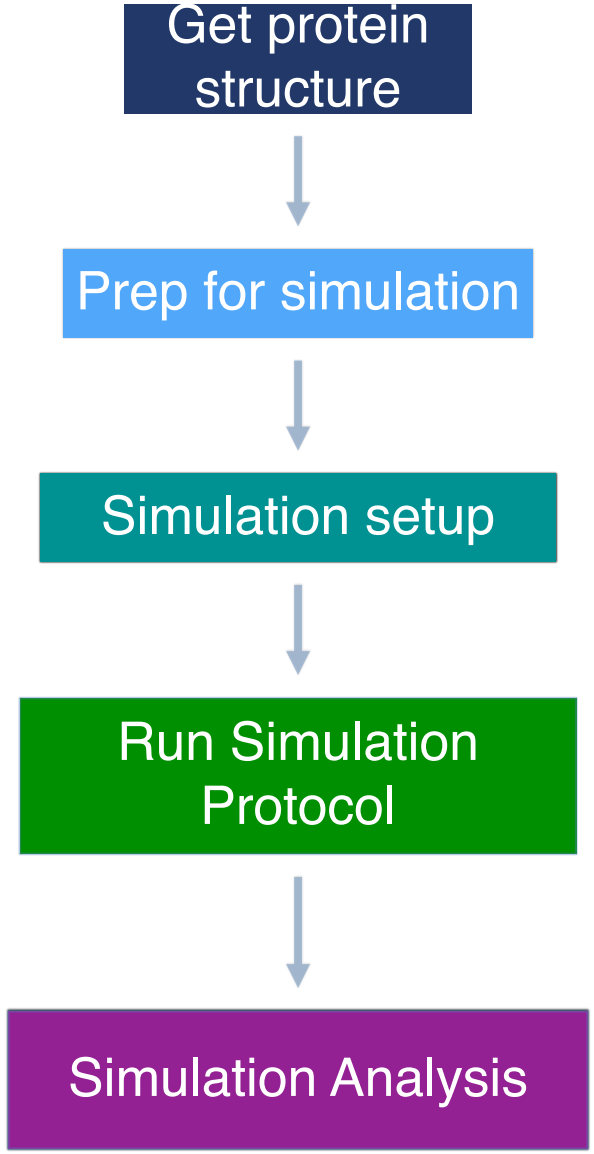
mostly C++ command line

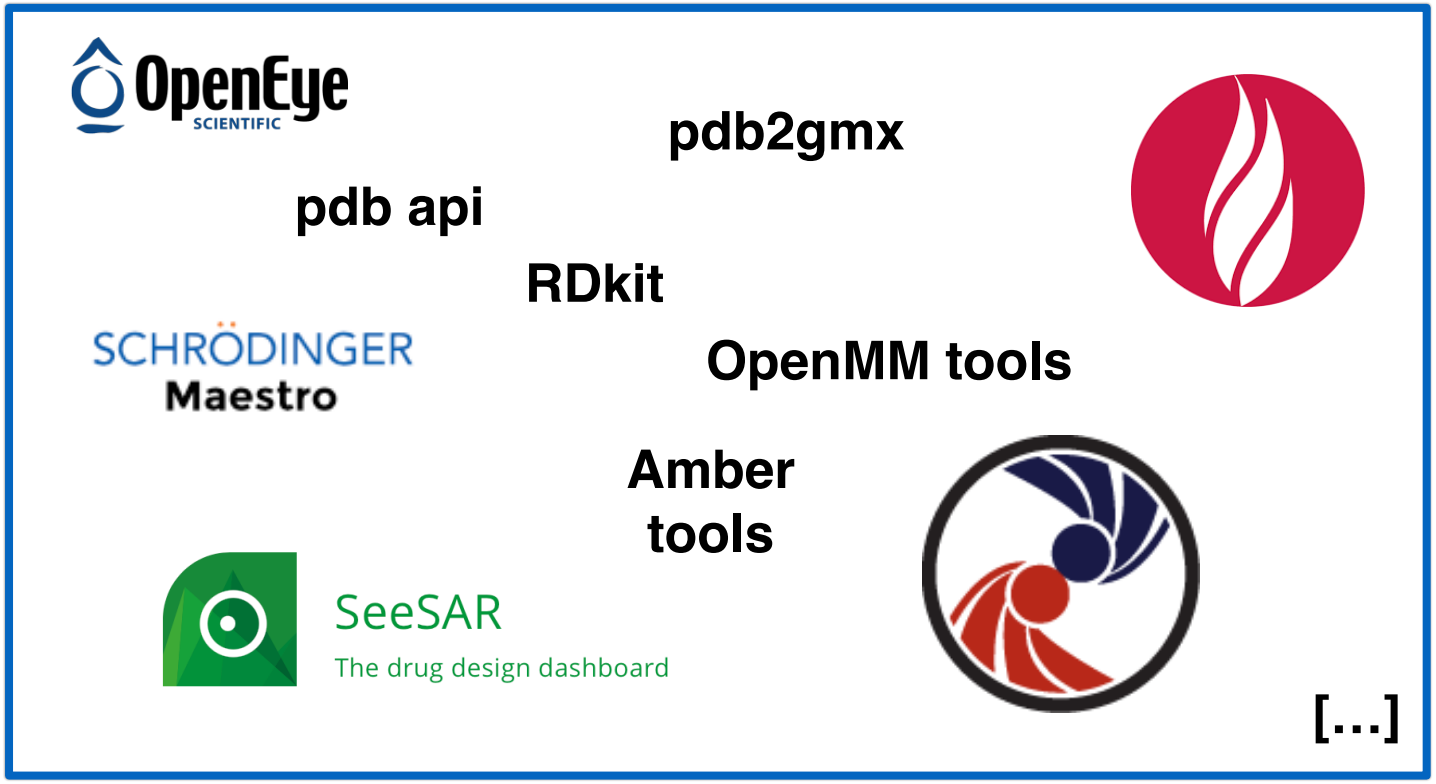
Python API

PyEMMA
MDAnalysis
md-analysis-tools
MDTraj
VMD
[…]

GUI, TCL, bash, Python, Perl, … get creative

Let’s get started with understanding protein structures





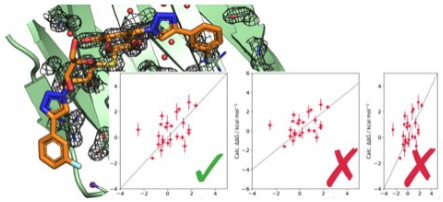
GUI, TCL, bash, python, Perl, … get creative

Crystal structures are models with potential errors

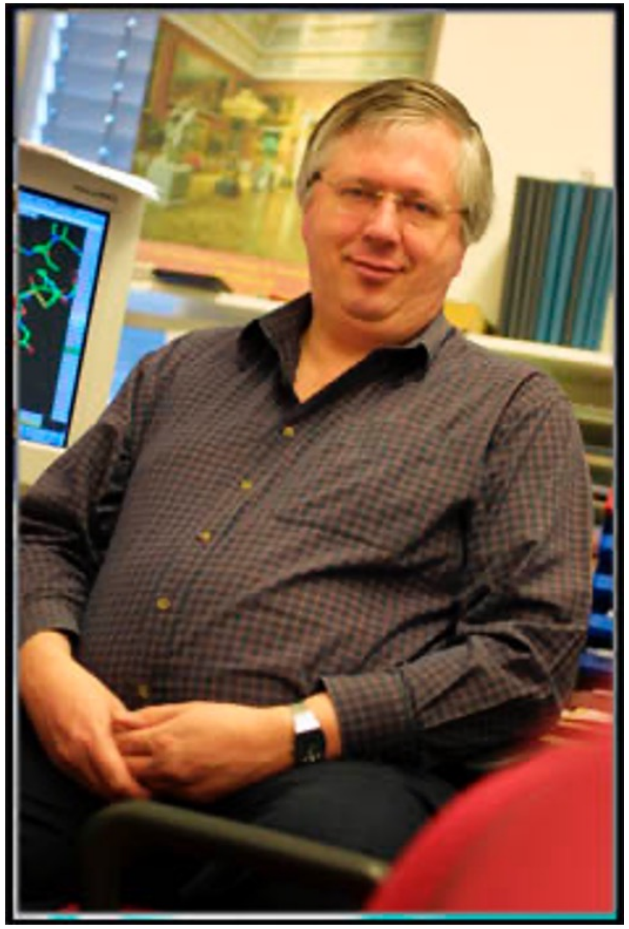
HOME / ARCHIVES / VOL. 4 NO. 1 (2022) / Articles

Best Practices for Constructing, Preparing, and Evaluating Protein-Ligand Binding Affinity Benchmarks [Article v1.0]

David F. Hahn
Computational Chemistry, Janssen Research & Development, Turnhoutseweg 30, Beerse B-2340, Belgium
https://orcid.org/0000-0003-2830-6880



Examples of errors/oddities in PDB structures:
https://swift.cmbi.umcn.nl/teach/pdbad/

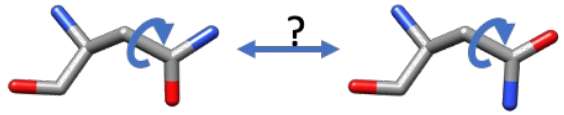


Gert Vriend (author of WHAT_CHECK)

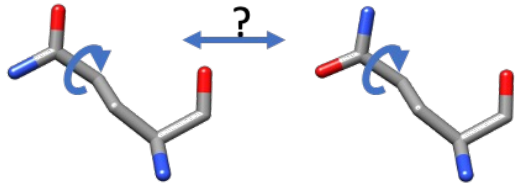
Alternative conformations of side chains - NGH flips

Typically, the crystallographers will have assigned the orientation based on potential H-bonding interactions, etc.

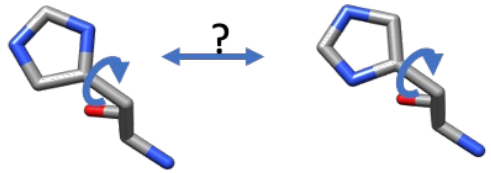
No standard protocol for this: always worth double-checking.



Asparagine (N)



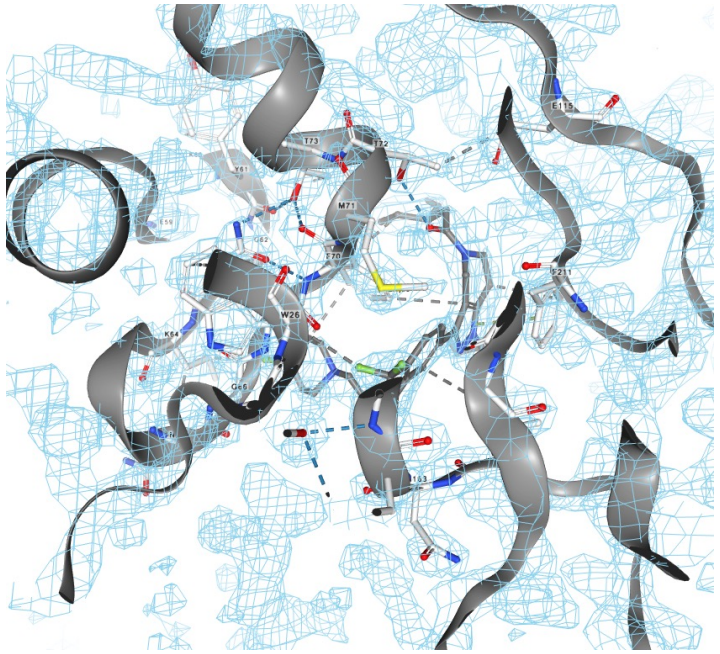
Glutamine (Q)



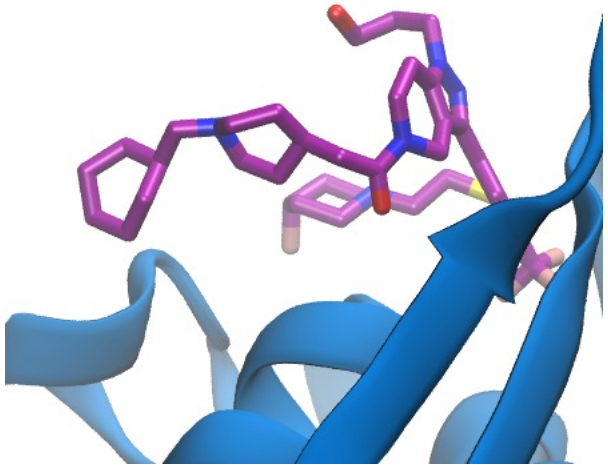
Histidine (H)

Ligand densities can also have creative input from the crystallographer

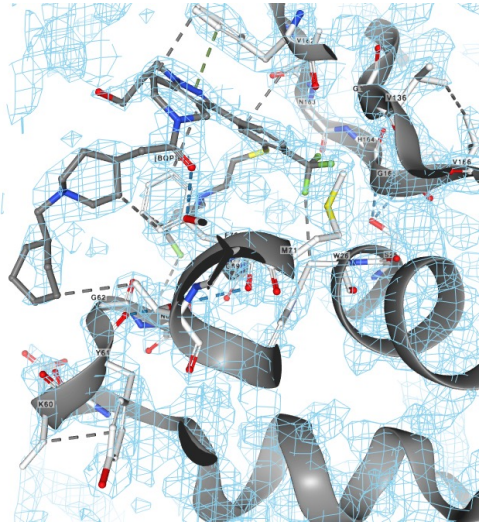
Cathepsin S



Cathepsin S with ligand

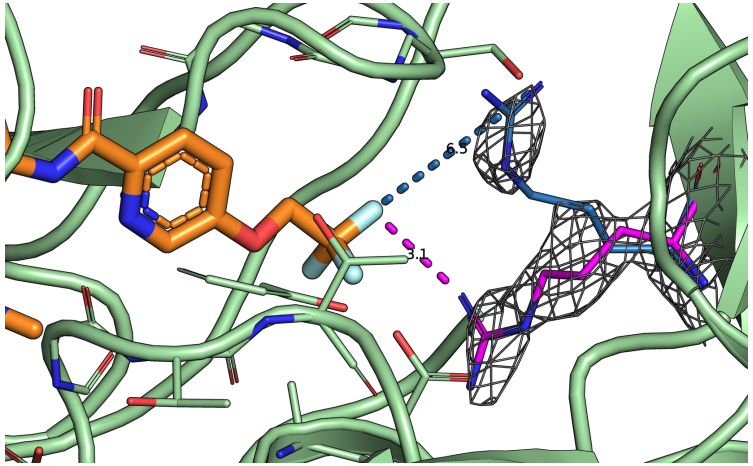


boat/chair?

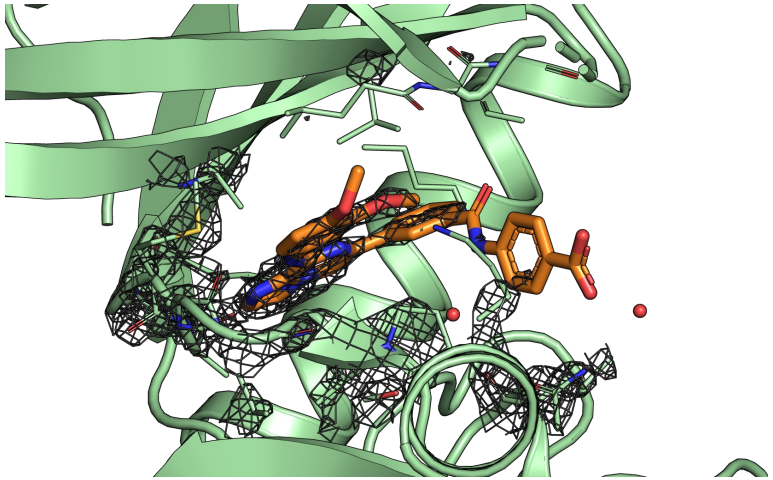


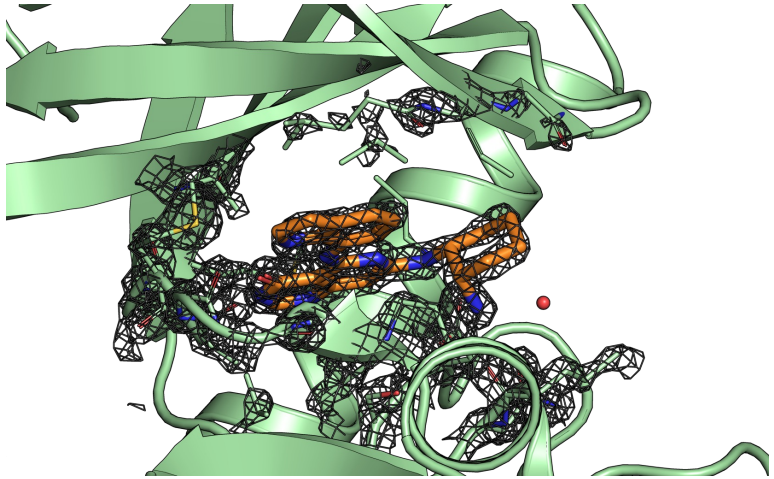
Picking the best crystal structure requires care

BACE (Hunt)



spleen tyrosine kinase



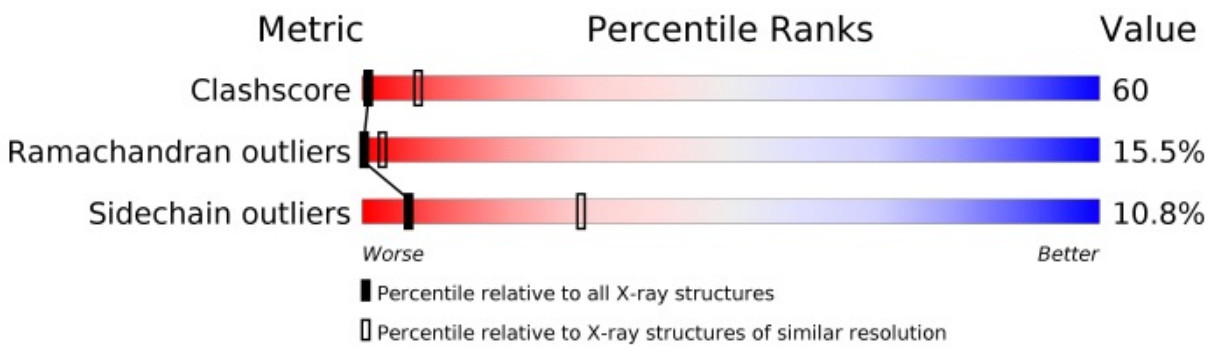


Which protein structure?

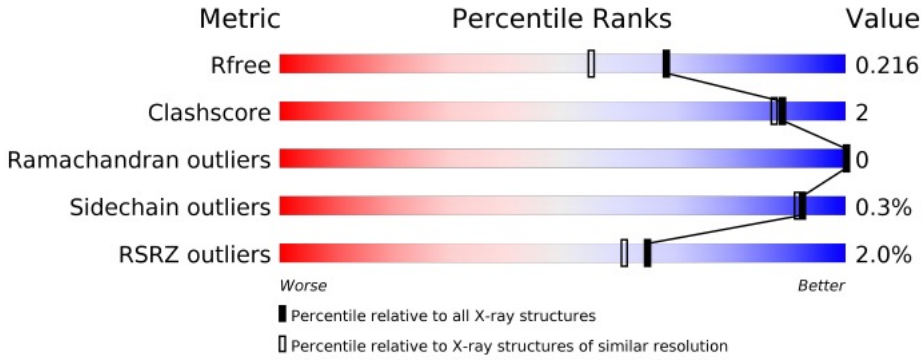
- Alternative side chain conformations need to be assessed carefully
- Active site residue densities are important for choosing the right crystal structure

The RCSB PDB report can help with choosing structures

Jnk1 - 2GMX

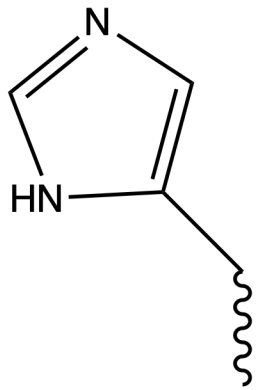


Jnk1 - 3ELJ

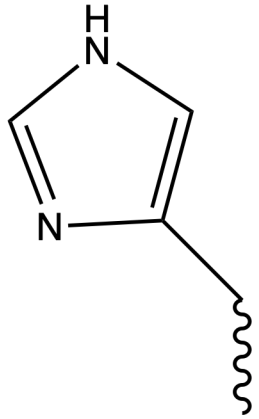


How to choose the right tautomer?

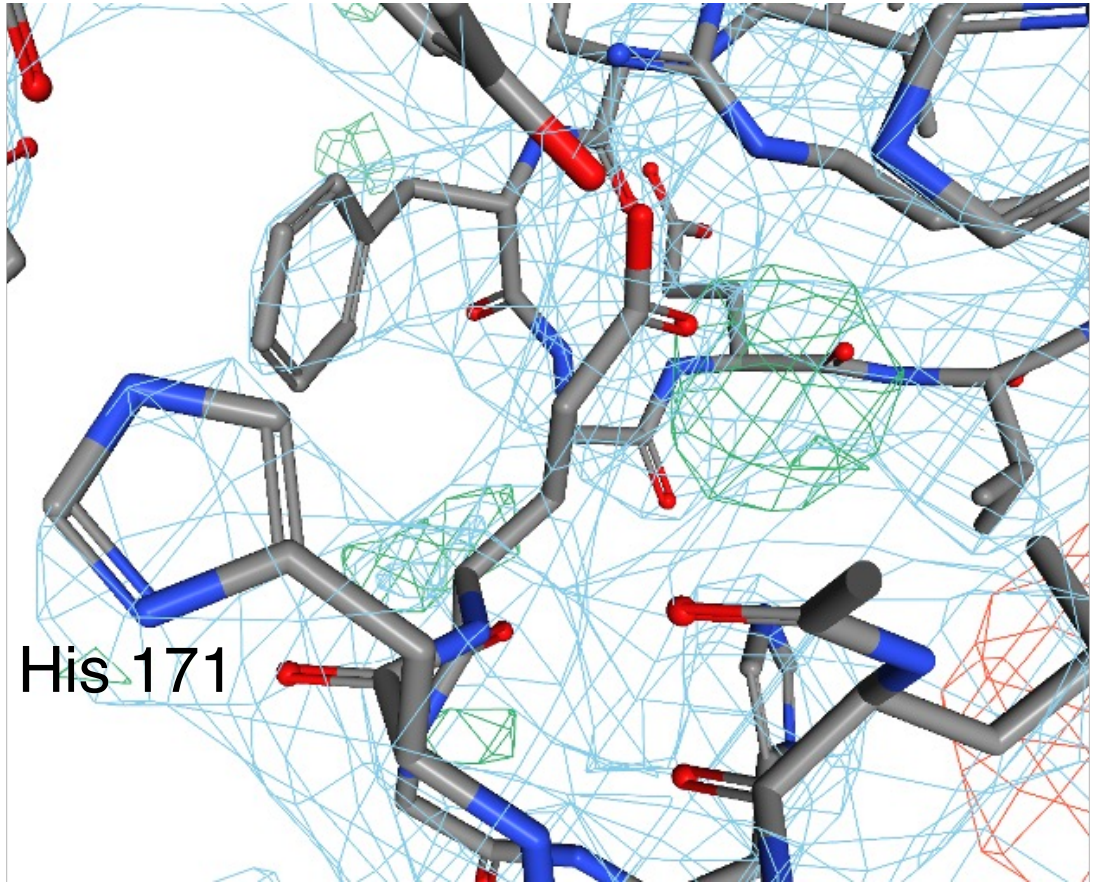
Which tautomer?



δ-tautomer

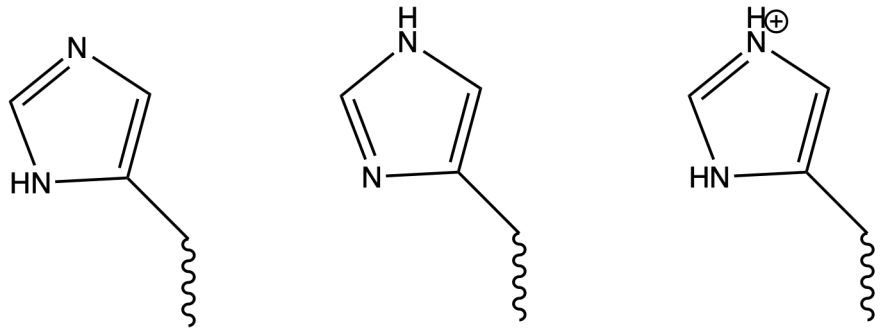


ε-tautomer

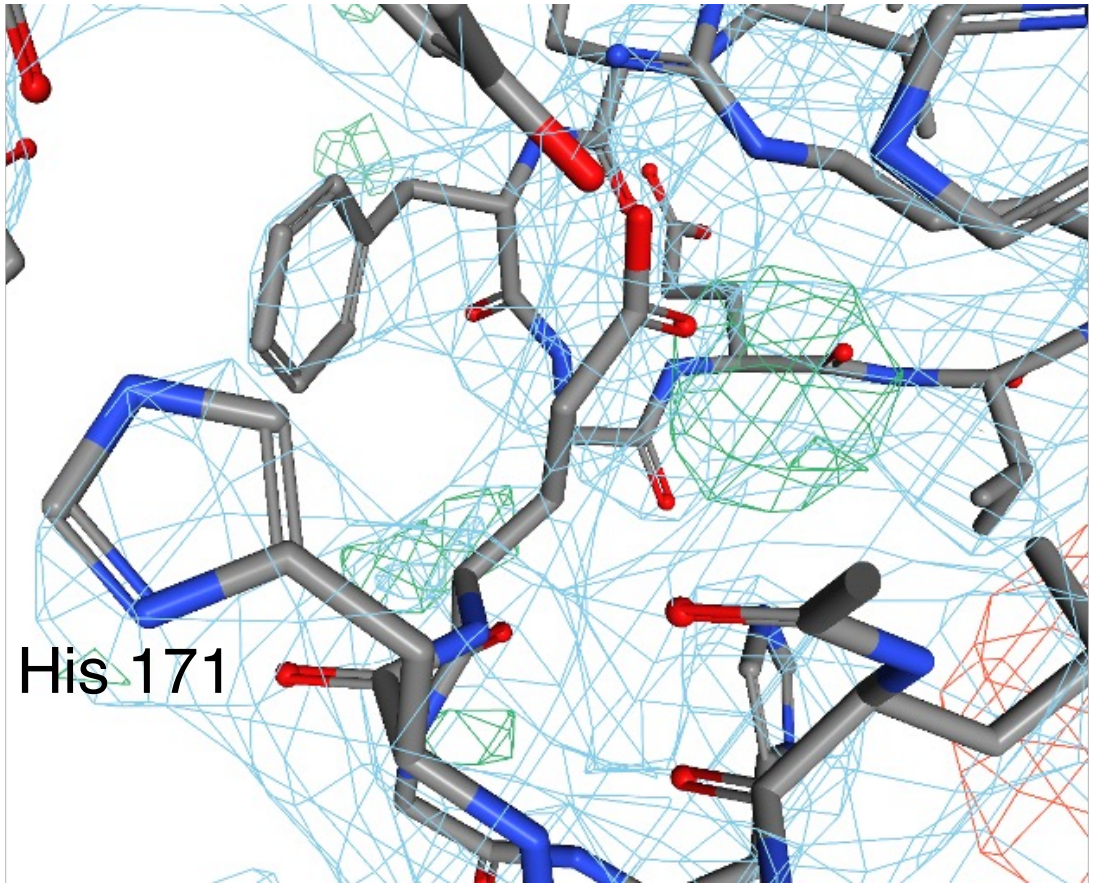


How to choose the right tautomer and protonation state?

Which tautomer and protonation?

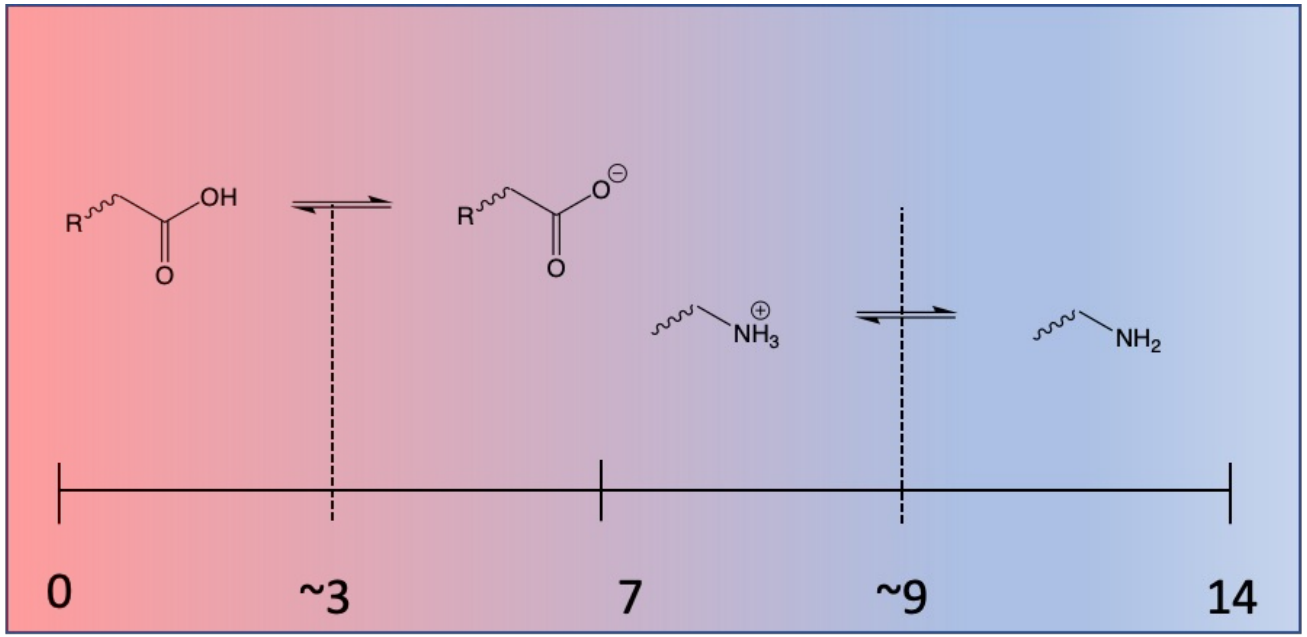


δ-tautomer ε-tautomer protonated



We need to worry about pKa to decide protonation

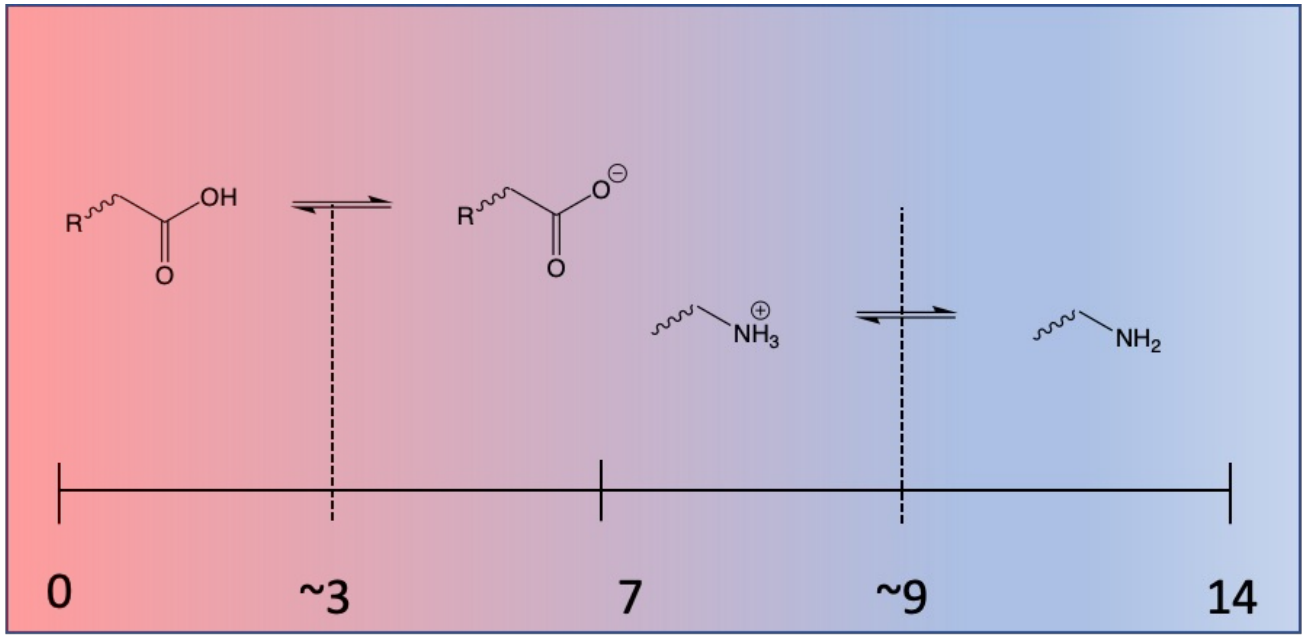
- pKa: pH at which an acidic/basic group is 50% protonated/deprotonated.
- pKas are not fixed things!
- “Standard” values refer to the situation when the group is in dilute aqueous solution.
- Groups buried in the centre of hydrophobic proteins or close to other charged groups can show large pKa shifts.



We need to worry about pKa to decide protonation

- Finding the right protonation state for proteins and ligands can be challenging!

- Poorly chosen protonation states
  - can lead to inaccurate simulations by (de)stabilising interactions
  - can lead to incorrectly estimating binding affinities or protein local or collective dynamics.

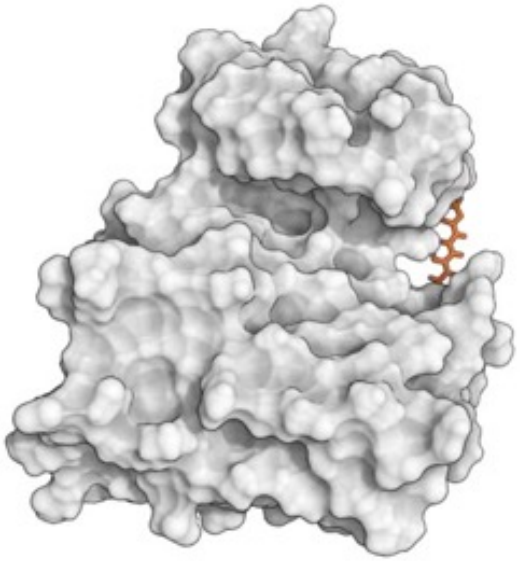


Amino acids that need protonation state consideration

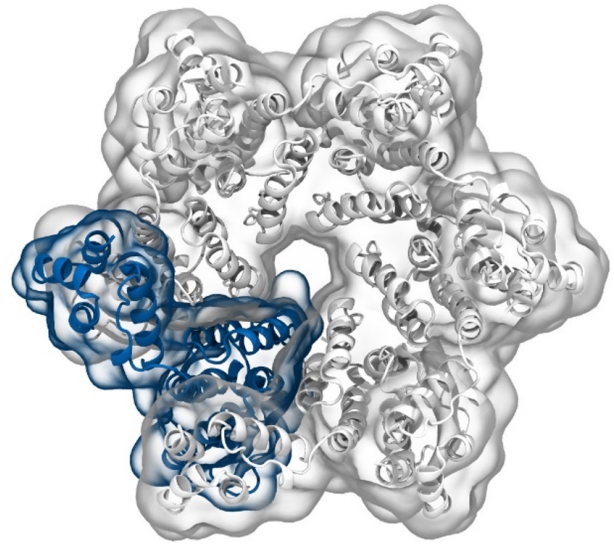
Amino acid	pKa	options	Significance*
Aspartic acid	3.65	–COOH instead of COO^-?	possible
Glutamic acid	4.25	–COOH instead of COO^-?	possible
Histidine	6.00	protonated instead of neutral?	very possible
Cysteine	8.18	$–S^-$ instead of SH?	very possible
Tyrosine	10.07	$–O^-$ instead of OH?	possible
Lysine	10.53	$–NH_2$ instead of NH_3^+?	possible
Arginine	12.48	neutral instead of protonated?	unlikely

* for a simulation around physiological pH

Simulation of Biomolecules



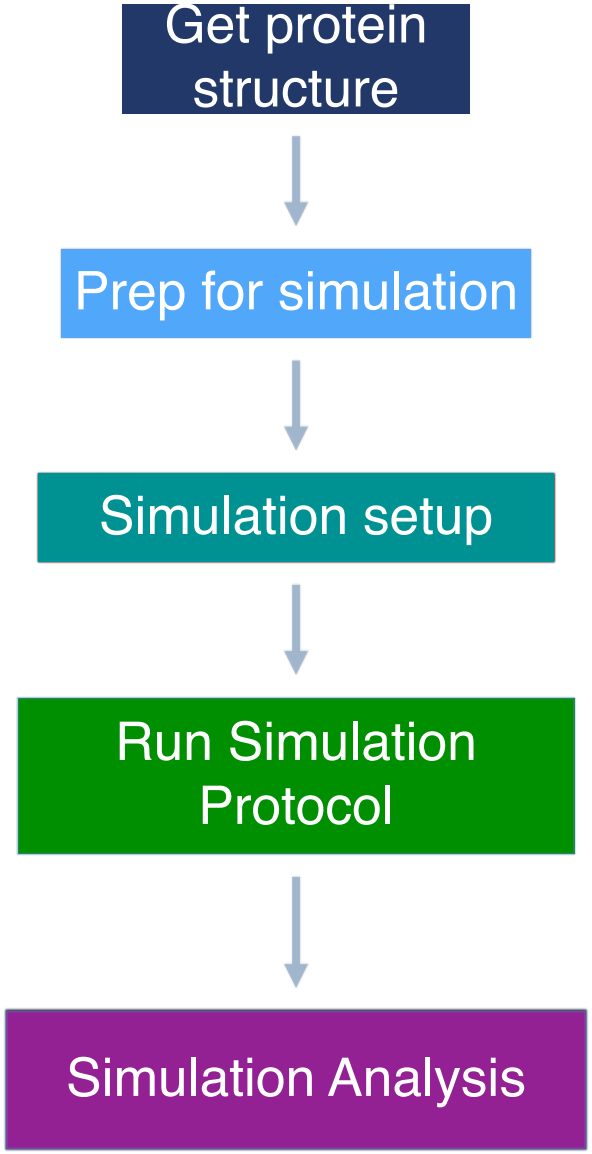
Setting up a protein simulation



Delivered by **Tim Spankie**
University of Edinburgh
January 2025

Based on content from a ten-hour series
by Matteo Degiacomi (matteo.degiacomi@ed.ac.uk) and Antonia Mey (antonia.mey@ed.ac.uk)

A typical workflow for molecular dynamics



Getting your protein structure

AlphaFold Protein Structure Database

RCSB PDB PROTEIN DATA BANK

Getting ligands/co-factors

ZINC20

RCSB PDB PROTEIN DATA BANK

Disclaimer!

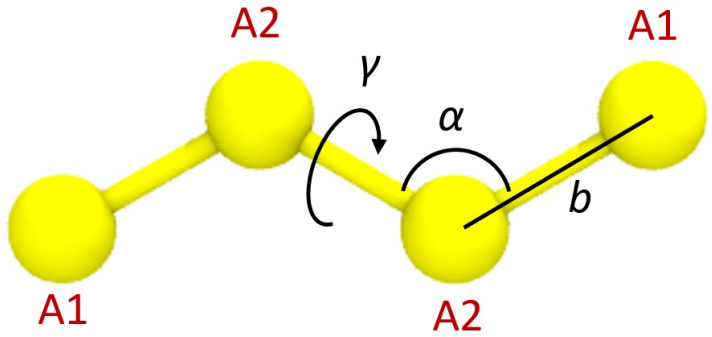
Running biomolecular MD can take days on specialised hardware.

Today we will *not* run any of them, and instead will focus on fundamental principles using small molecules.

Evaluating interatomic interactions

MD simulation require defining a system *topology.* This includes information on:

- *Connectivity* between atoms (covalent bond network)
- *Atom types* (depend on chemical environment, e.g., sp2 or sp3 Carbon)



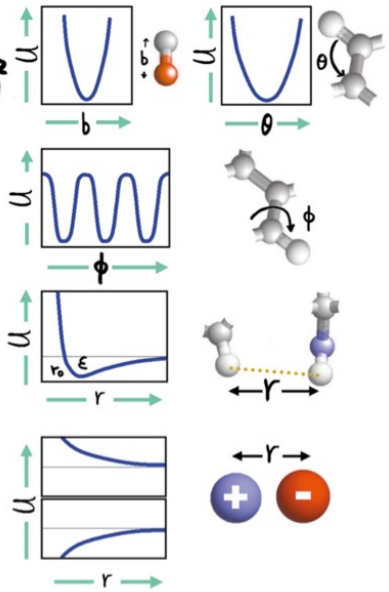
$$\vec{F}_i = -\frac{dU}{d\vec{x}_i}$$

$$U = \sum_{\text{All Bonds}} \tfrac{1}{2}K_b(b-b_o)^2 + \sum_{\text{All Angles}} \tfrac{1}{2}K_\theta(\theta-\theta_o)^2$$

$$+ \sum_{\text{All Torsion Angles}} K_\phi[1-\cos(n\phi+\delta)]$$

$$+ \sum_{\text{All nonbonded pairs}} \epsilon\left[\left(\tfrac{r_o}{r}\right)^{12} - 2\left(\tfrac{r_o}{r}\right)^{6}\right]$$

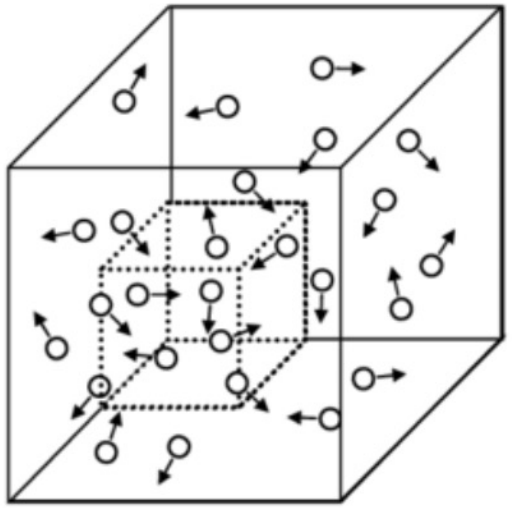
$$+ \sum_{\text{All partial charges}} 332 q_i q_j / r$$



The functional form of interatomic interactions and their strengths is defined by a *force field.*

Choosing your thermodynamic ensemble

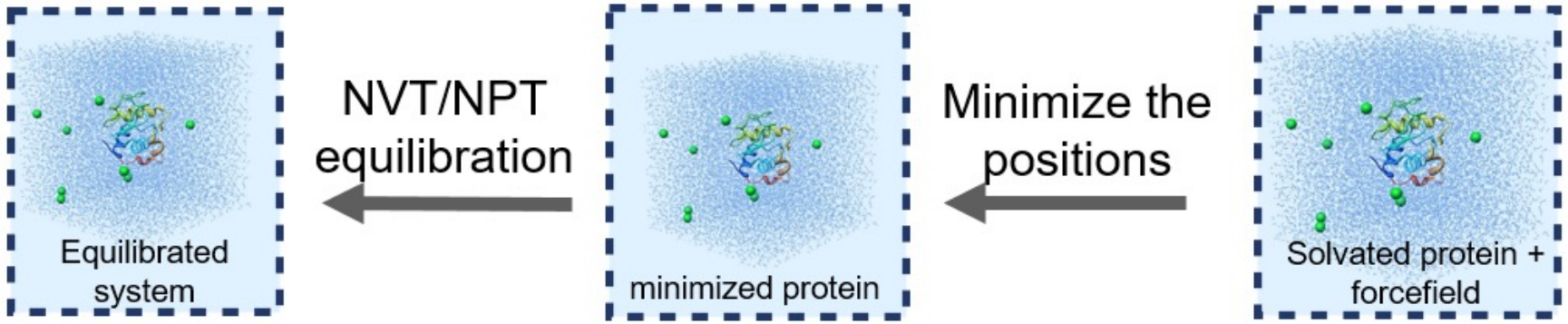
Simulations replicate a specific *thermodynamic ensemble* (typically NVT or NPT), or even grand canonical (μVT)



μVT

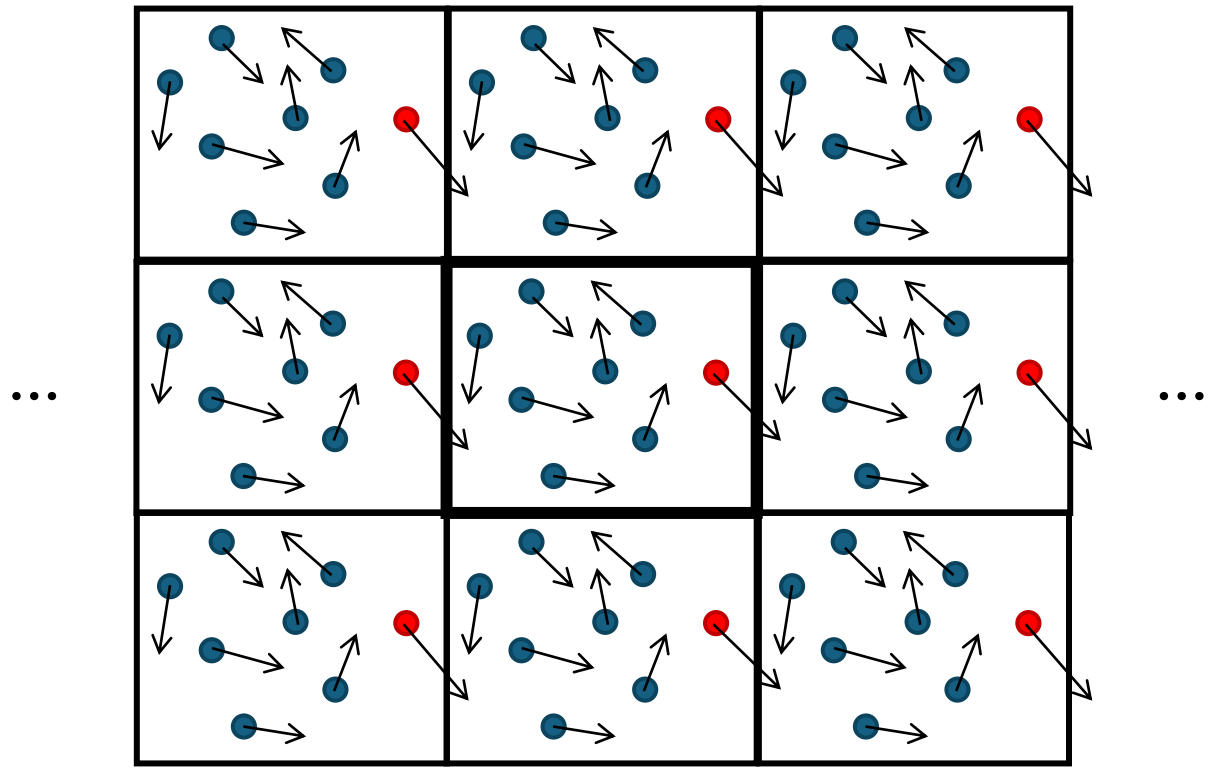
You will have different options to include *thermostats* (scaling atom velocities) and *barostats* (scaling positions) in your calculations:

- Nose-Hoover
- Berendsen
- Parrinello-Rahman
- Langevin piston
- …

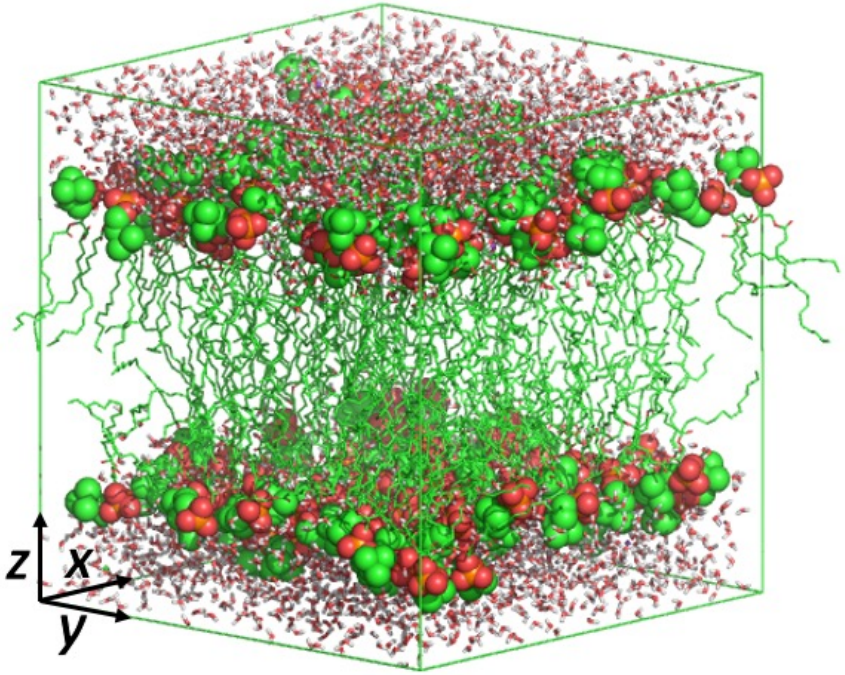


Periodic boundary conditions (PBC) and pressure coupling

Useful to reduce finite-size effect and simulate bulk

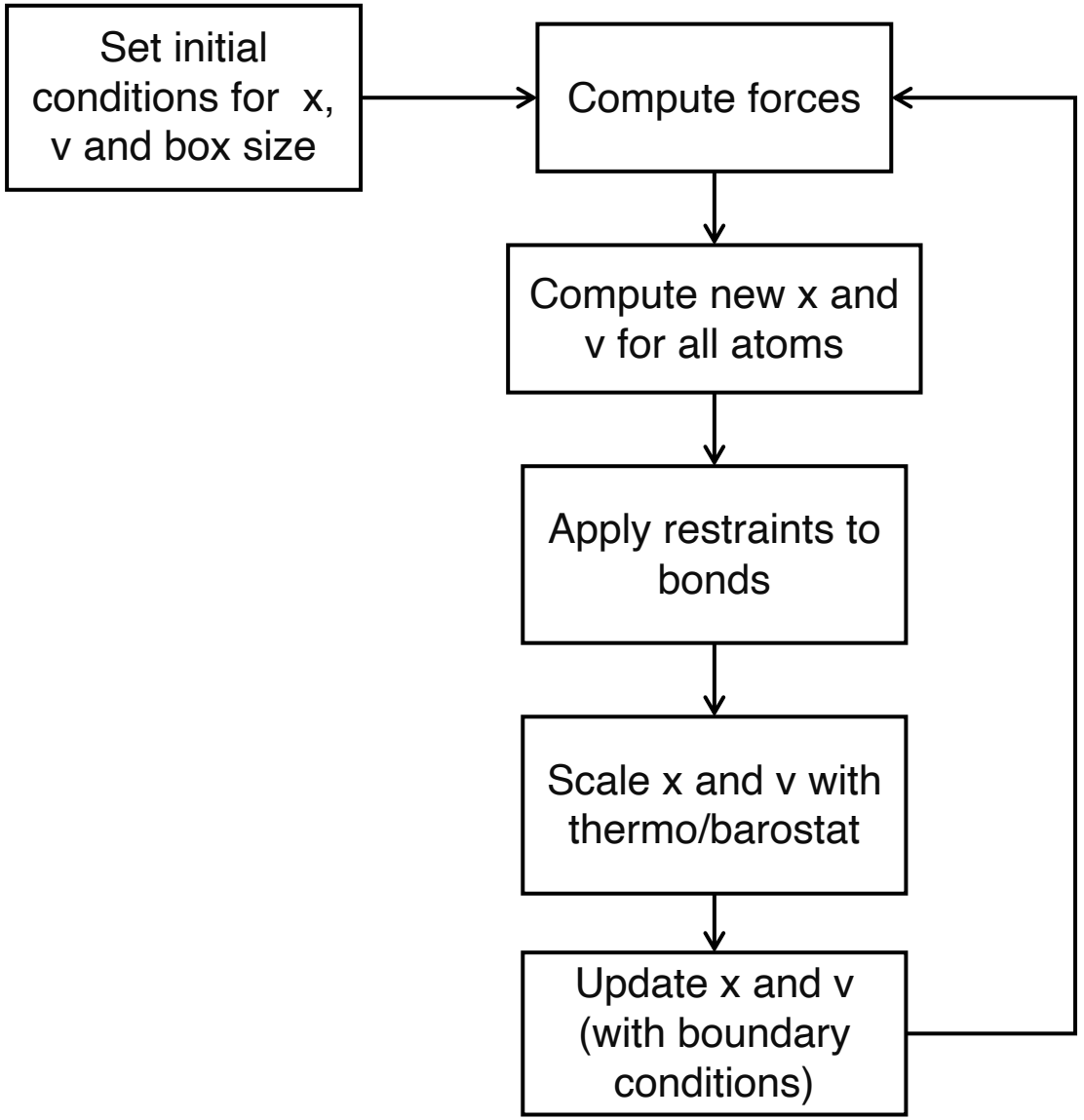


Typically, PBC applied in *x*, *y* and *z* direction



For membrane systems, use semi-isotropic pressure coupling (x, $y \neq z$, lipids compressibility is direction-dependent)

A Molecular Dynamics timestep



Molecular dynamics require multiple steps for the setup of simulations

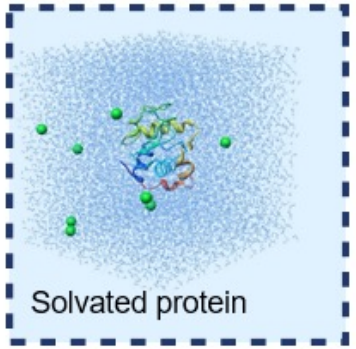
Disulphide bridges

Protonation

Crystal Waters Cofactors

Alternative side chains

Solvate in water and add ions (*typically, neutralize the system*)



Apply forcefield description

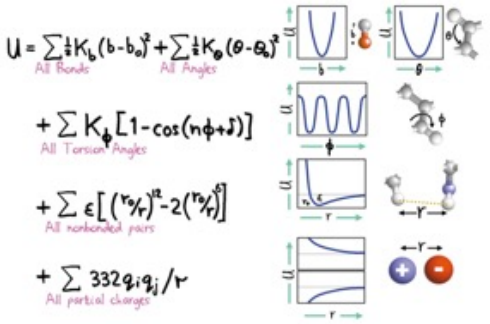
$$\vec{F}_i = -\frac{dU}{d\vec{x}_i}$$

$$U = \sum_{\text{All Bonds}} \tfrac{1}{2}K_b(b-b_0)^2 + \sum_{\text{All Angles}} \tfrac{1}{2}K_\theta(\theta-\theta_0)^2$$

$$+ \sum_{\text{All Torsion Angles}} K_\phi[1-\cos(n\phi+\delta)]$$

$$+ \sum_{\text{All nonbonded pairs}} \epsilon\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right]$$

$$+ \sum_{\text{All partial charges}} 332 q_i q_j / r$$



Production simulation

Equilibrated system

NVT/NPT equilibration

minimized protein

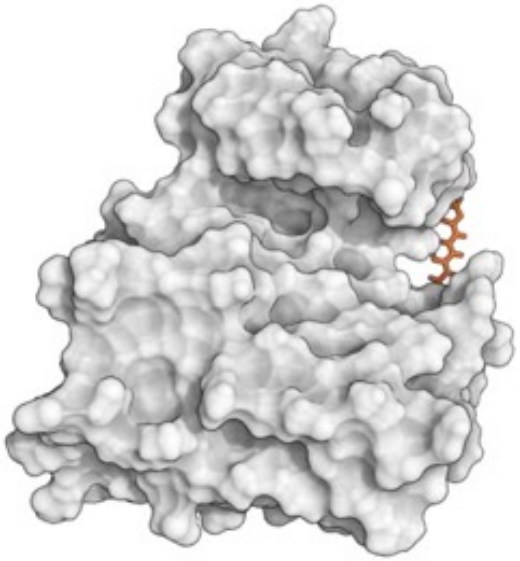
Minimize the positions

Solvated protein + forcefield

http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/lysozyme/01_pdb2gmx.html

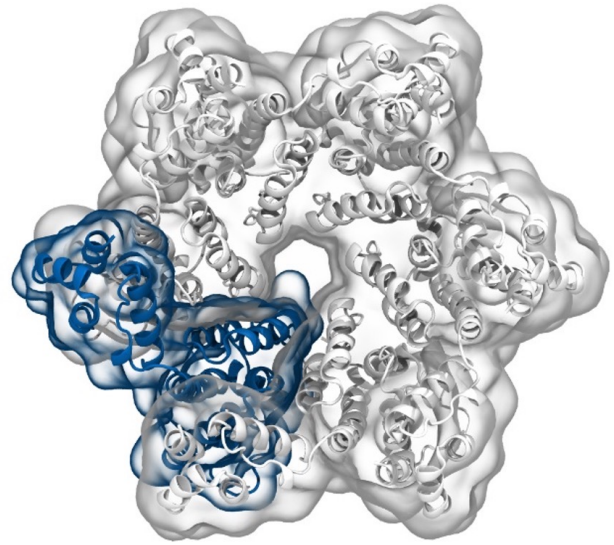
Levitt, ***Nature Struct. Biol.***, 2001

Simulation of Biomolecules



Basic Simulation Analysis

Delivered by **Tim Spankie**
University of Edinburgh
January 2025



Based on content from a ten-hour series https://github.com/CCPBioSim/BioSim-analysis-workshop
by Matteo Degiacomi (matteo.degiacomi@ed.ac.uk) and Antonia Mey (antonia.mey@ed.ac.uk)

Volume and pressure equilibration

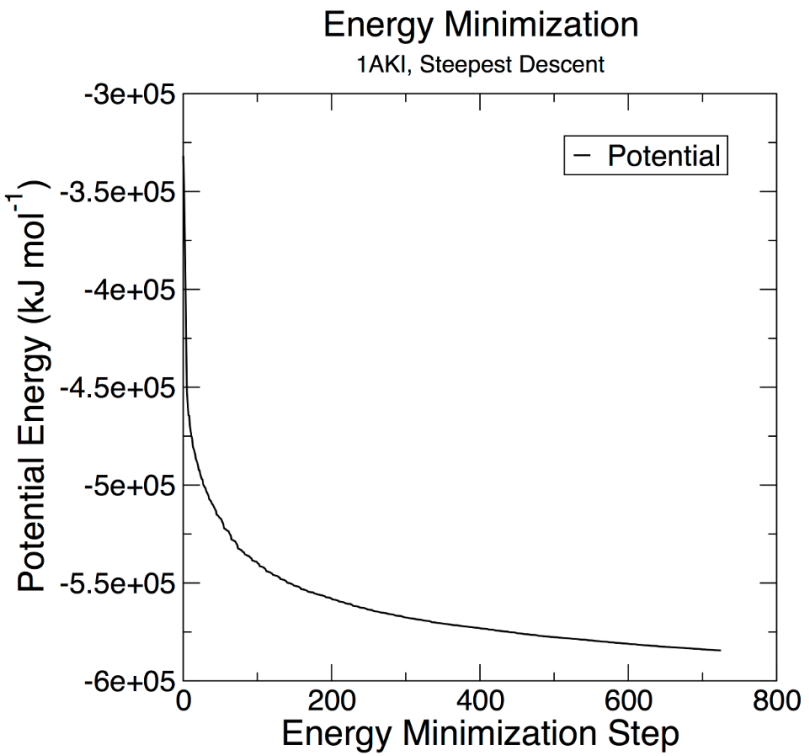
Steps until production:

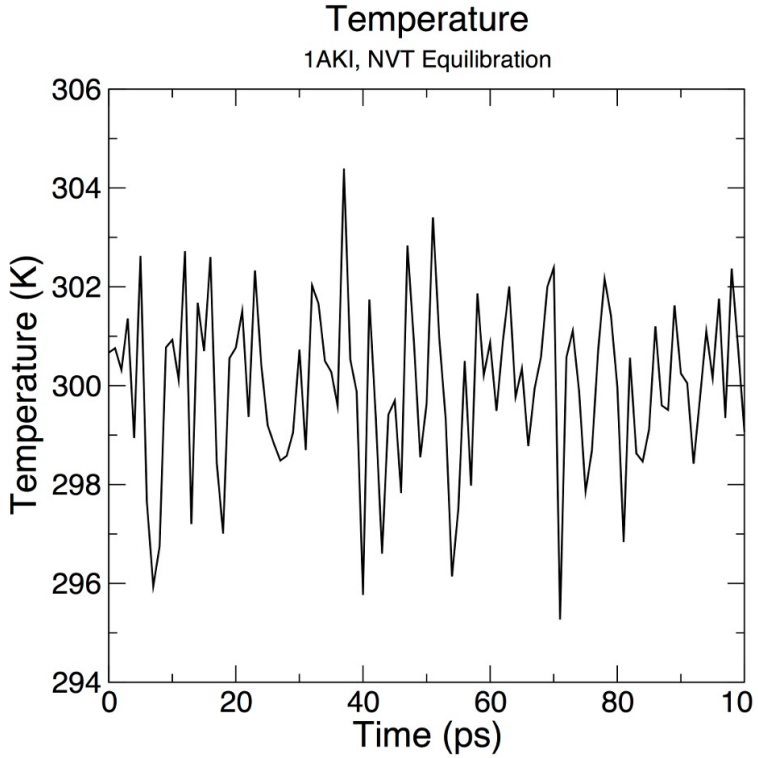
Production ensemble

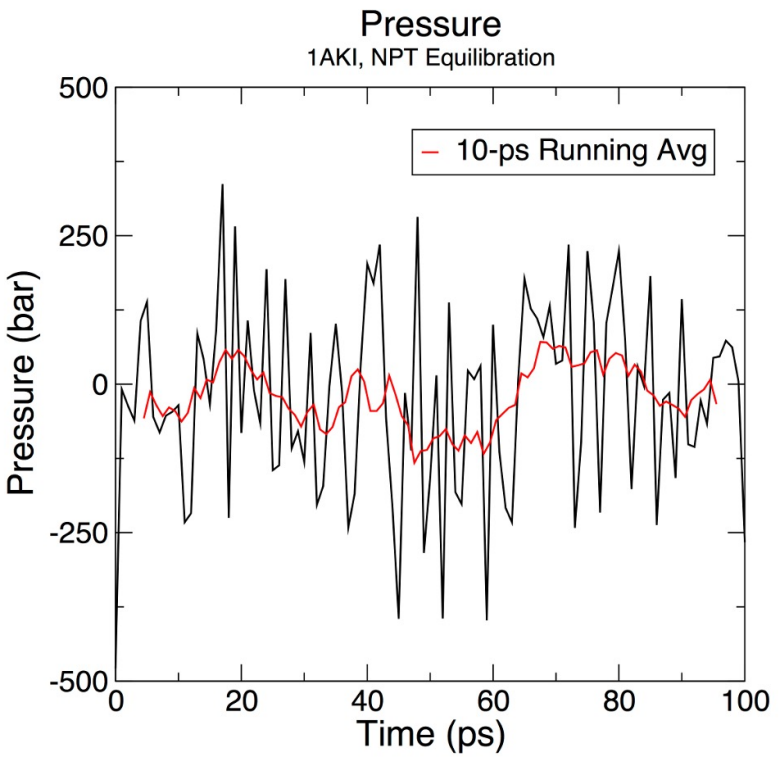
Minimisation → NVT → NPT → NPT

NVT: (short simulation to relax to temperature of interest)

NPT: (short simulation to relax to density of interest)

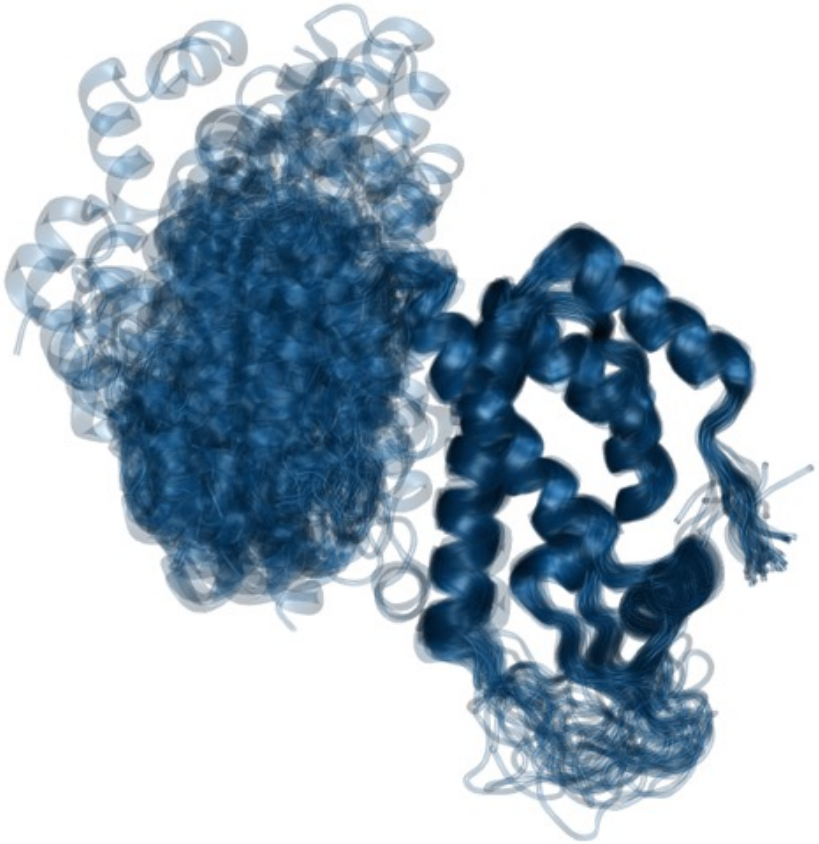


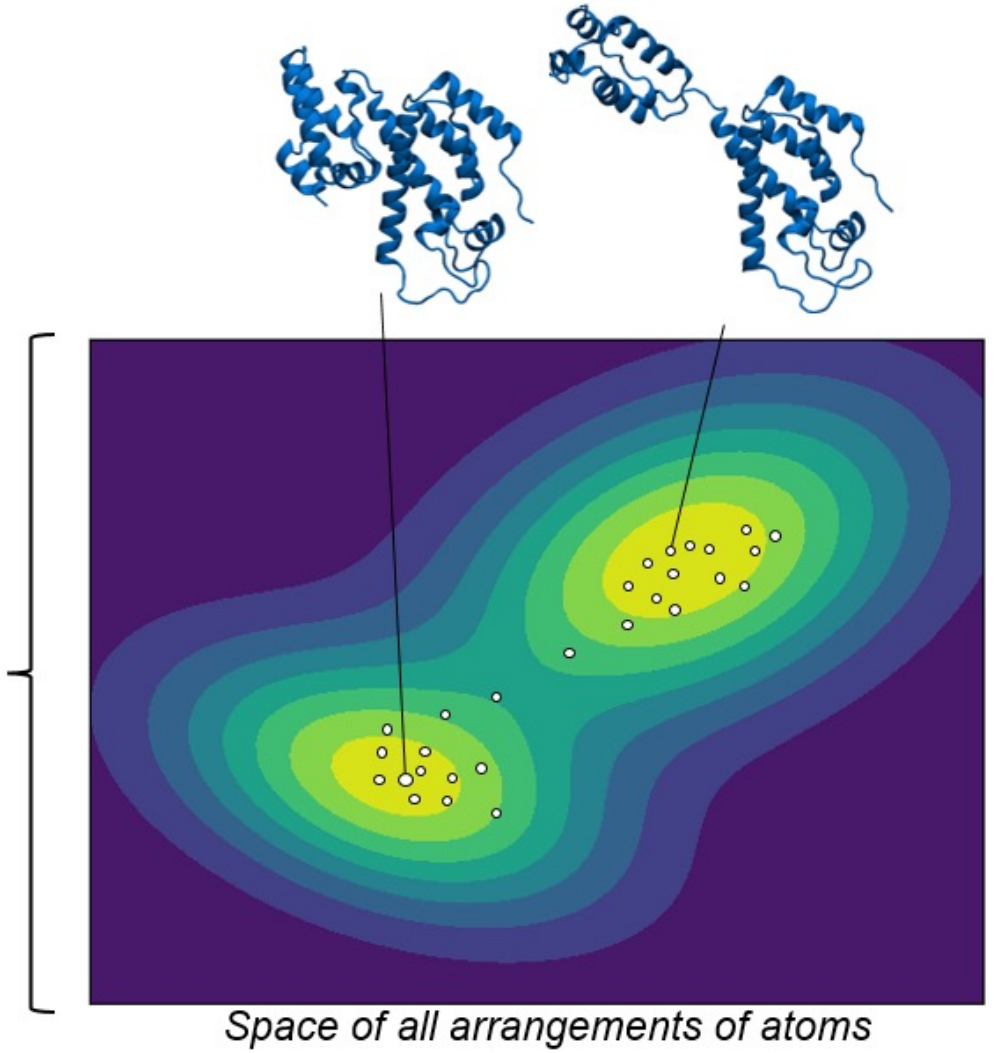




*Living J.Comp.Mol.Sci. 2019,**1**,5957*

[Recall] Sampling the conformational space



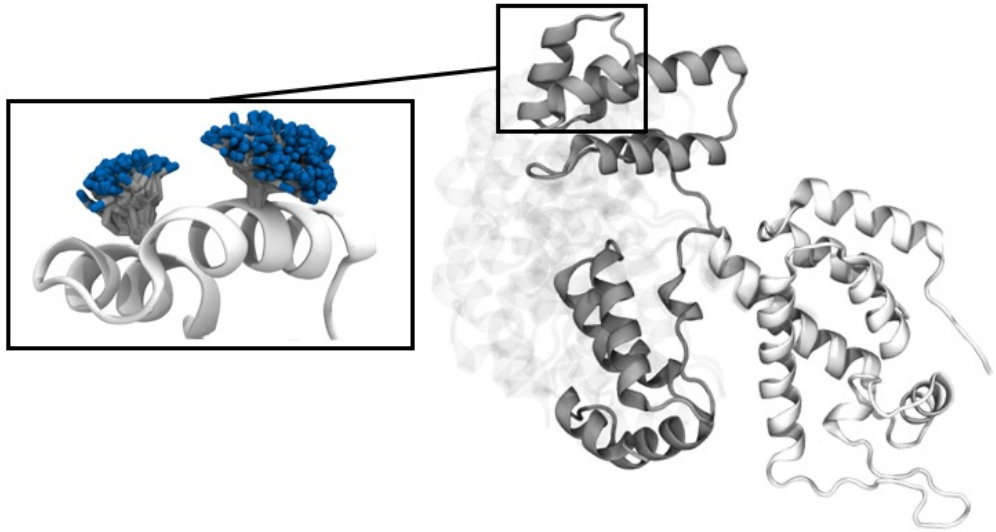


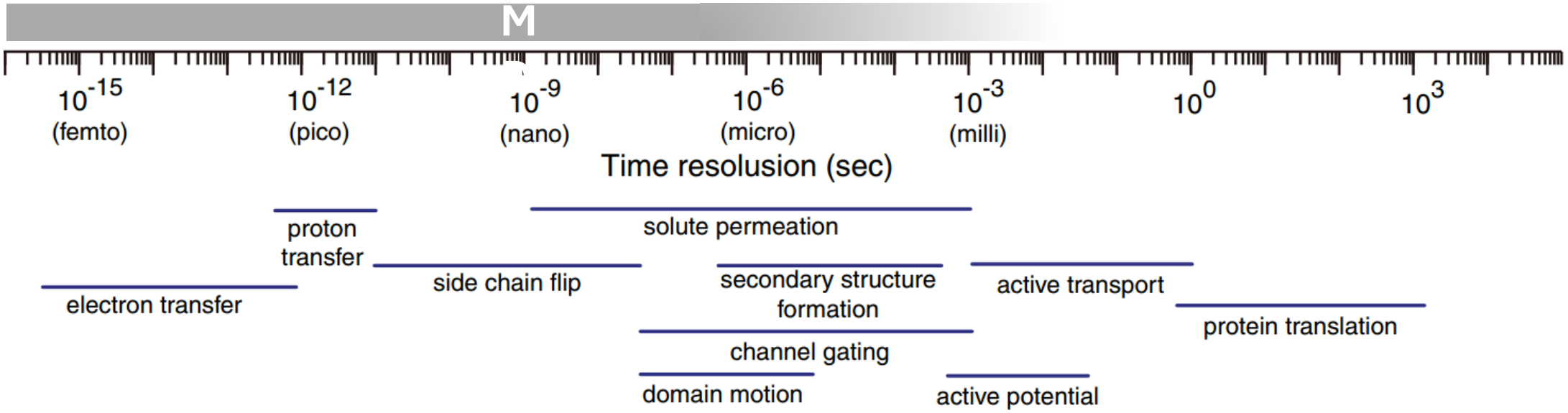
Probability of sampling a conformation is inversely proportional to its energy: $p_i \propto e^{-\varepsilon_i / kT}$

Timescales in biology

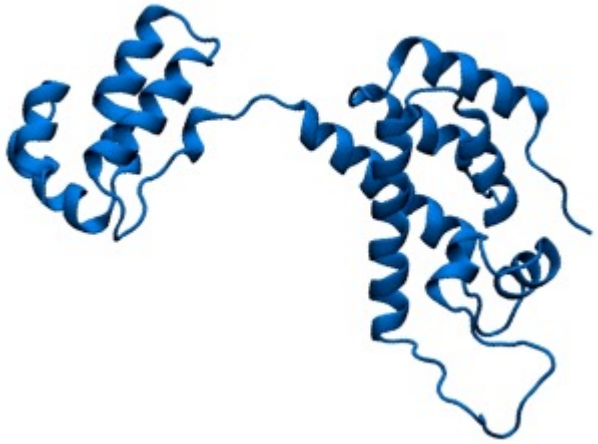
Different regions, different timescales:

- Side chains faster than backbone
- Loops faster than helices and sheets
- backbone faster than side chains
- Protein surface faster than core





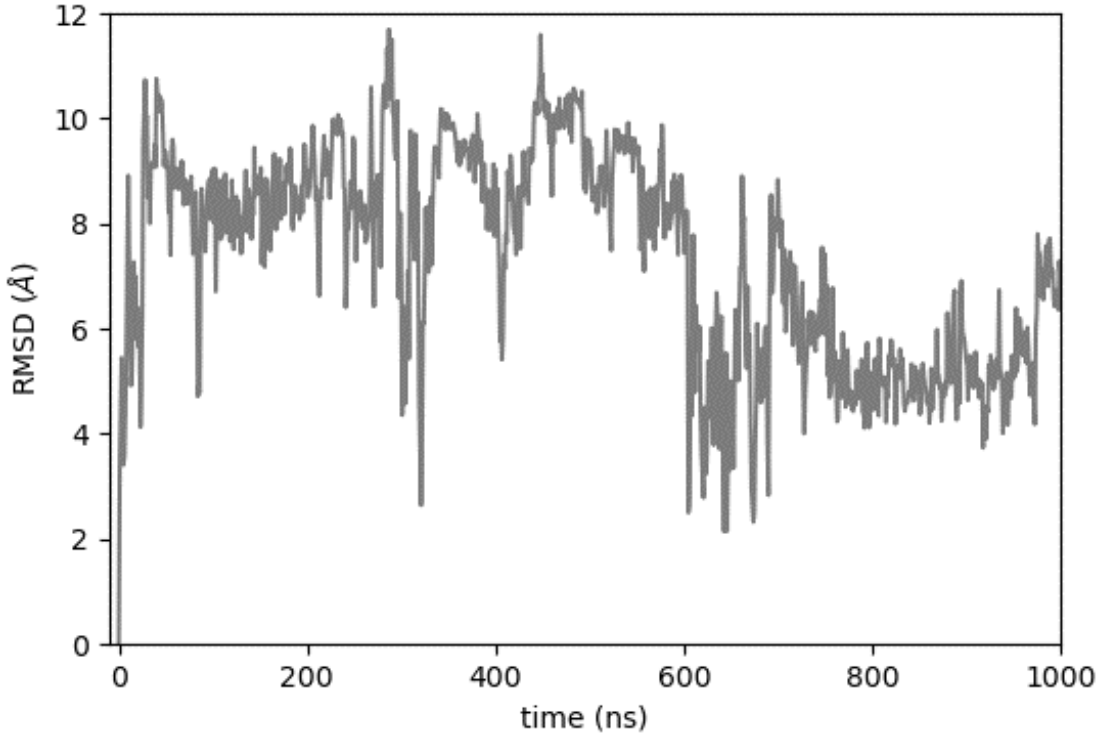
Root Mean Square Deviation (RMSD)



Given a system with *N* atoms, and a reference arrangement x_0:

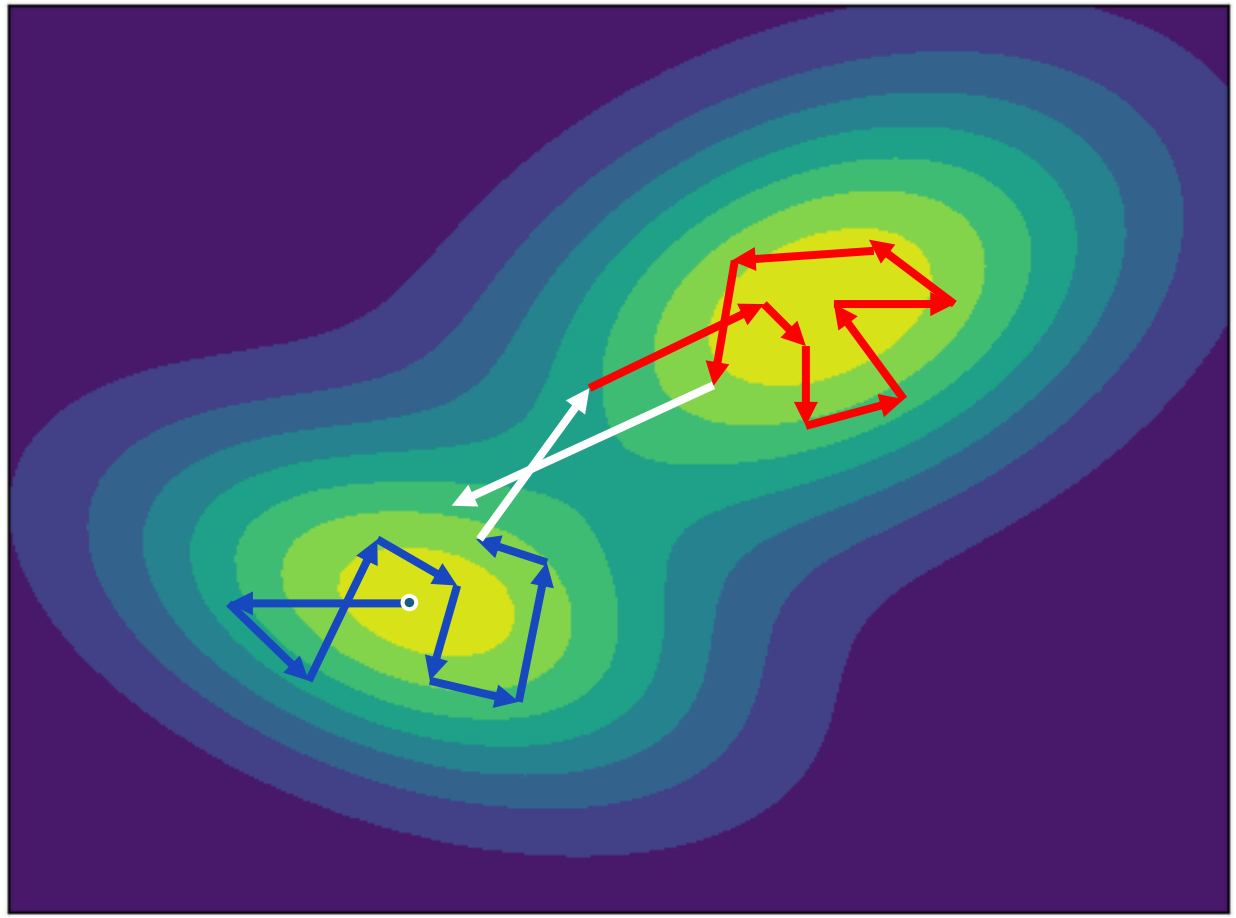
$$RMSD = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(X_i - x_0)^2}$$

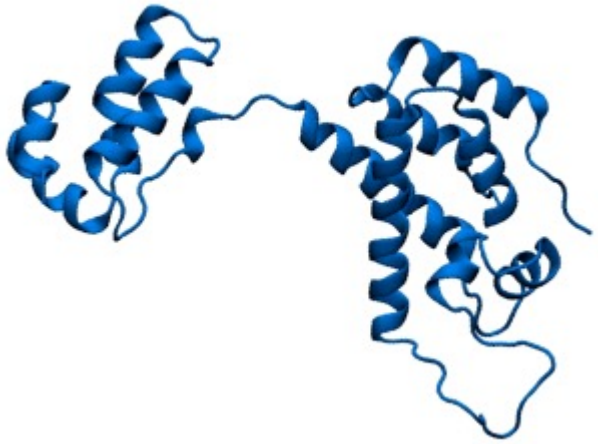
In MD, x_0 is often the first conformation in the simulation.

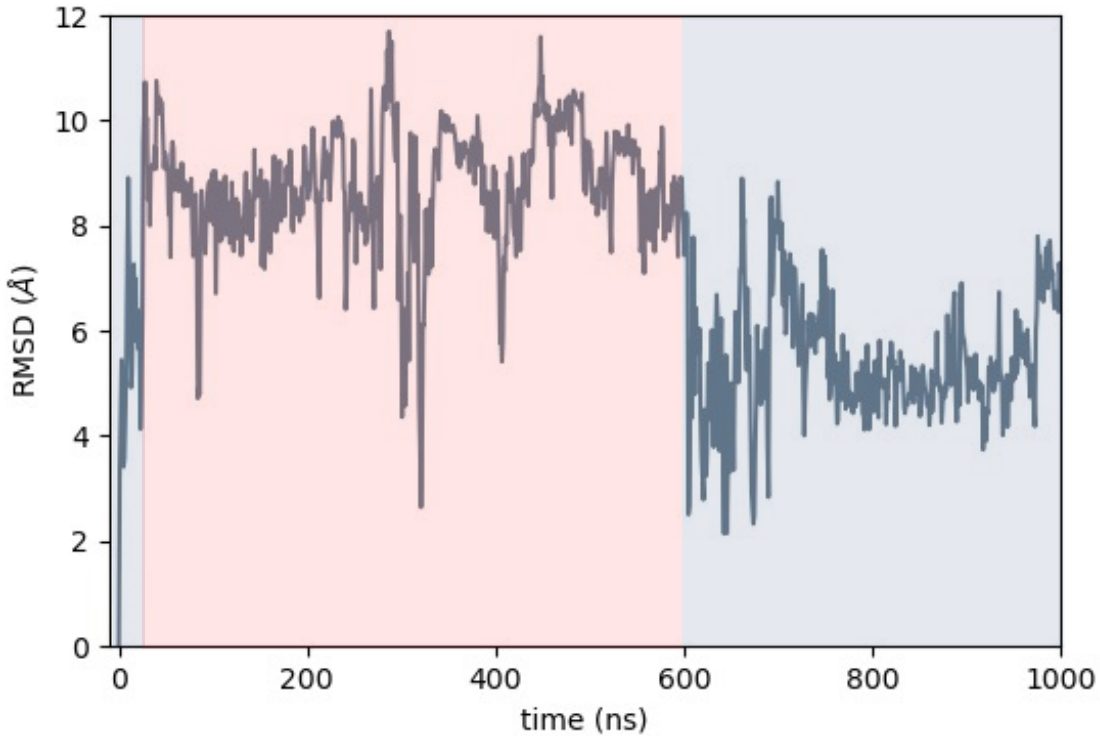


HIV-1 capsomer simulation from: *Degiacomi & Dal Peraro,* ***Structure,***

Convergence?



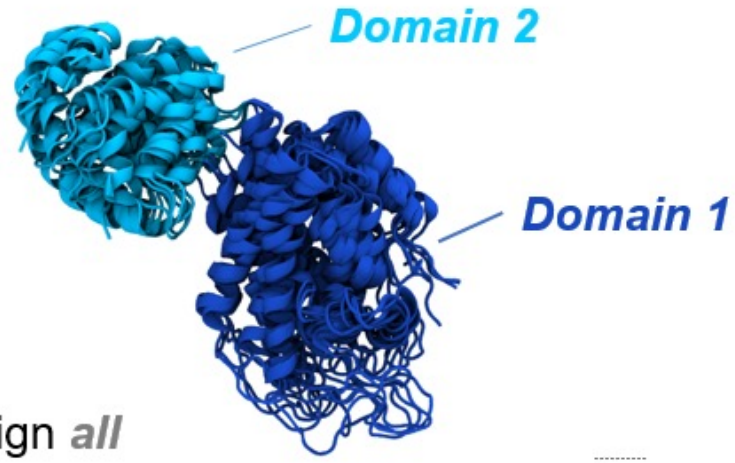




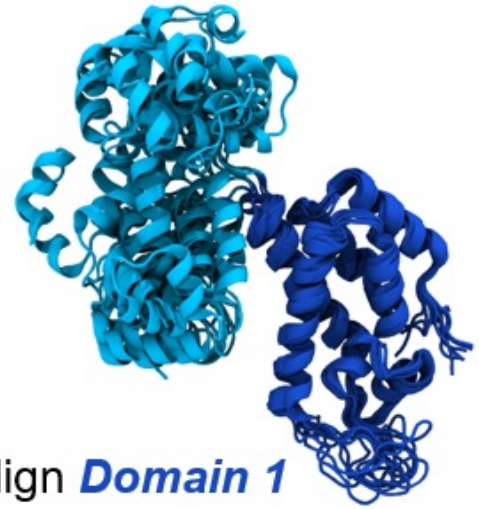
Refrain from using RMSD as a single indicator of simulation convergence.

HIV-1 capsomer simulation from: *Degiacomi & Dal Peraro,* ***Structure,***

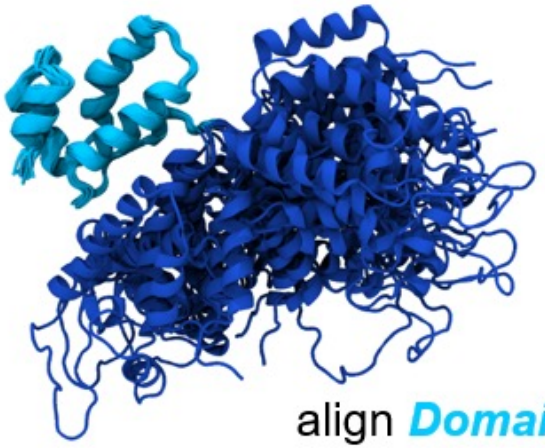
RMSD is alignment- and selection-dependent



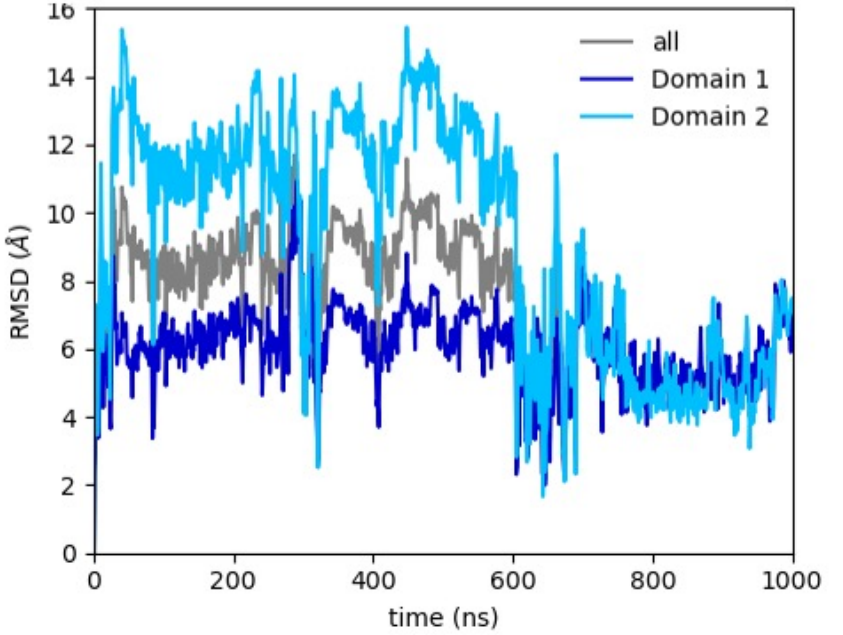
align *all*



align *Domain 1*

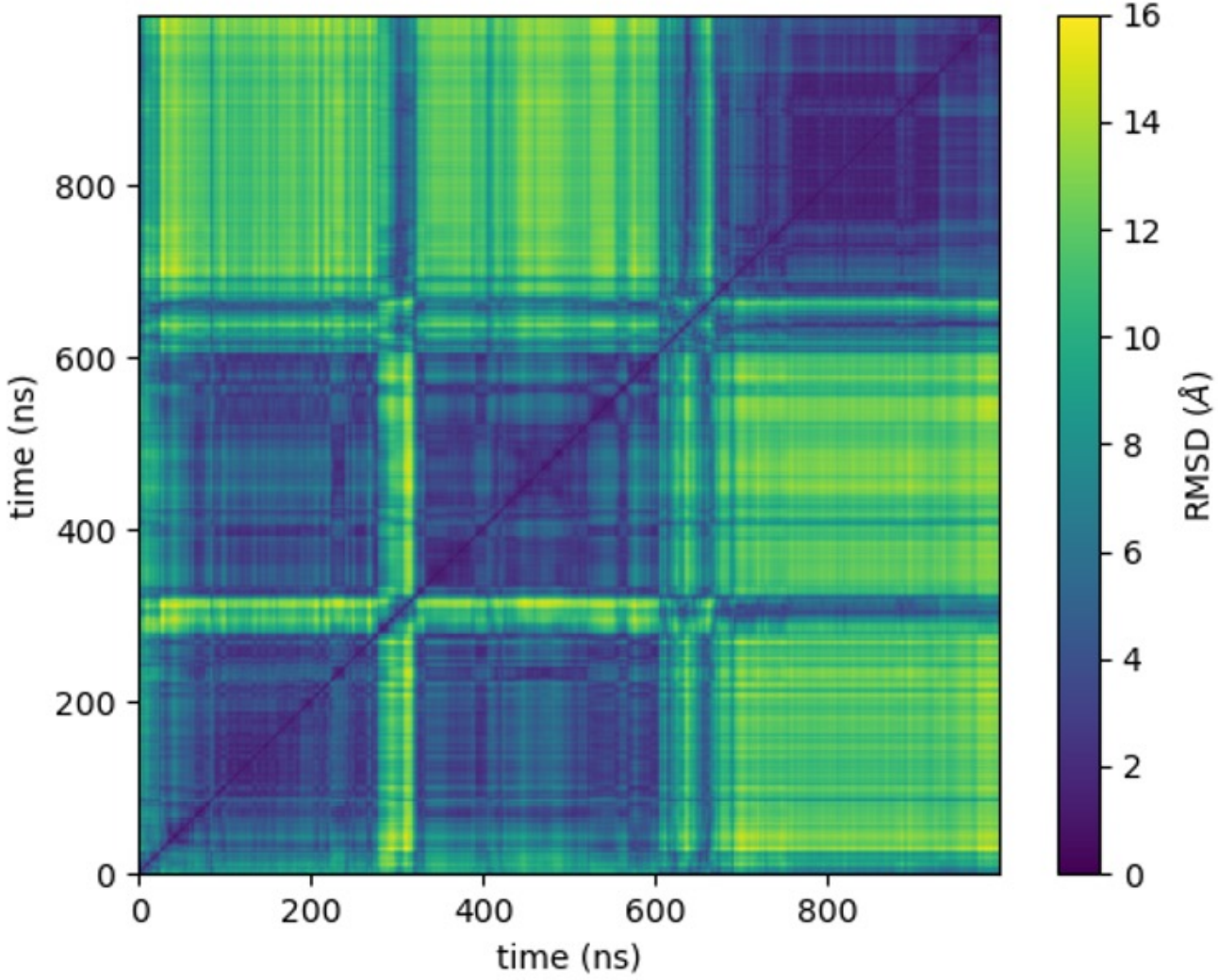


align *Domain 2*



Pairwise RMSD

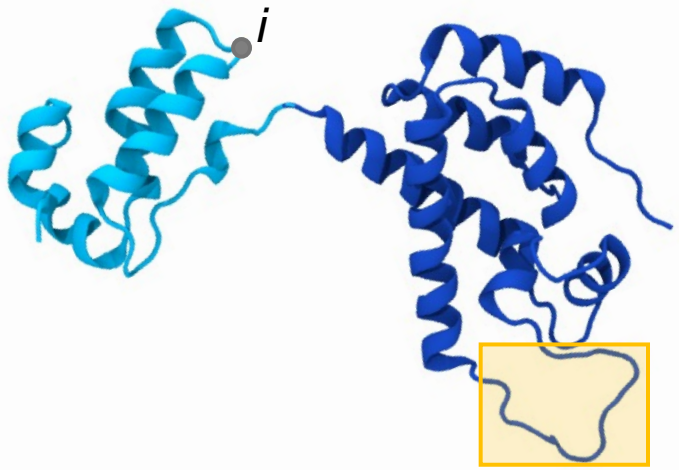
- Two structures with same RMSD from a reference are not forcefully similar to each other.
- Pairwise RMSD helps seeing if protein re-visits conformations throughout the simulation.

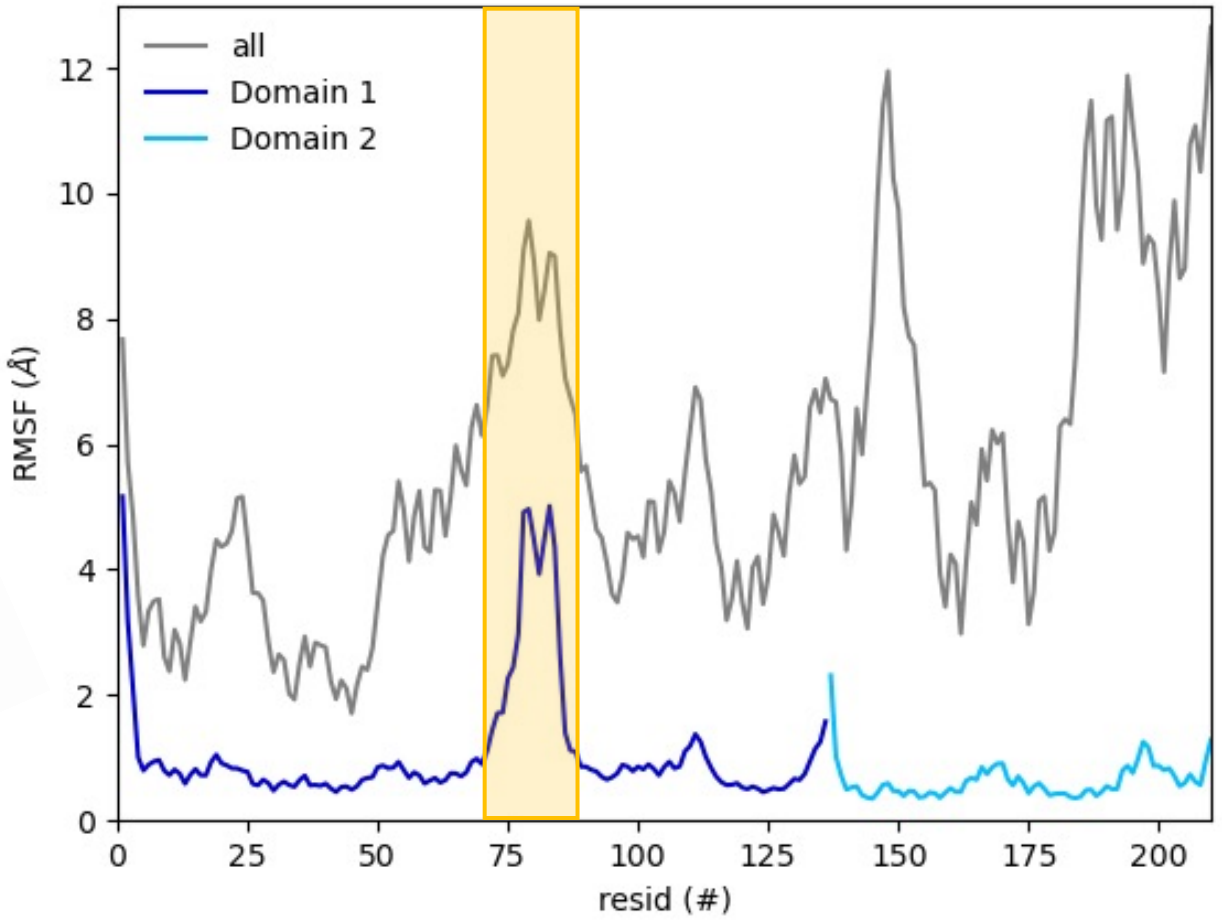


Root Mean Square Fluctuation (RMSF)

The RMSF σ_i of atom *i* calculates how much it fluctuates around its mean position $\langle X_i \rangle$.

$$\sigma_i = \sqrt{\langle (X_i - \langle X_i \rangle)^2 \rangle}$$



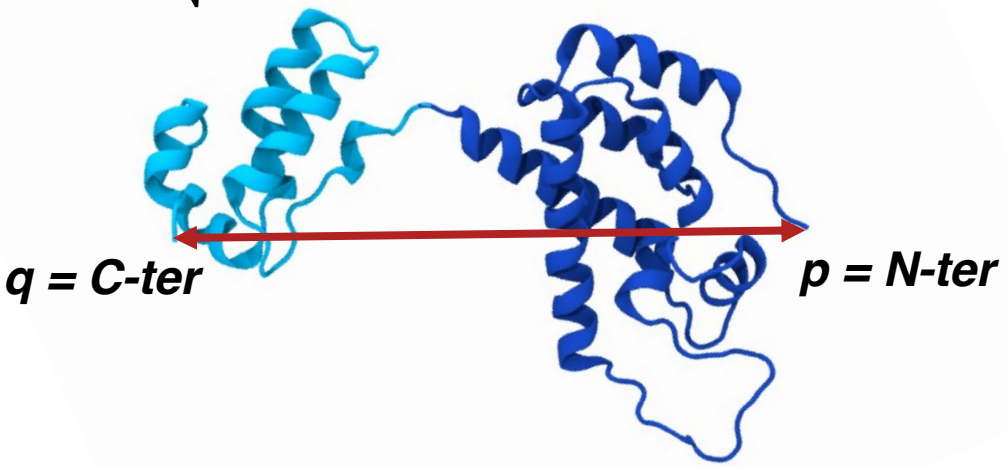


Helps identifying flexible/regid regions.
Typically done on C_α atoms.
Warning: result depends on alignment!

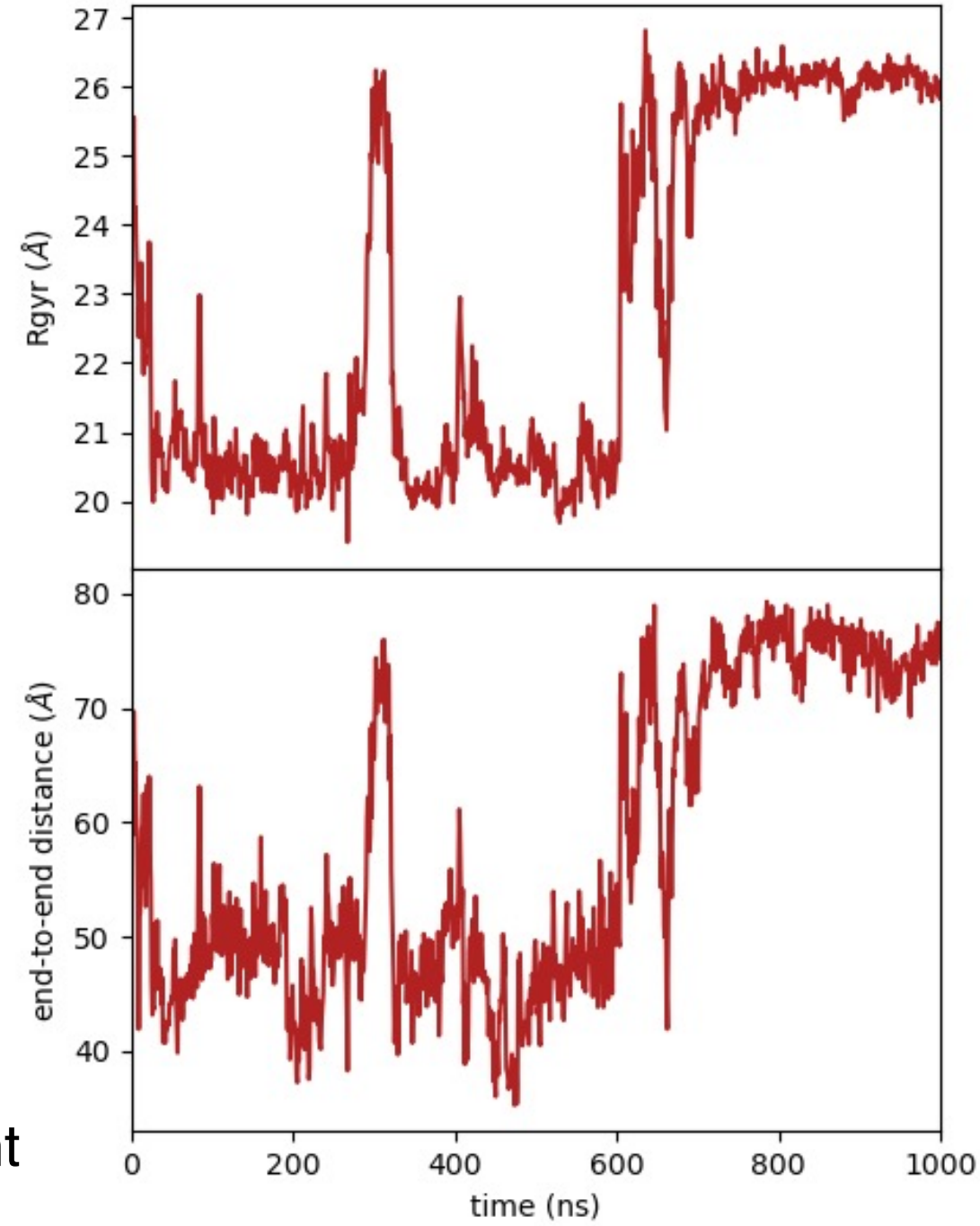
end-to-end distance and Radius of Gyration (Rg)

$$d(p,q) = \|p - q\|$$

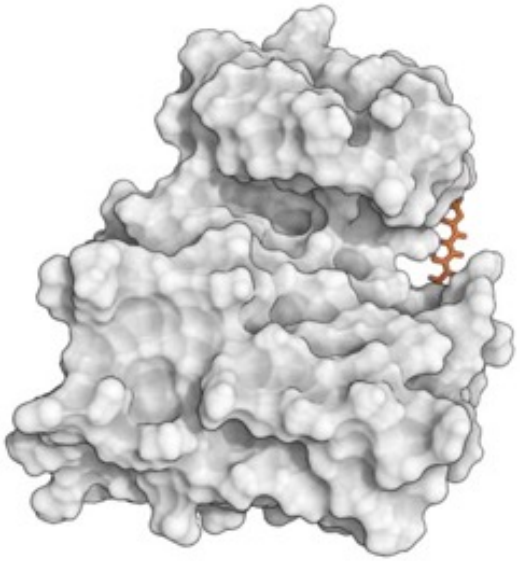
$$R_g = \sqrt{\frac{1}{N}|r_k - r_{mean}|^2}$$



Help quantifying protein compaction.
Internal properties: do *not* depend on alignment

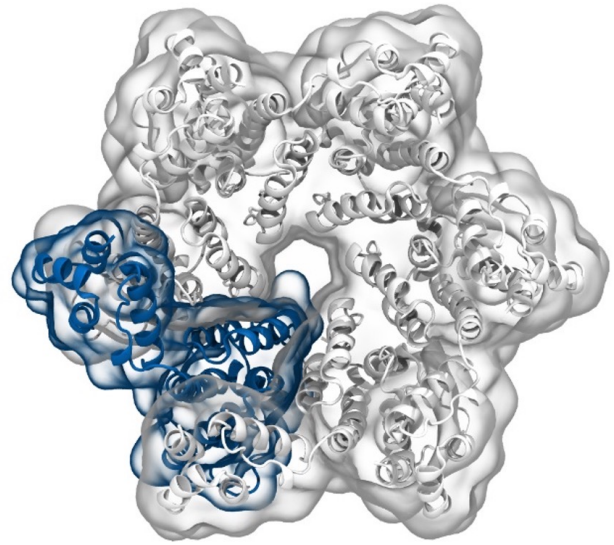


Simulation of Biomolecules



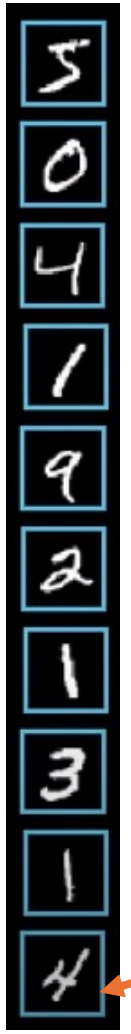
Dimensionality Reduction

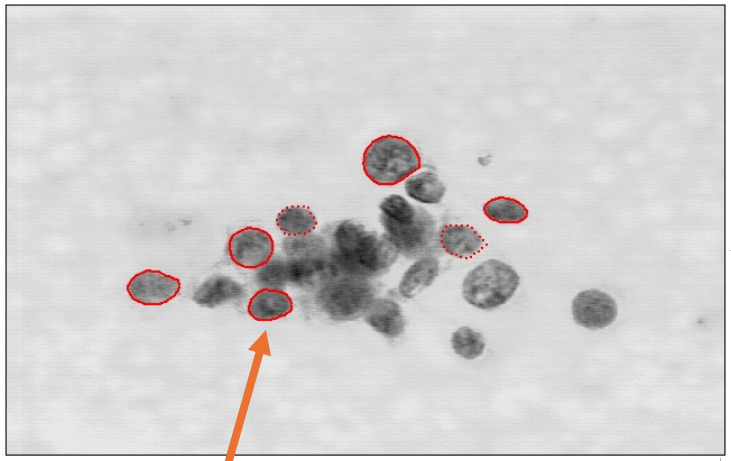
Delivered by **Tim Spankie**
University of Edinburgh
January 2025



Based on content from a ten-hour series https://github.com/CCPBioSim/BioSim-analysis-workshop
by Matteo Degiacomi (matteo.degiacomi@ed.ac.uk) and Antonia Mey (antonia.mey@ed.ac.uk)

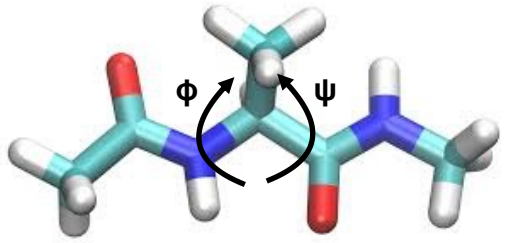
features are possible ways to represent data

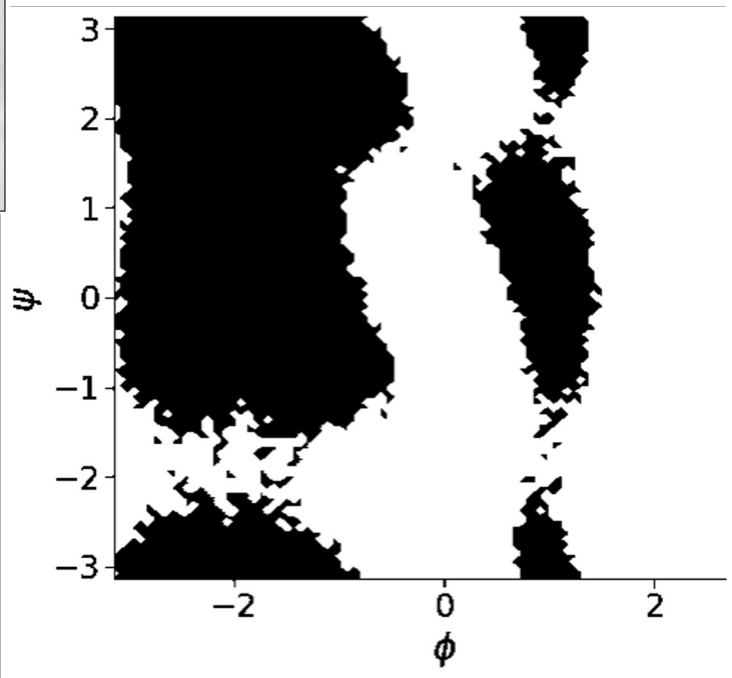




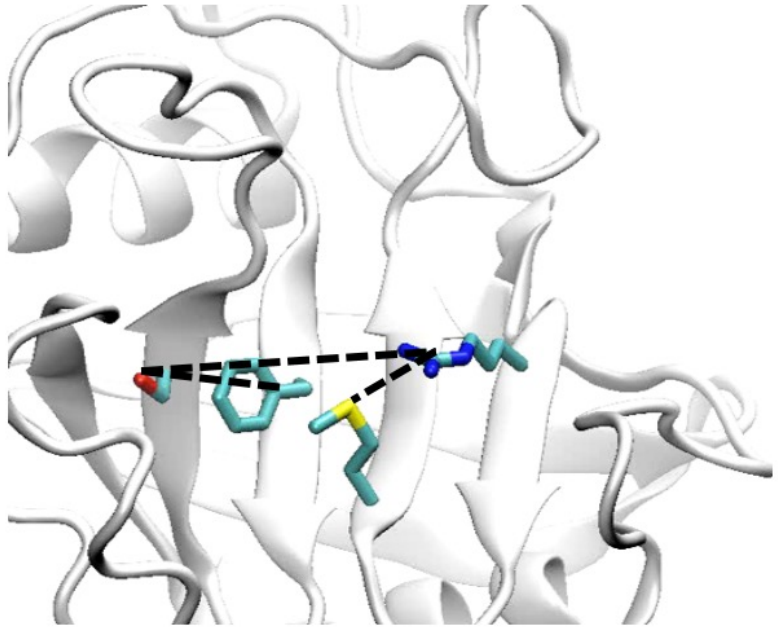
Sphericity (***1***)

Pixels colour
(28x28 = ***784***)





Torsional angles (***2***)



atomic positions (***459***)
atomic distances (***3***)

Not all features are useful

Task: predict the weather in Edinburgh using historical data

data = {X, Y, Z} → sun, rain, snow

{Temperature (C), ~~Temperature (K)~~, Humidity (g/m^3)}

2 decorrelated features

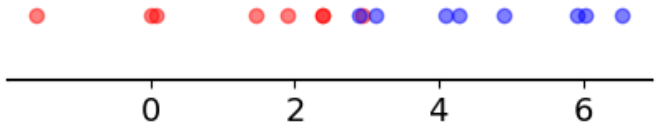
{Temperature (C), ~~Swiss cheese export (£)~~, Humidity (g/m^3)}

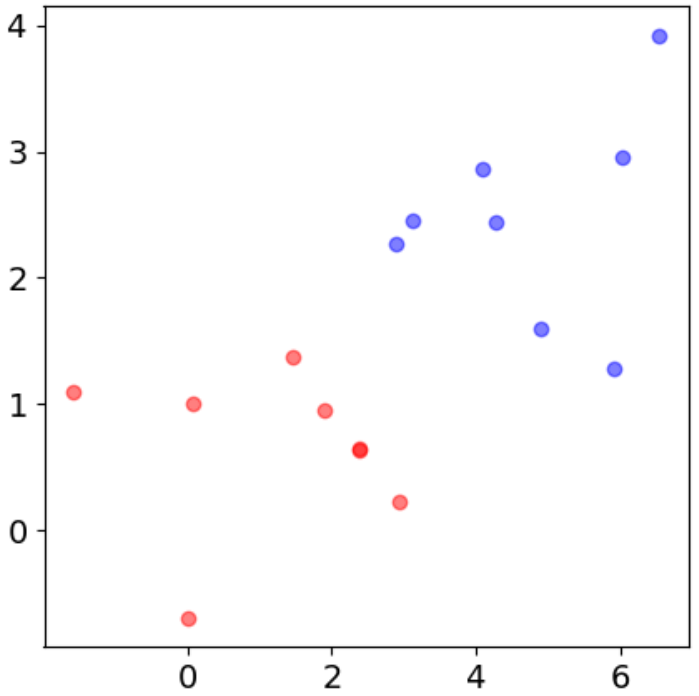
2 relevant features

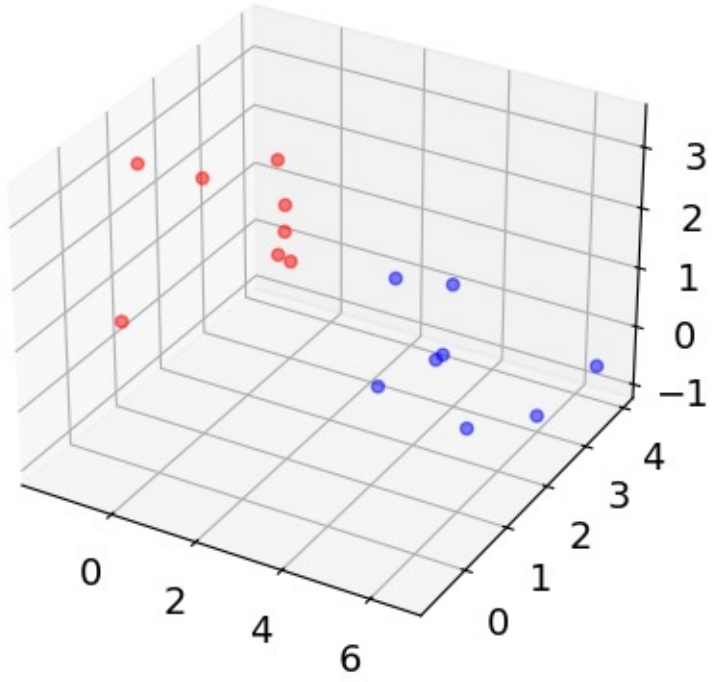
{Avg. gas expenditure (£), Heat strokes (#), Slipping accidents (#), Sunscreen sold (£)}

4 features connected to another quantity: temperature

Curse of dimensionality







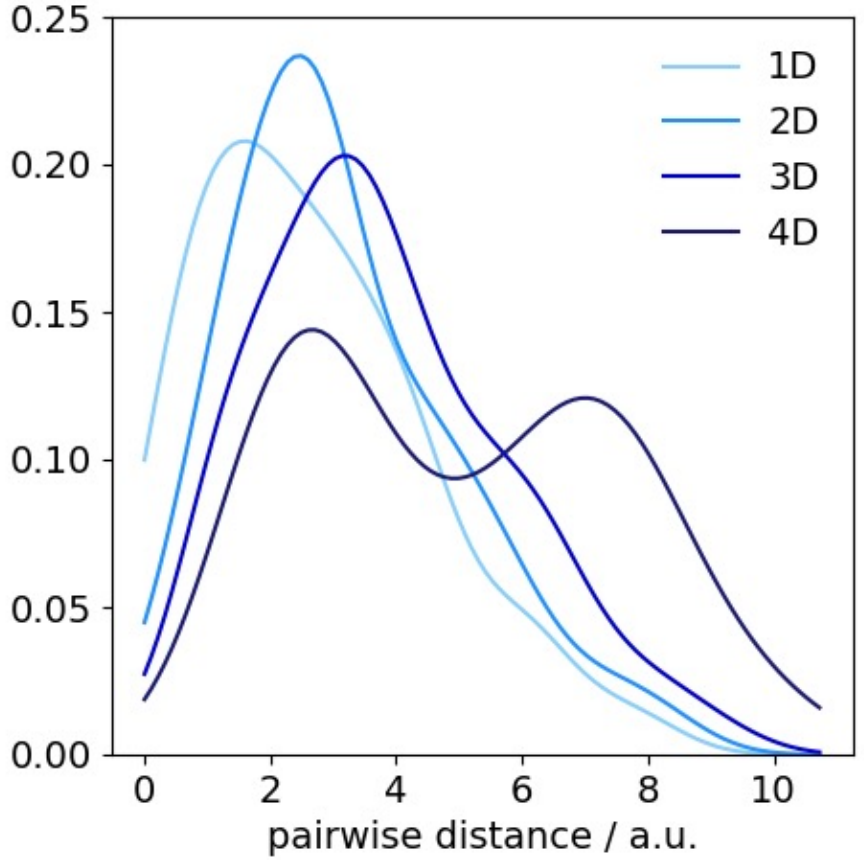
$$\|x\|_2 = \sqrt{\sum_{i=1}^{N} x_i^2}$$

Distances between M data points $\boldsymbol{x} \in \mathbb{R}^N$ increase, when N increases

Problem: less data density increases uncertainty on underlying data structure

[Extra] Curse of dimensionality

$$\|x\|_2 = \sqrt{\sum_{i=1}^{N} x_i^2}$$



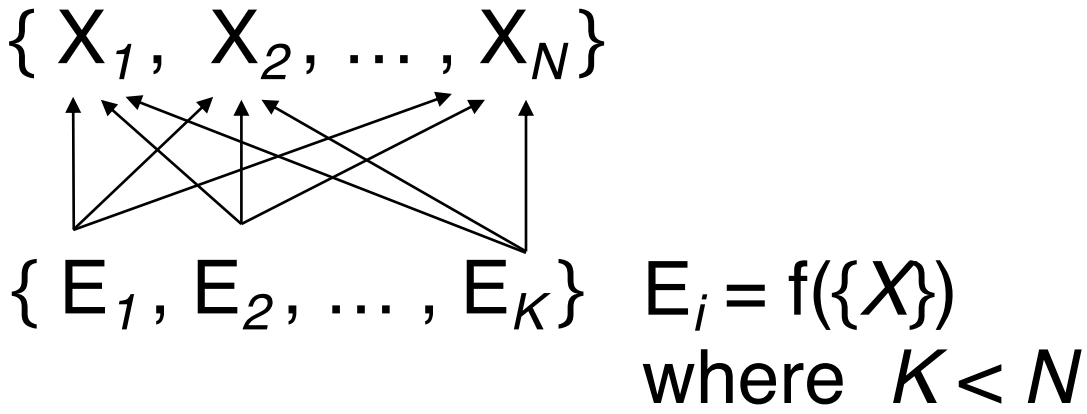
Distribution of pairwise distances between points shown in previous slide

Distances between *M* data points $\boldsymbol{x} \in \mathbb{R}^N$ increase, when *N* increases

Problem: less data density increases uncertainty on underlying data structure

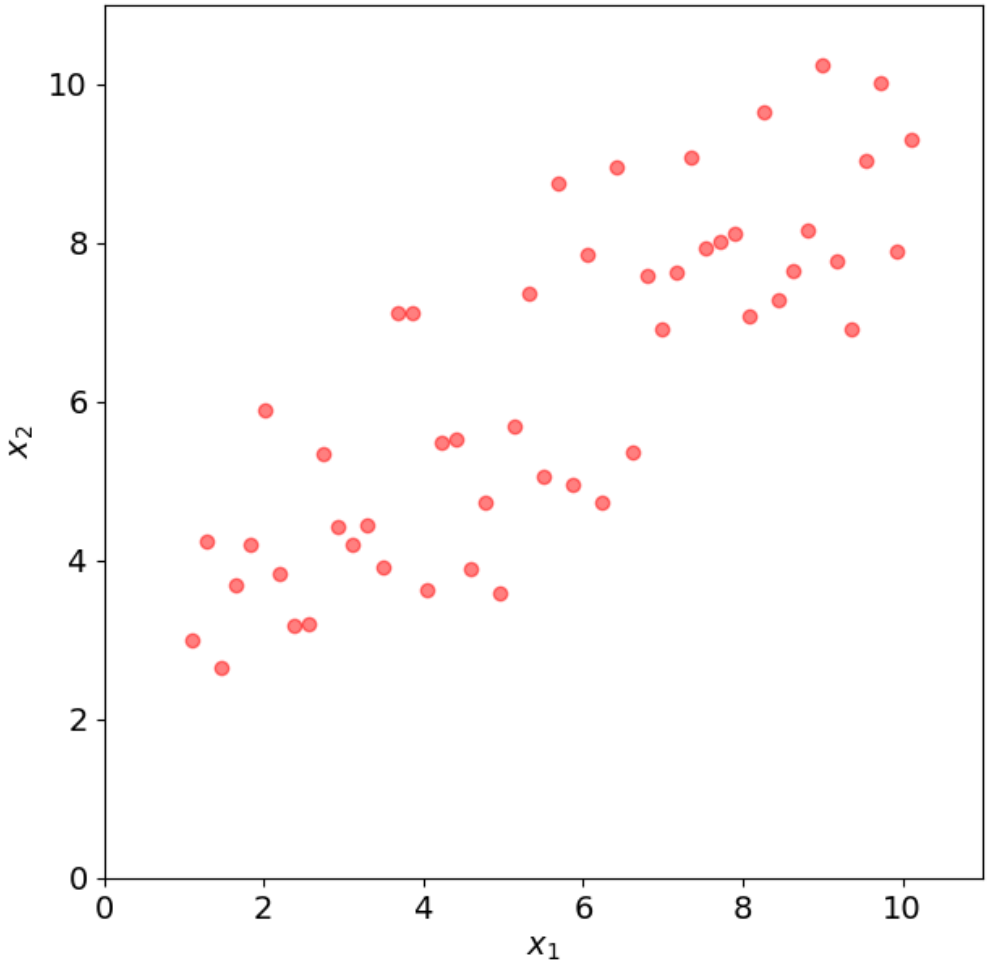
Reducing features increases data density

- Chose appropriate features [expert user]
- Remove features
- Find a lower dimensional representation *E* of features *X*



Principal Components Analysis (PCA)

Let $\boldsymbol{X}$ a dataset of M datapoints in N dimensions (here, M=50 and N=2)



- Center data:

$$\boldsymbol{X}' = \boldsymbol{X} - \mu$$

- Compute data covariance matrix $\boldsymbol{C}$:

$$c_{i,j} = \frac{1}{M}\sum_{k=1}^{M} \boldsymbol{x}'_i \boldsymbol{x}'_j$$

- Calculate eigenvalue decomposition:

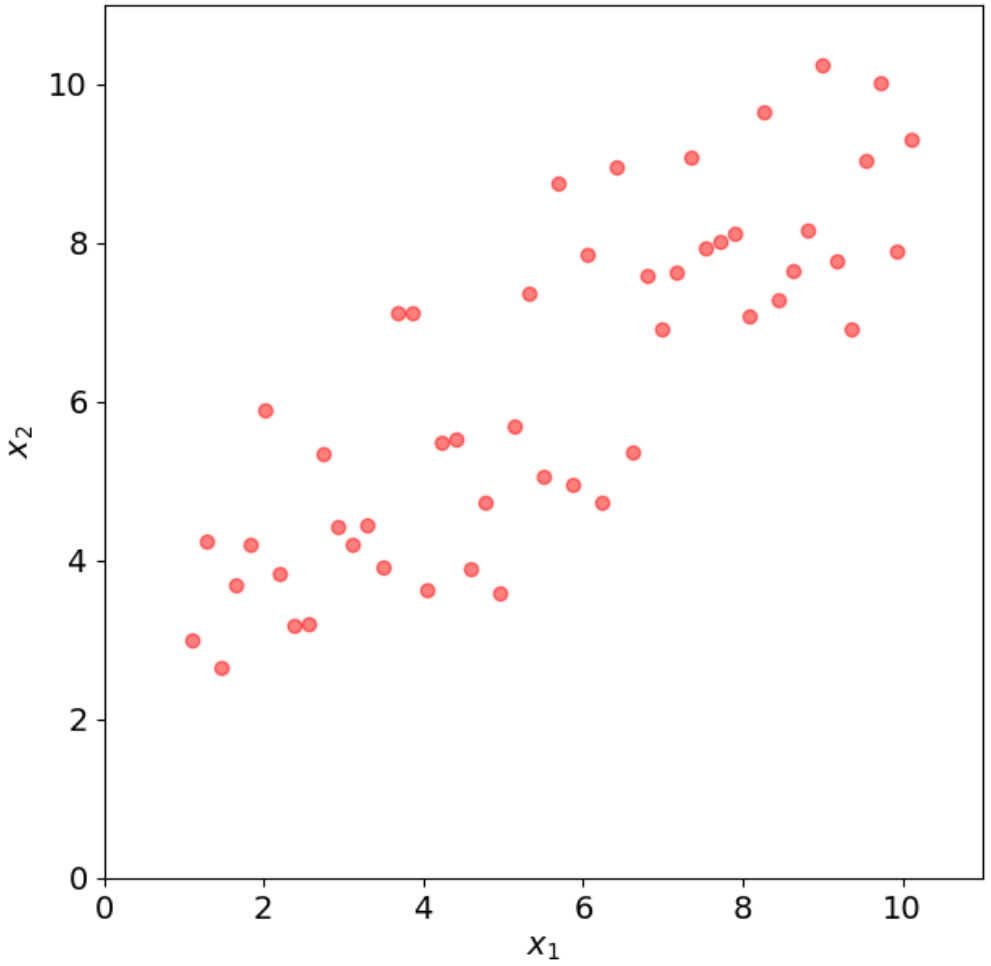
$$\boldsymbol{C} = \boldsymbol{V}\boldsymbol{\lambda}\boldsymbol{V}^{-1}$$

*N*x*N* matrix of ***eigenvectors***

*N*x*N* diagonal matrix of ***eigenvalues***

[Extra] Principal Components Analysis (PCA)

Let $\boldsymbol{X}$ a dataset of M datapoints in N dimensions (here, M=50 and N=2)



An eigenvector $\boldsymbol{v}$ of $\boldsymbol{C}$ respects:

$$\boldsymbol{C}\boldsymbol{v} = \boldsymbol{\lambda}\boldsymbol{v}$$

Find eigenvalues as the roots of the characteristic polynomial:

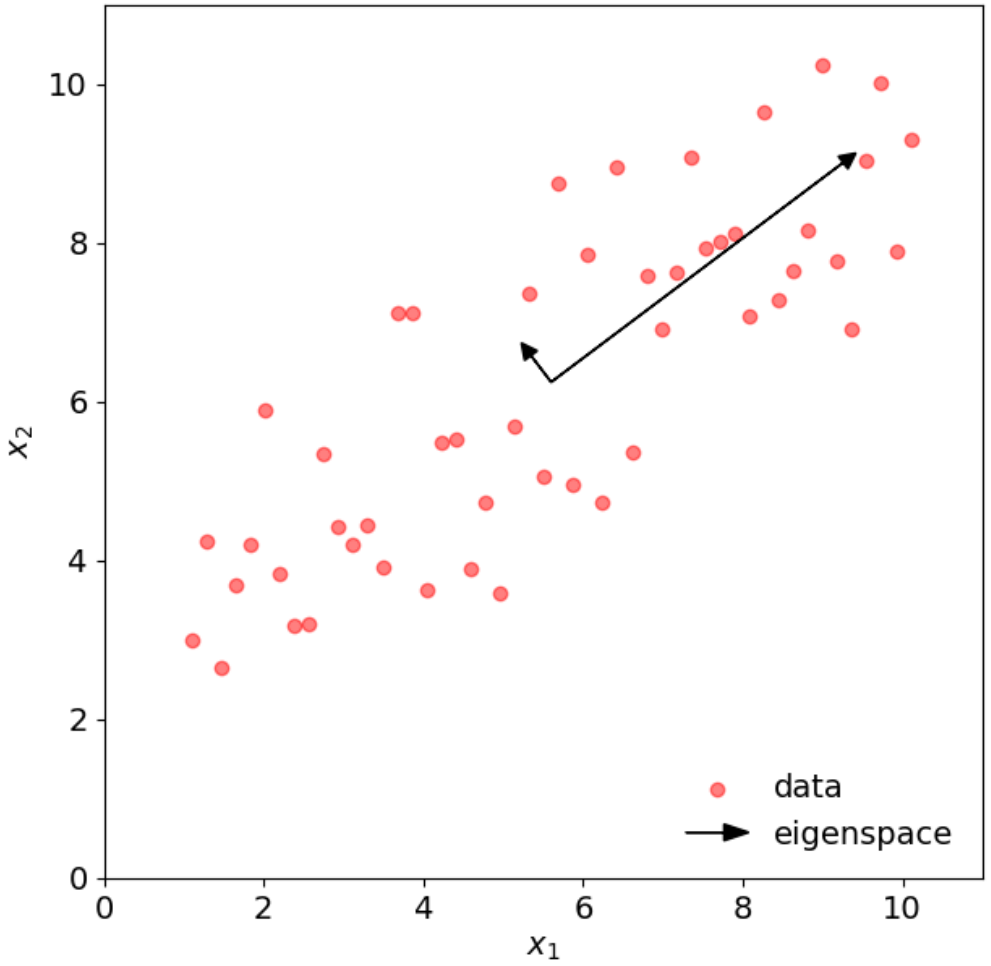
$$p(\boldsymbol{\lambda}) = \det(\boldsymbol{C} - \boldsymbol{\lambda}\boldsymbol{I}) = 0$$

The i-th eigenvector $\boldsymbol{v_i}$ is found by solving:

$$\boldsymbol{C}\boldsymbol{v_i} = \boldsymbol{\lambda_i}\boldsymbol{v_i}$$

Principal Components Analysis (PCA)

Let $\boldsymbol{X}$ a dataset of M datapoints in N dimensions (here, M=50 and N=2)



$$\boldsymbol{C = V\lambda V^{-1}}$$

$\boldsymbol{V = [v_1 \ldots v_N]}$ eigenvectors: orthonormal base

$\boldsymbol{\lambda}$ eigenvalues: scalars defining the importance of each eigenvector

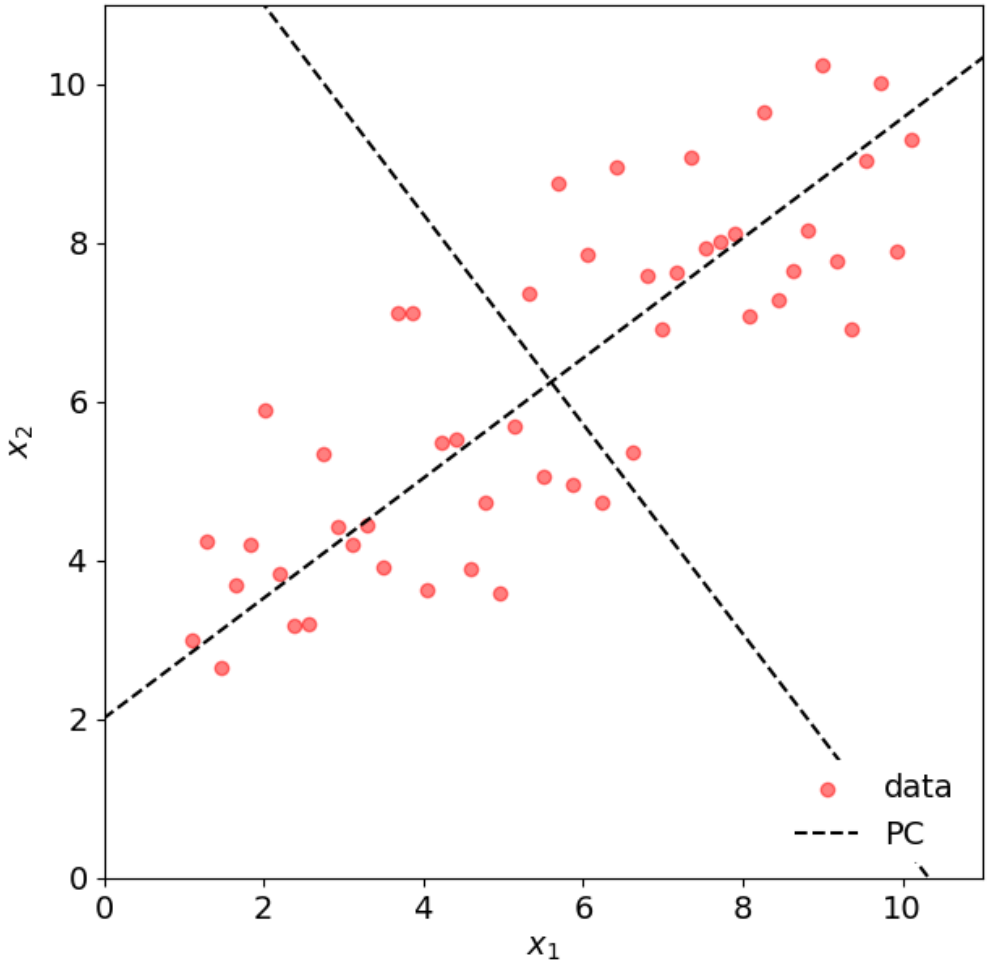
Importance $\boldsymbol{r_i}$ of each eigenvector $\boldsymbol{v_i}$:

$$\boldsymbol{r_i = \frac{\lambda_i}{\sum \lambda}}$$

Sort $\boldsymbol{V}$ *and* $\boldsymbol{\lambda}$ *according to* $\boldsymbol{\lambda}$

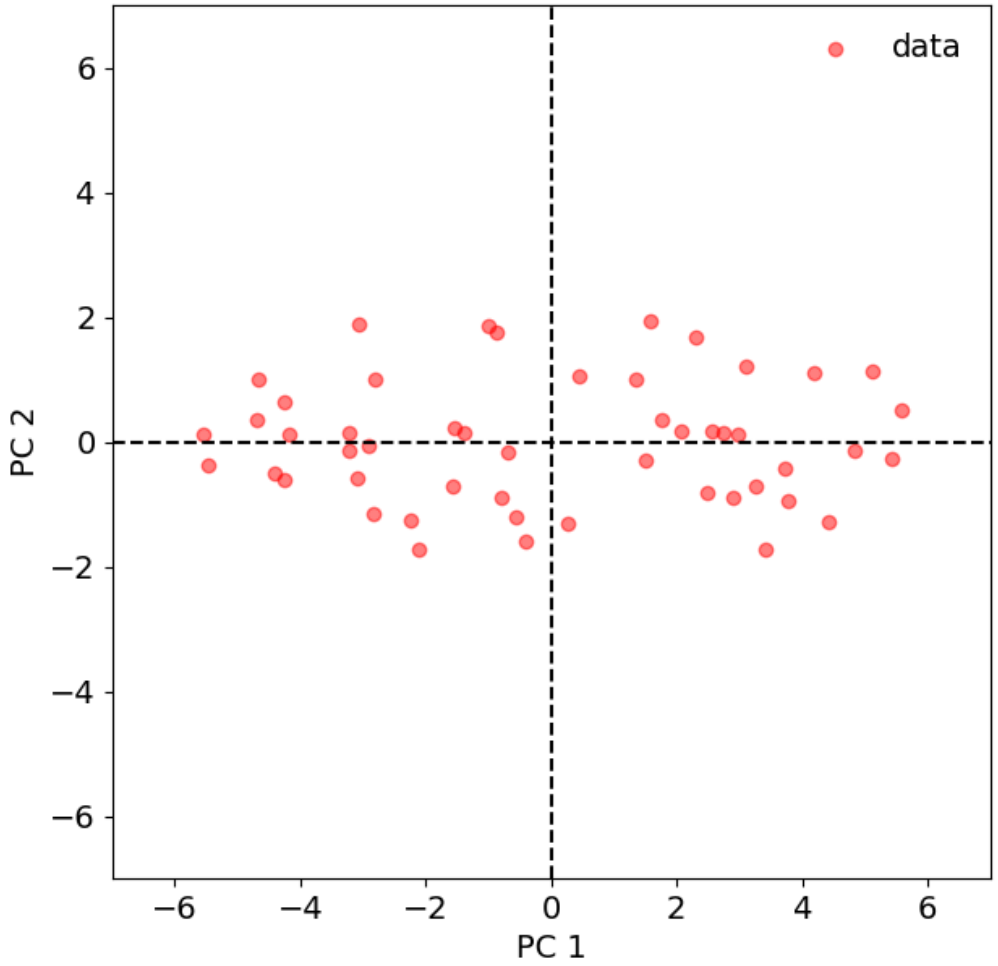
Projection into the eigenspace

Let $\boldsymbol{X}$ a dataset of M datapoints in N dimensions (here, M=50 and N=2)



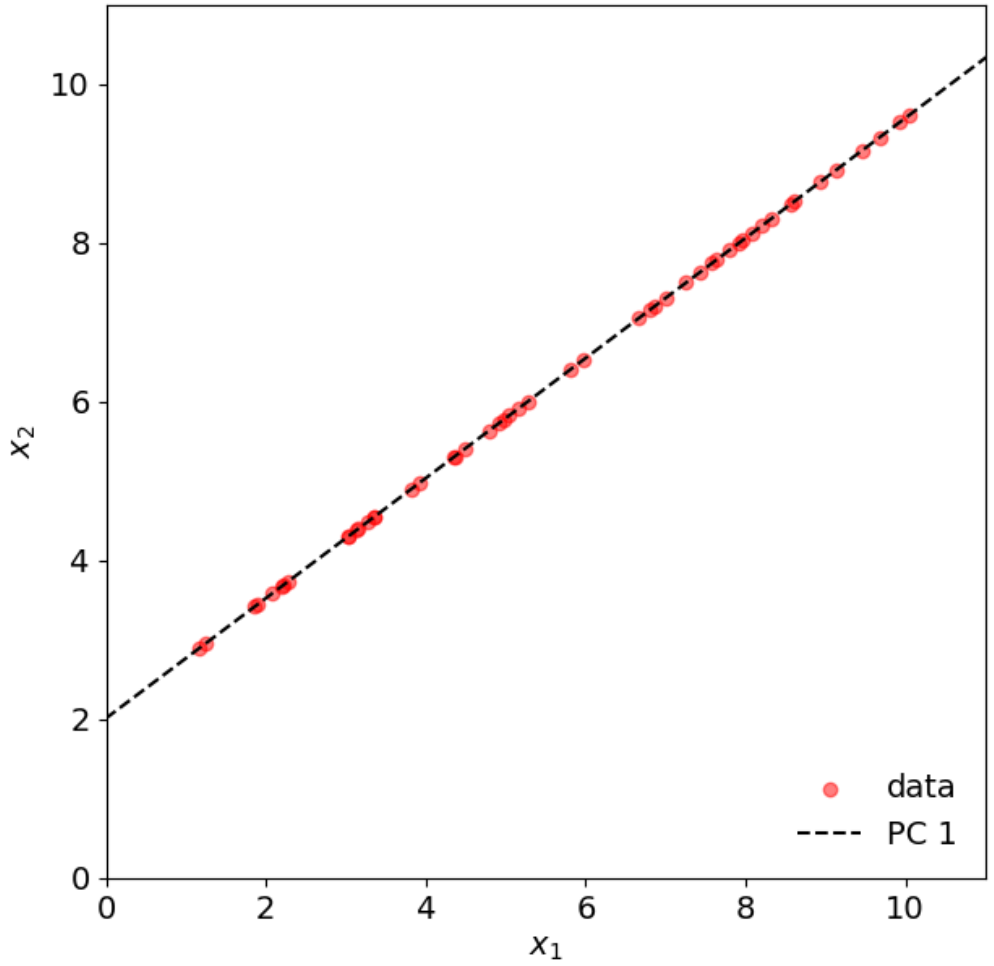
$$p = \boldsymbol{V}^T(\boldsymbol{x} - \mu)$$

transform



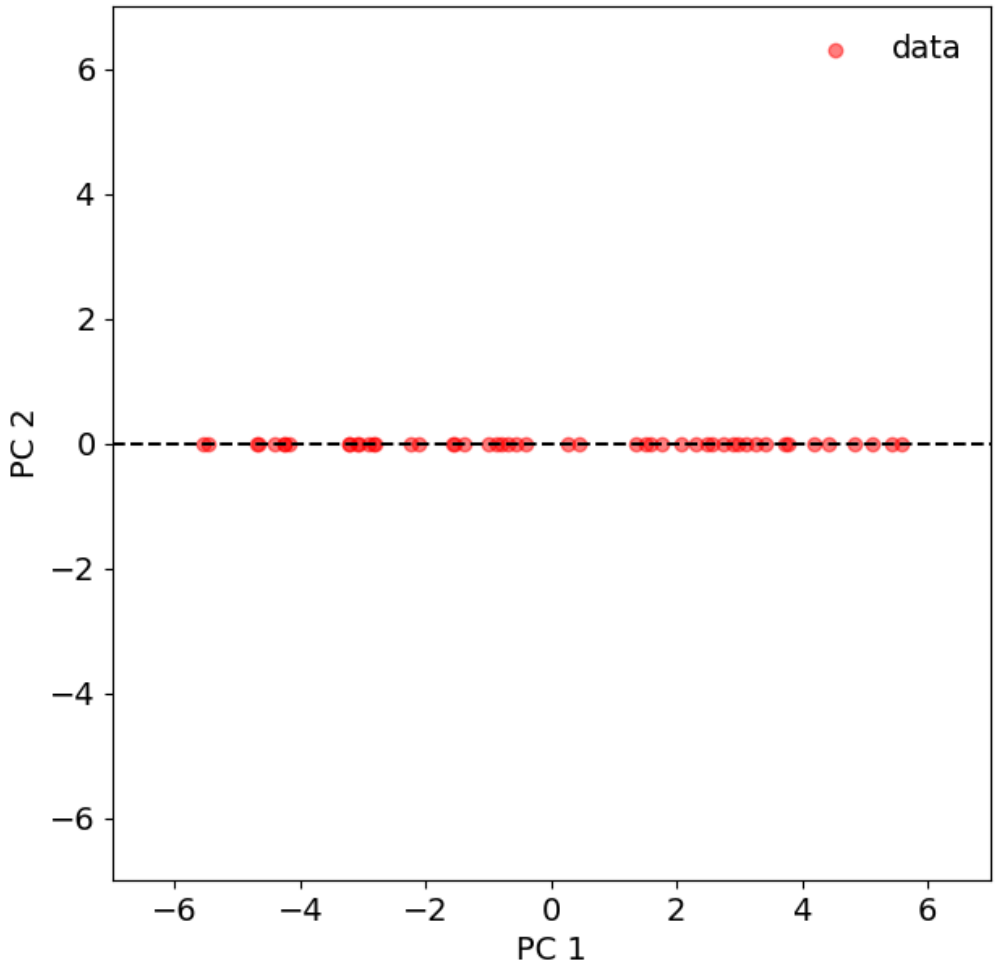
Dimensionality reduction

Remove dimensions that least contribute to data variance



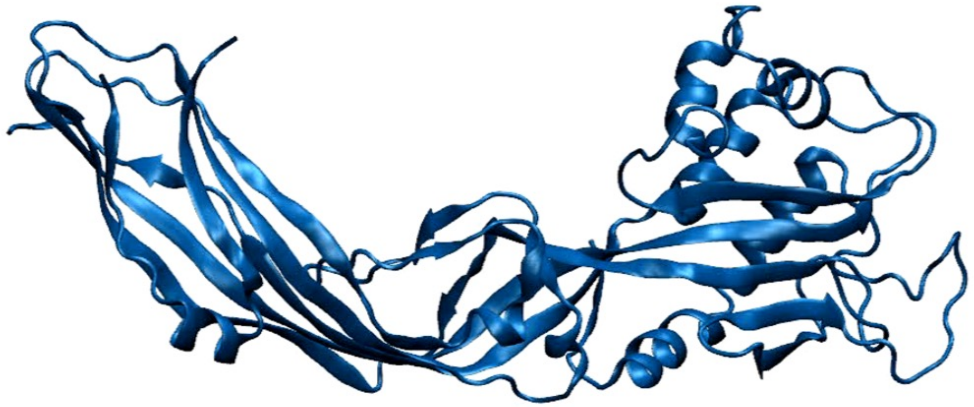
$$p = \boldsymbol{V}^T(\boldsymbol{x} - \mu)$$

transform

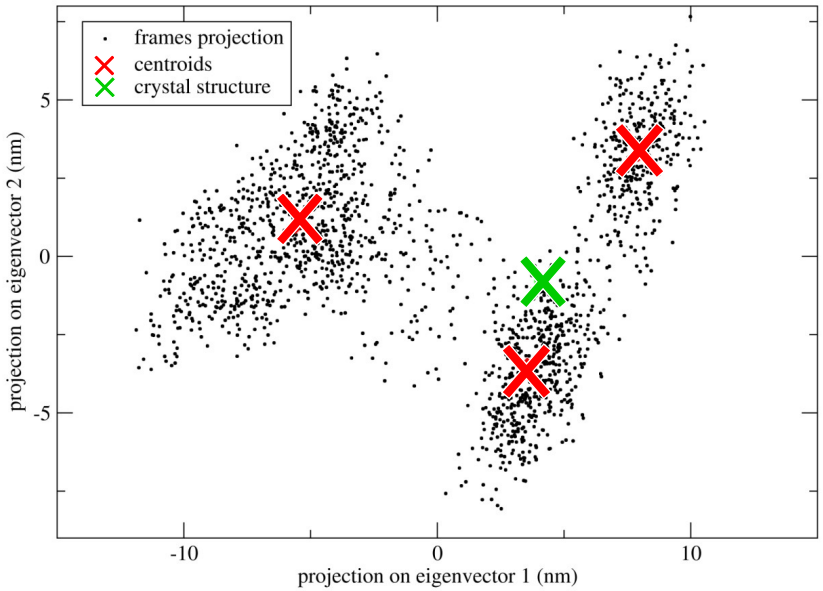


[Example 2] Identifying dominant motions in proteins

Protein MD simulation



Eigenspace of C_α coordinates

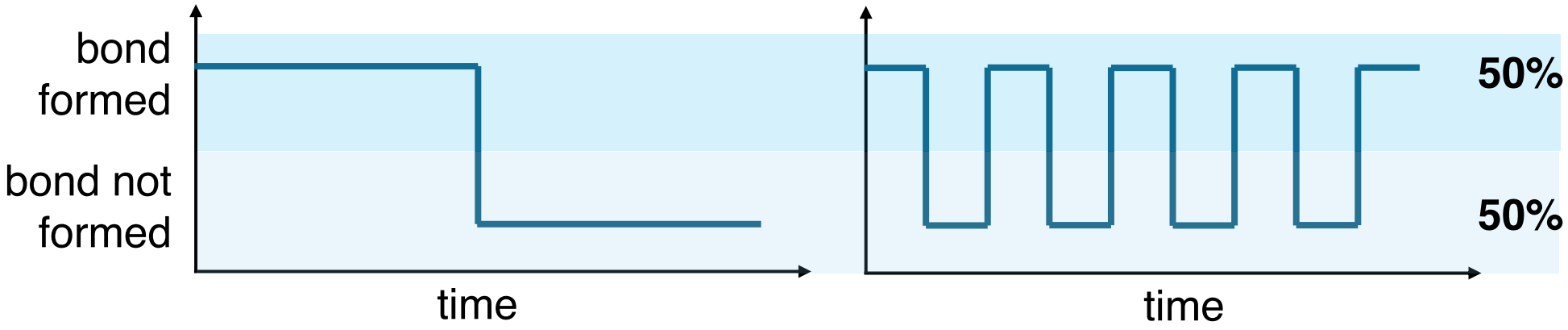


- Simulations are complex and noisy
- The first PCs capture large-scale collective motions, last ones capture noise

Degiacomi, M.T. & Dal Peraro, M.. *Macromolecular symmetric assembly prediction using swarm intelligence dynamic modeling*. ***Structure***, 2013

largest-scale motion ≠ slowest motion

Thought experiment reminder: typically hydrogen bond is considered established if donor-acceptor distance <2.5 Å, and donor-acceptor-hydrogen angle <20°.



- Biological function is often determined by the kinetics of a process (e.g., a specific conformational change).
- Two processes with different kinetics can have the same statistical distribution in terms of structure.
- PCA separates in terms of structure, not timescales!