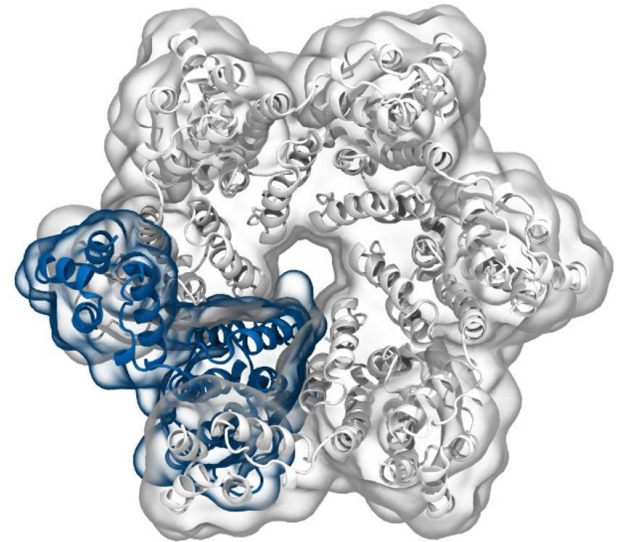
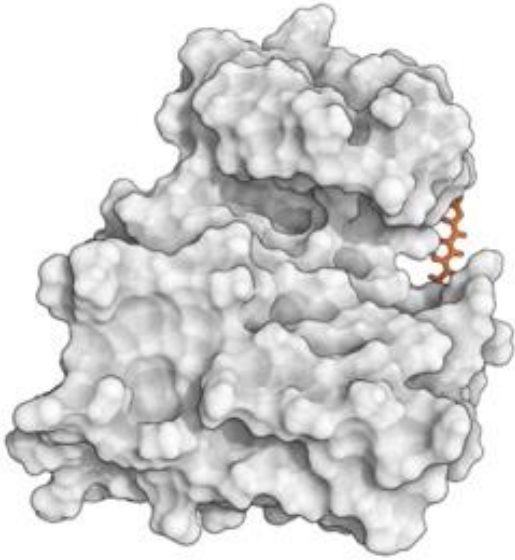


# Simulation of Biomolecules

## Basic Simulation Analysis

**2024 CCP5 Summer School**



Dr Matteo Degiacomi  
Durham University

[matteo.t.degiacomini@durham.ac.uk](mailto:matteo.t.degiacomini@durham.ac.uk)

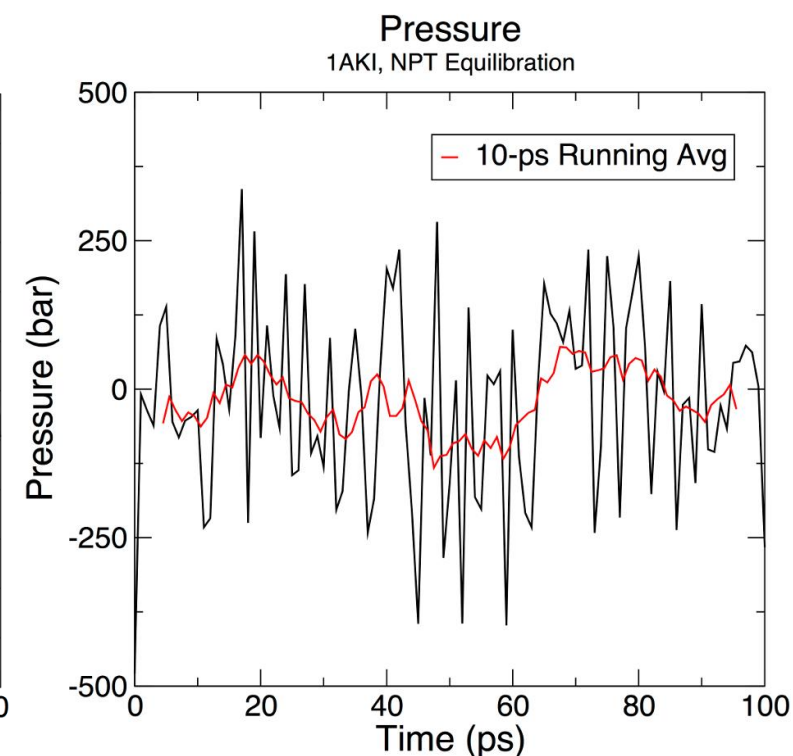
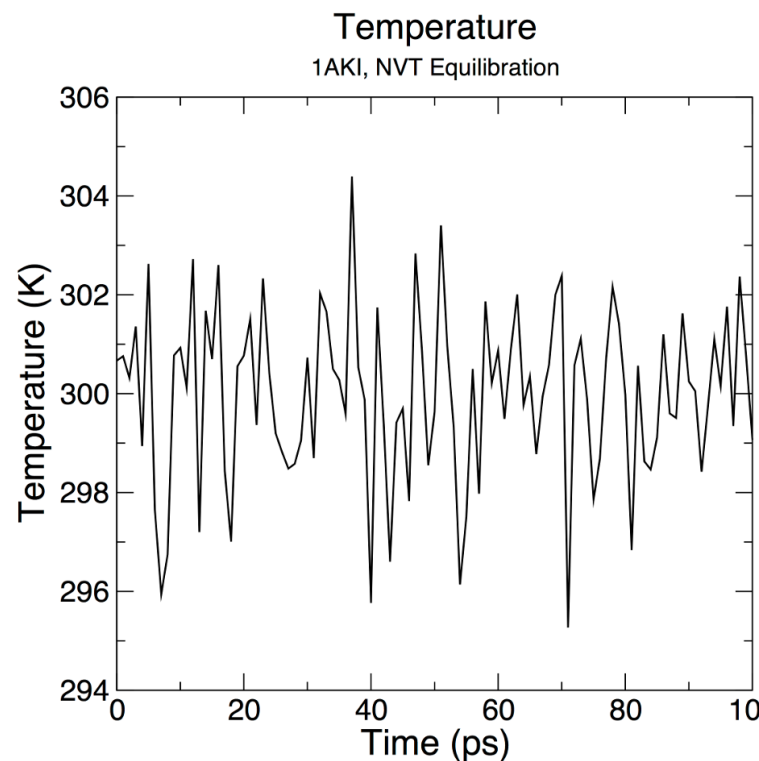
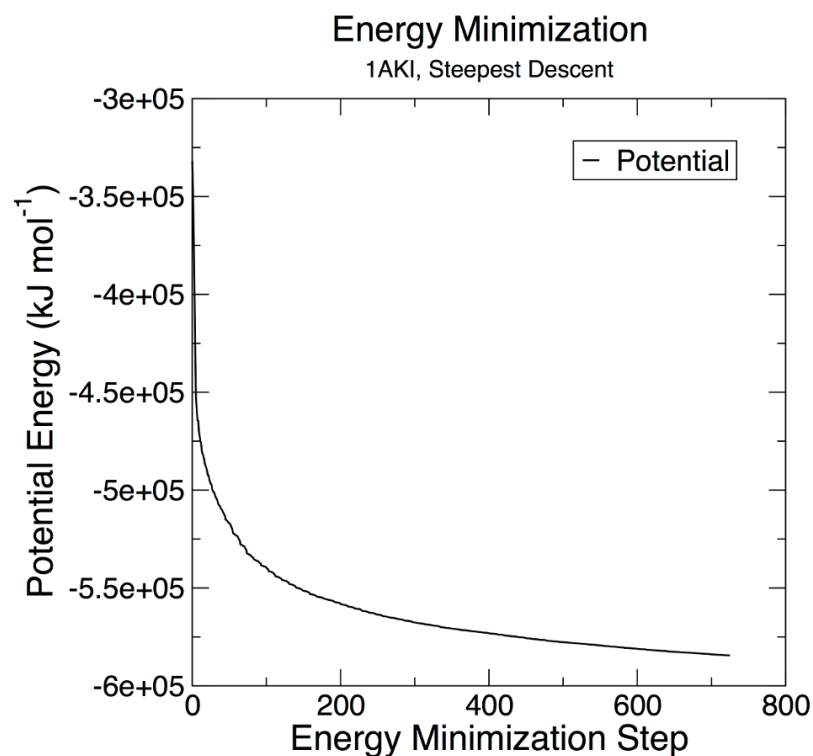
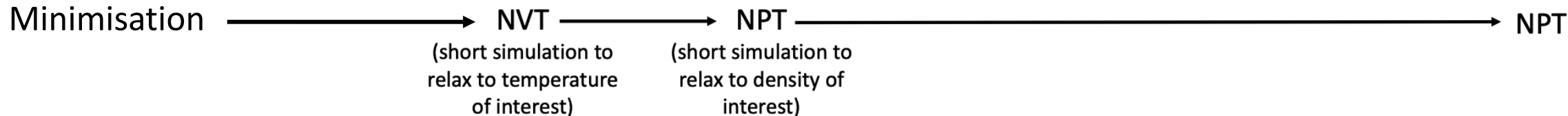
Dr Antonia Mey  
University of Edinburgh

[antonia.mey@ed.ac.uk](mailto:antonia.mey@ed.ac.uk)

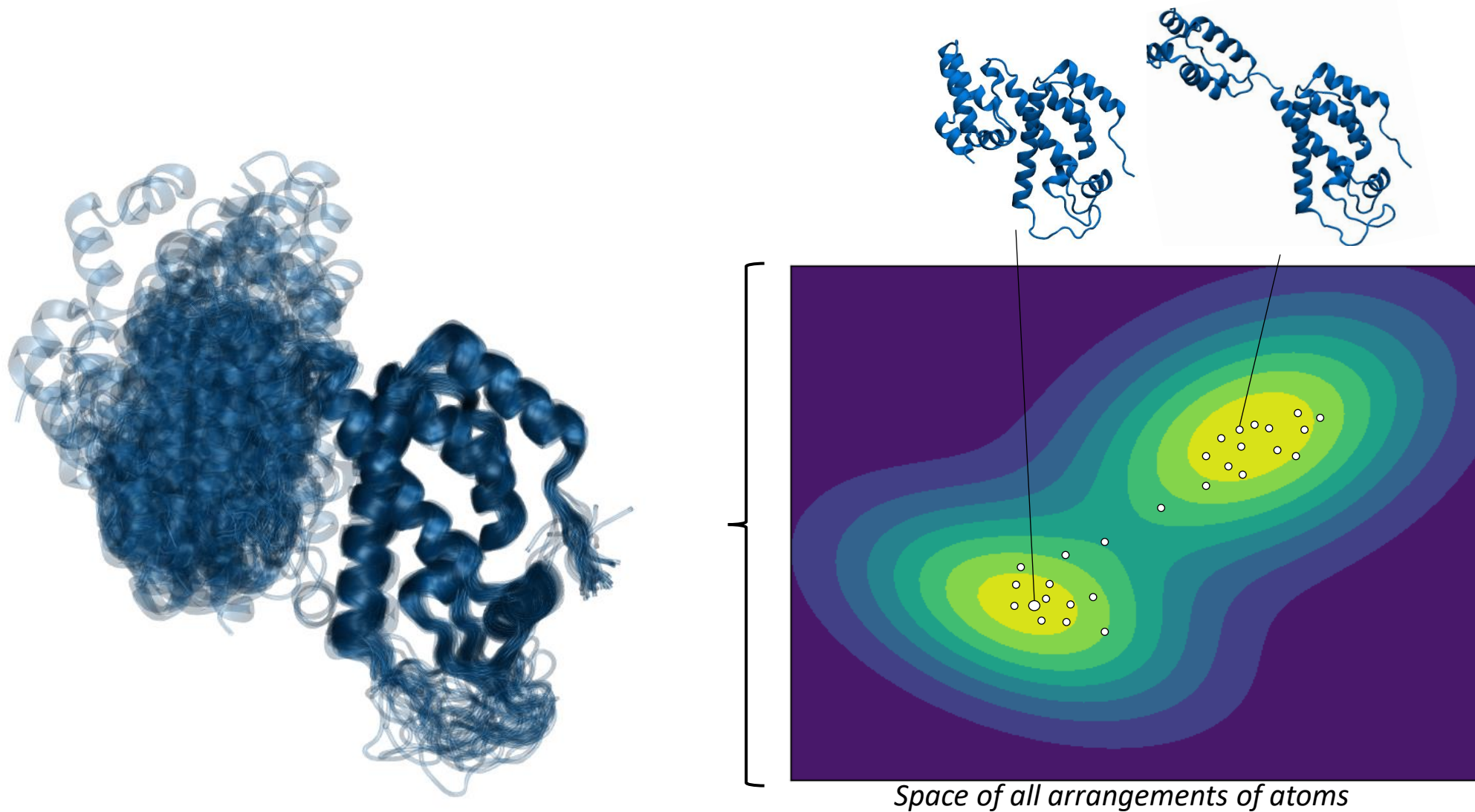
# Volume and pressure equilibration

Steps until production:

Production ensemble



# Sampling the conformational space

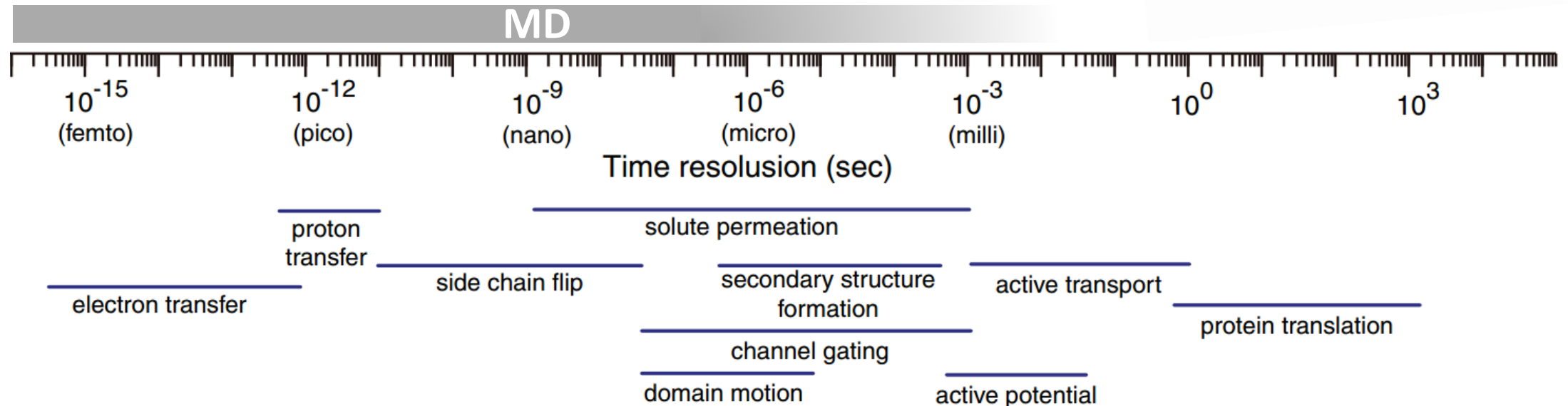
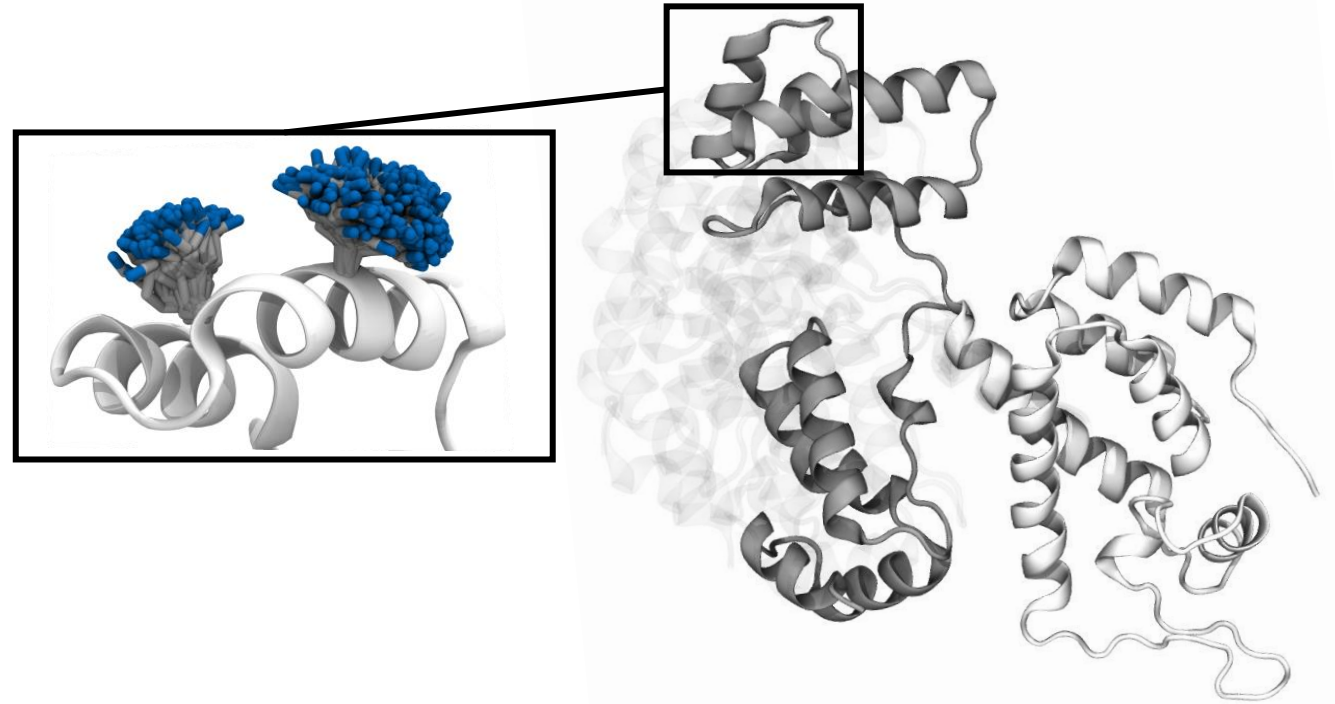


Probability of sampling a conformation is inversely proportional to its energy:  $p_i \propto e^{-\epsilon_i/kT}$

# 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

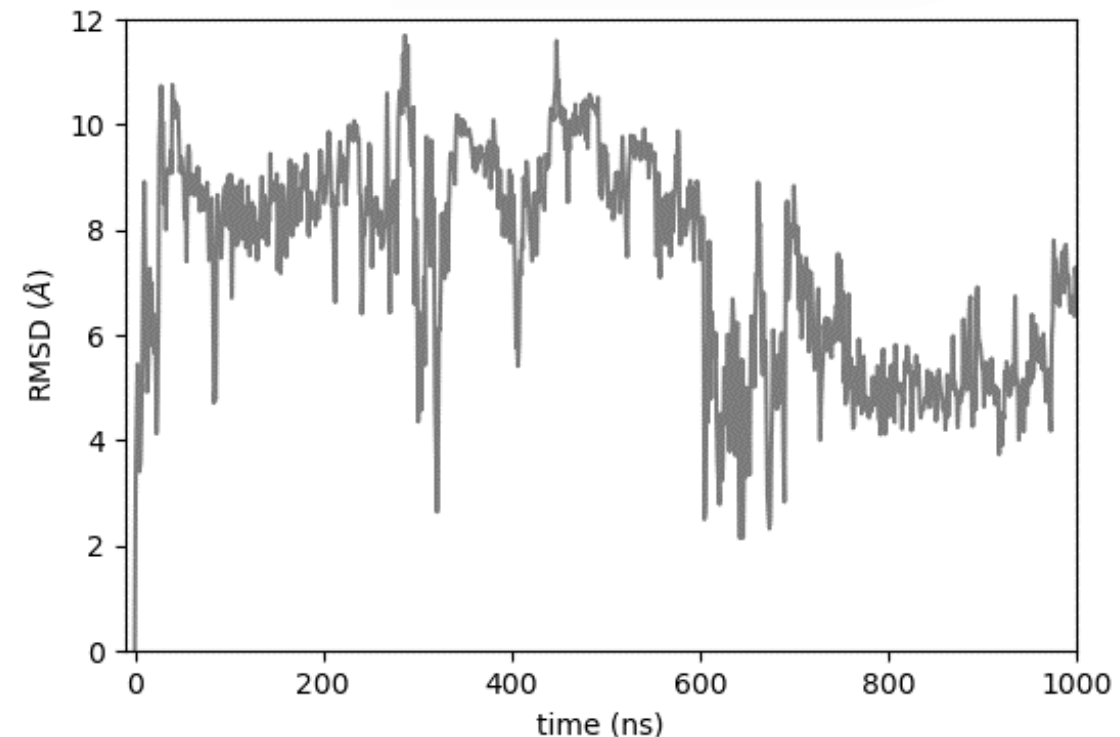
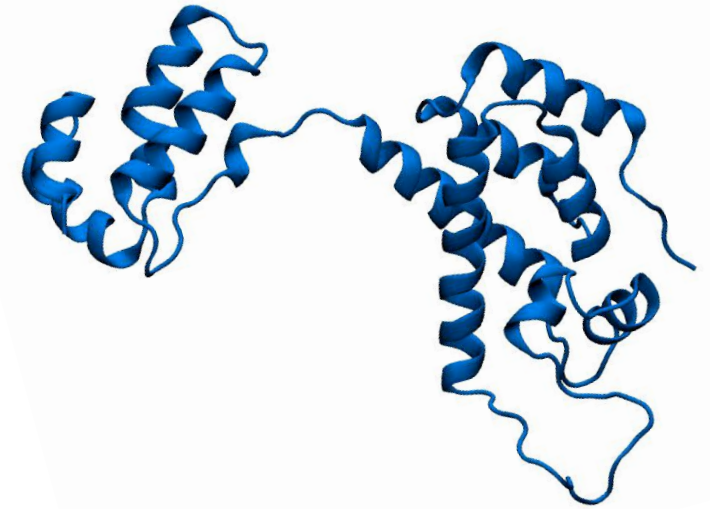


# Root Mean Square Deviation (RMSD)

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

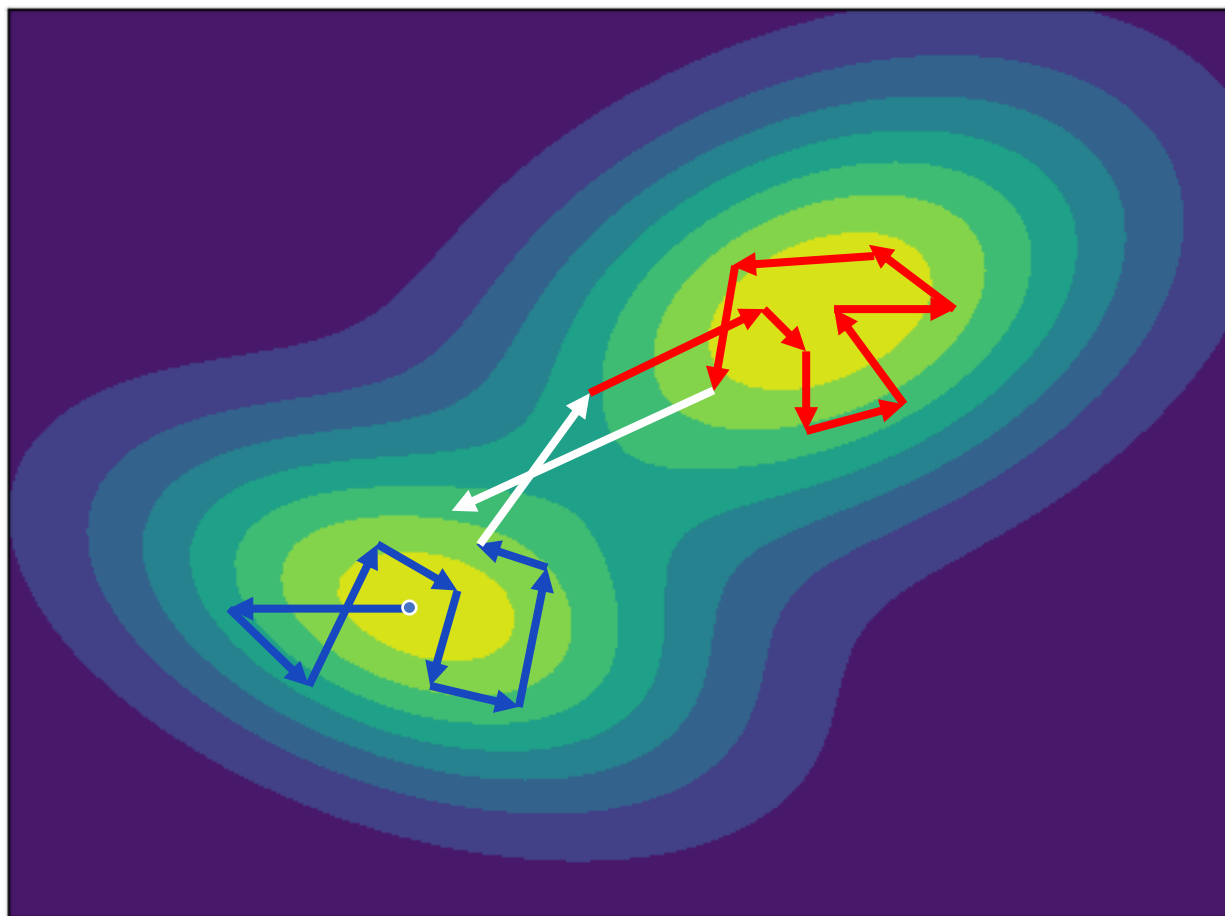
$$RMSD = \sqrt{\frac{1}{N} \sum_{i=0}^N (X_i - x_0)^2}$$

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

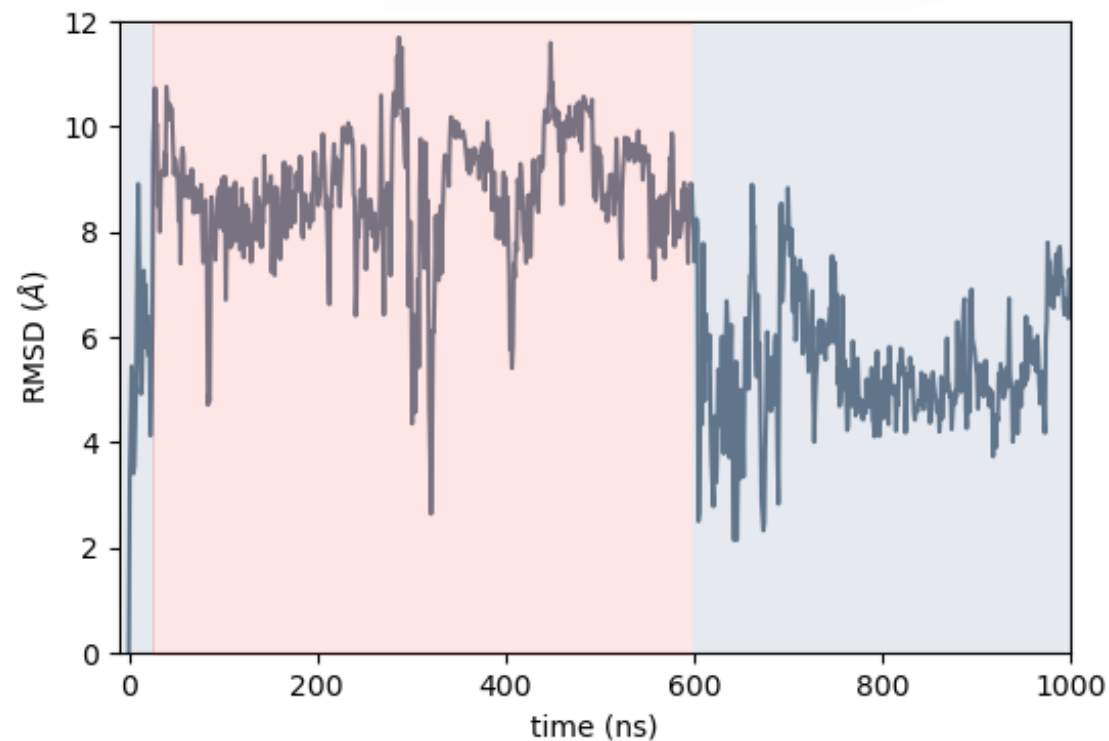
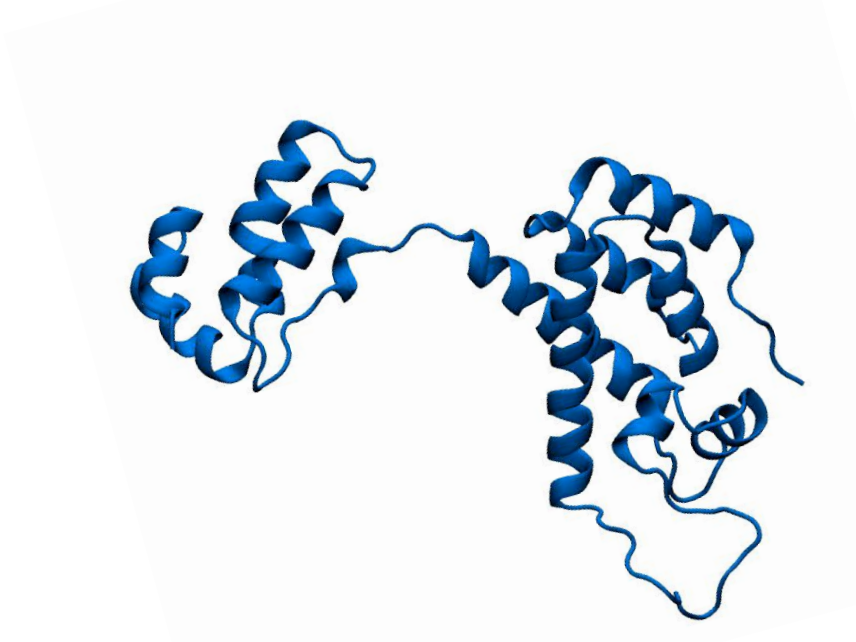




# Convergence?

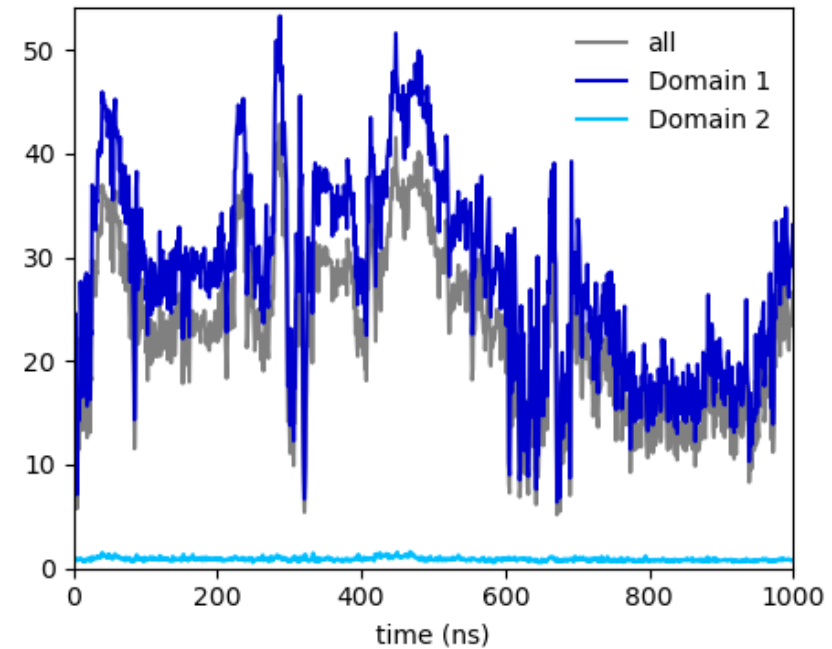
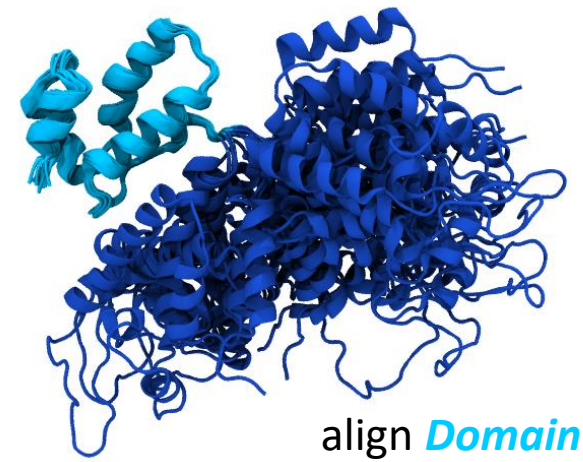
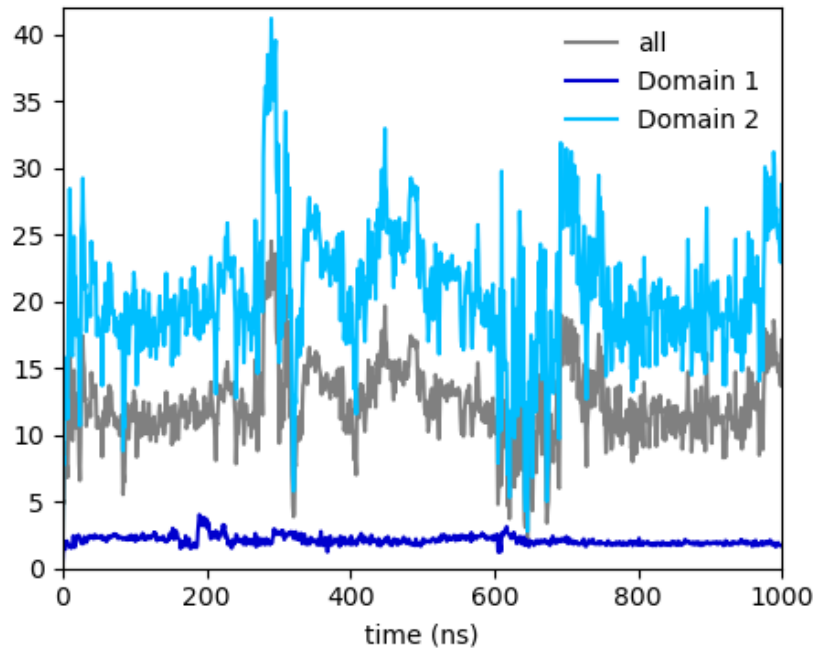
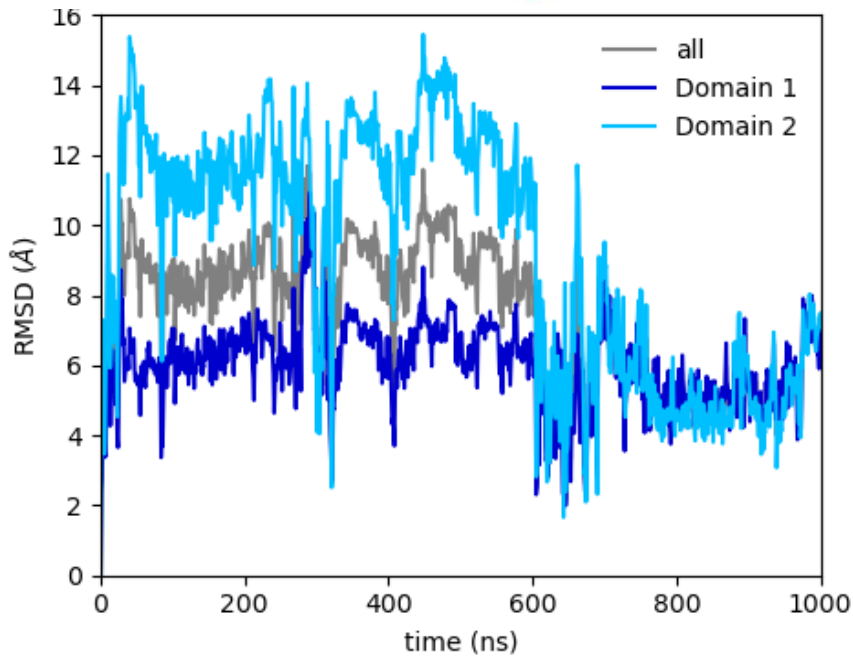
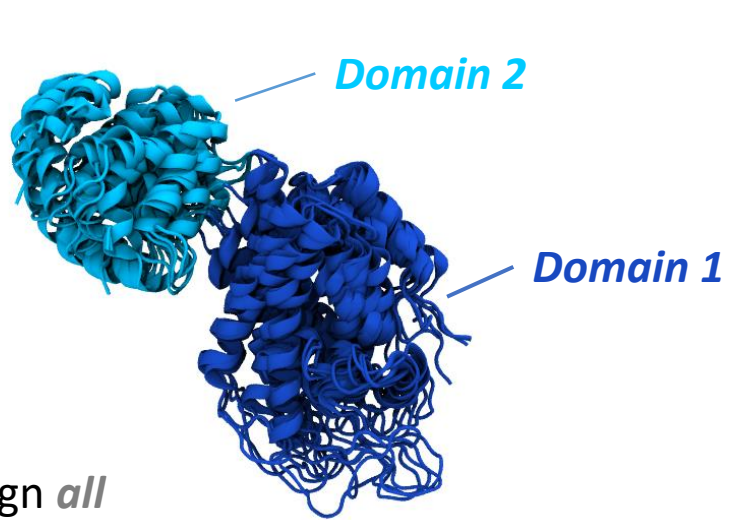


*Refrain from using RMSD as a single indicator of simulation convergence.*



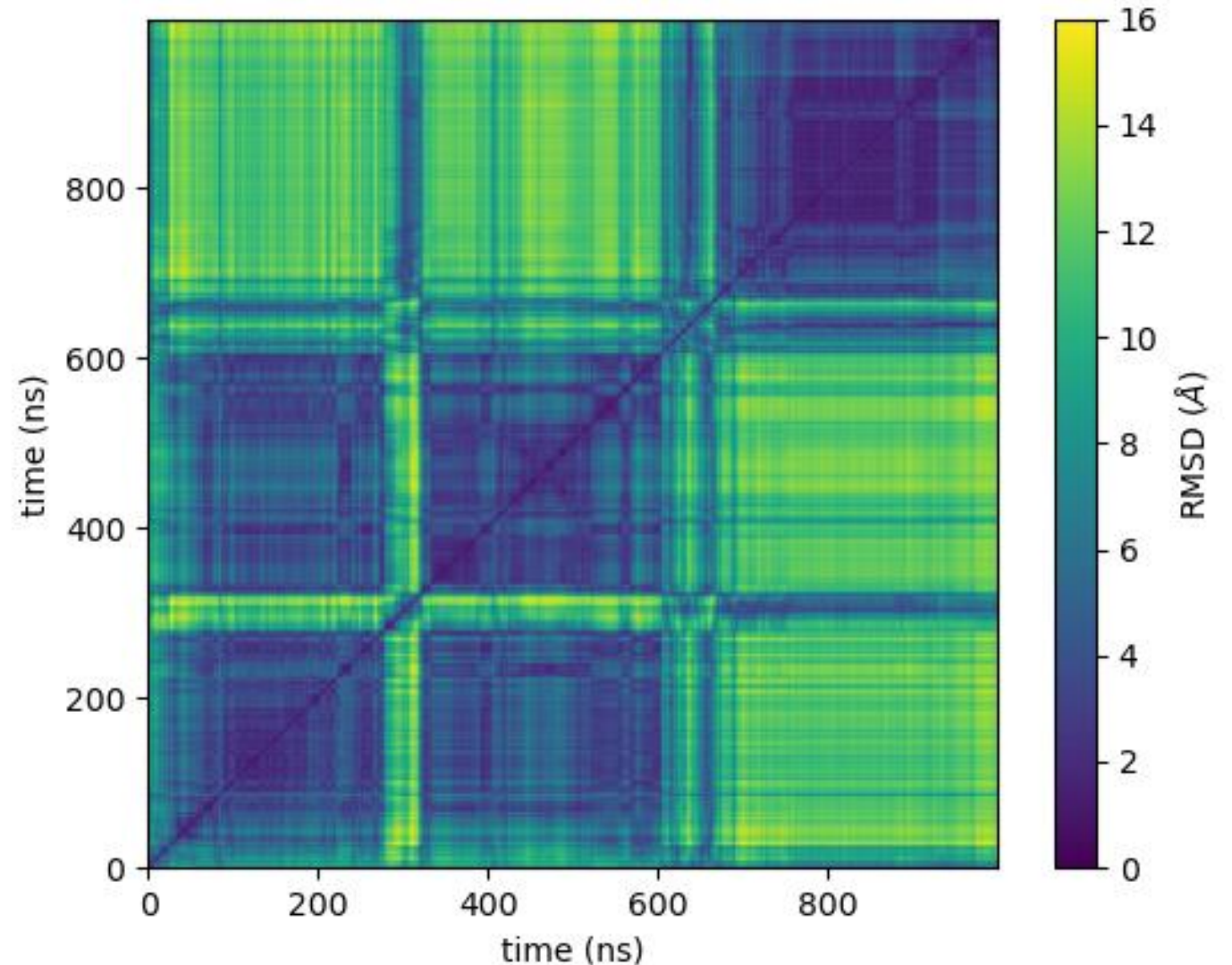
HIV-1 capsomer simulation from: *Degiacomi & Dal Peraro, Structure, 2013*

# RMSD is alignment- and selection-dependent



# 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.

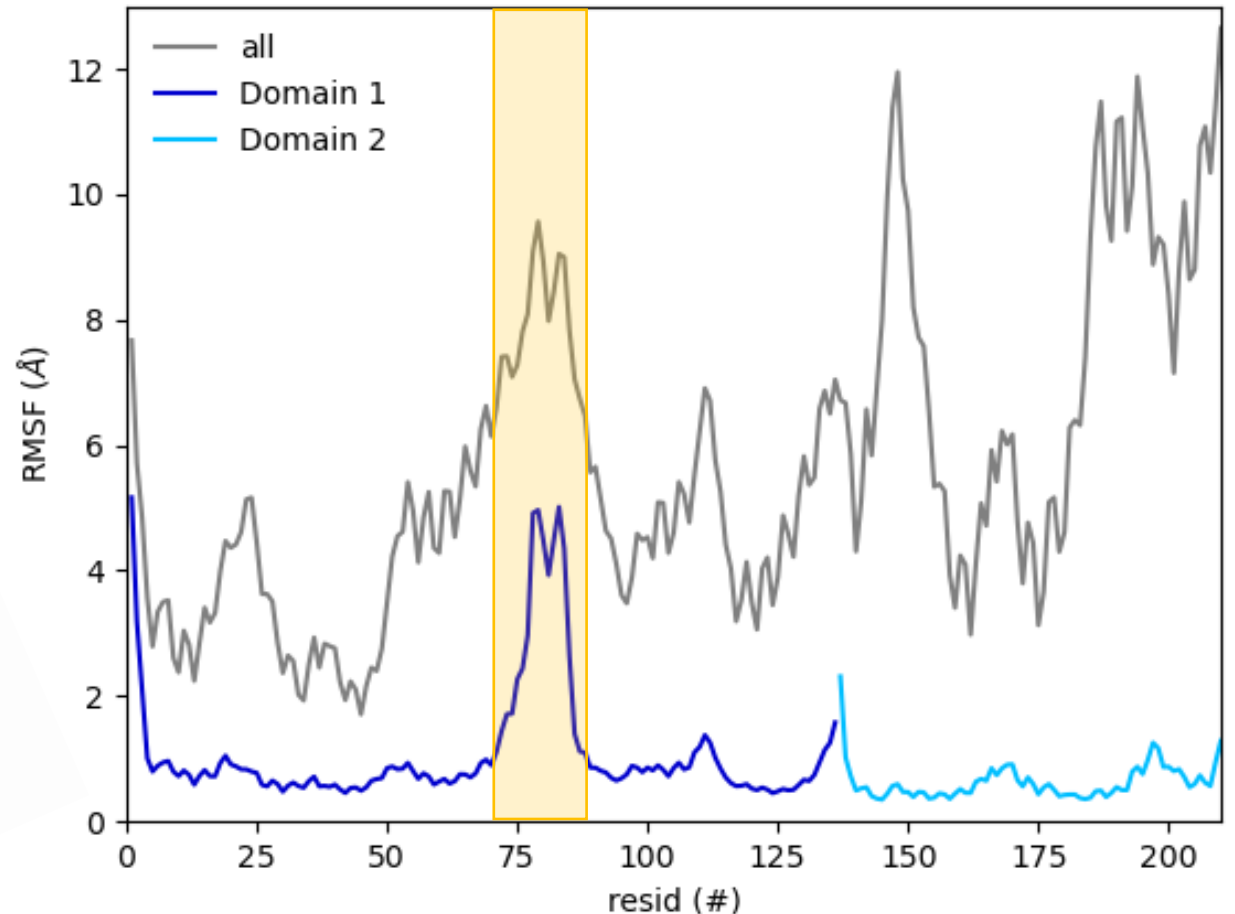
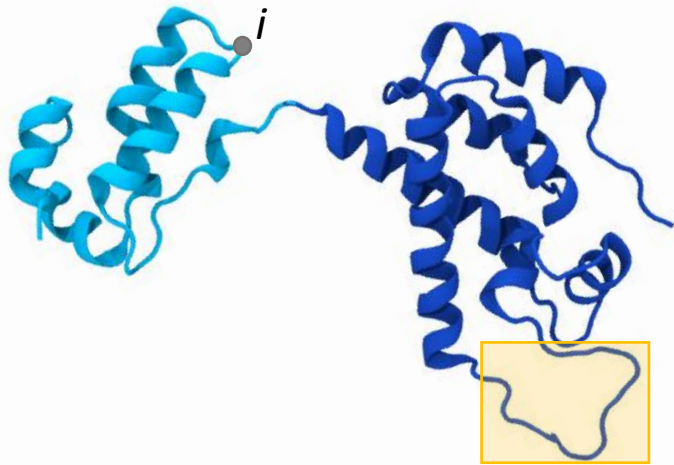




# Root Mean Square Fluctuation (RMSF)

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

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

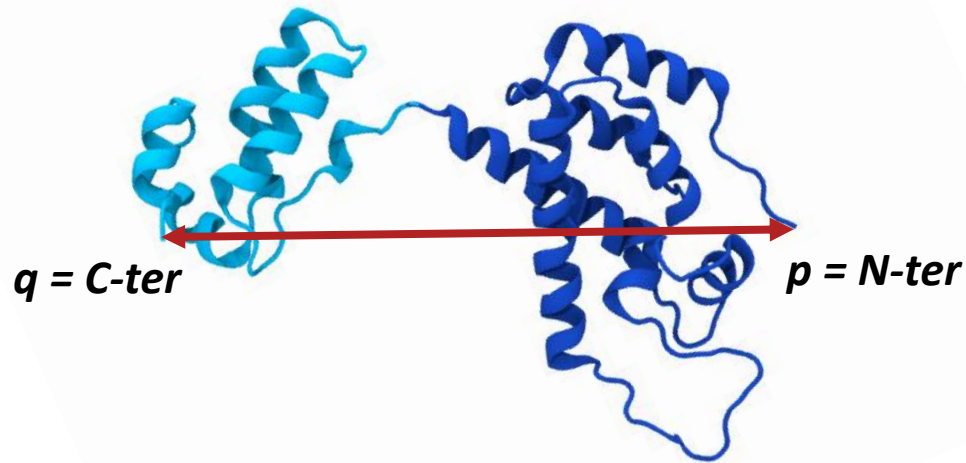


Typically done on  $C_\alpha$  atoms.  
Result depends on alignment!

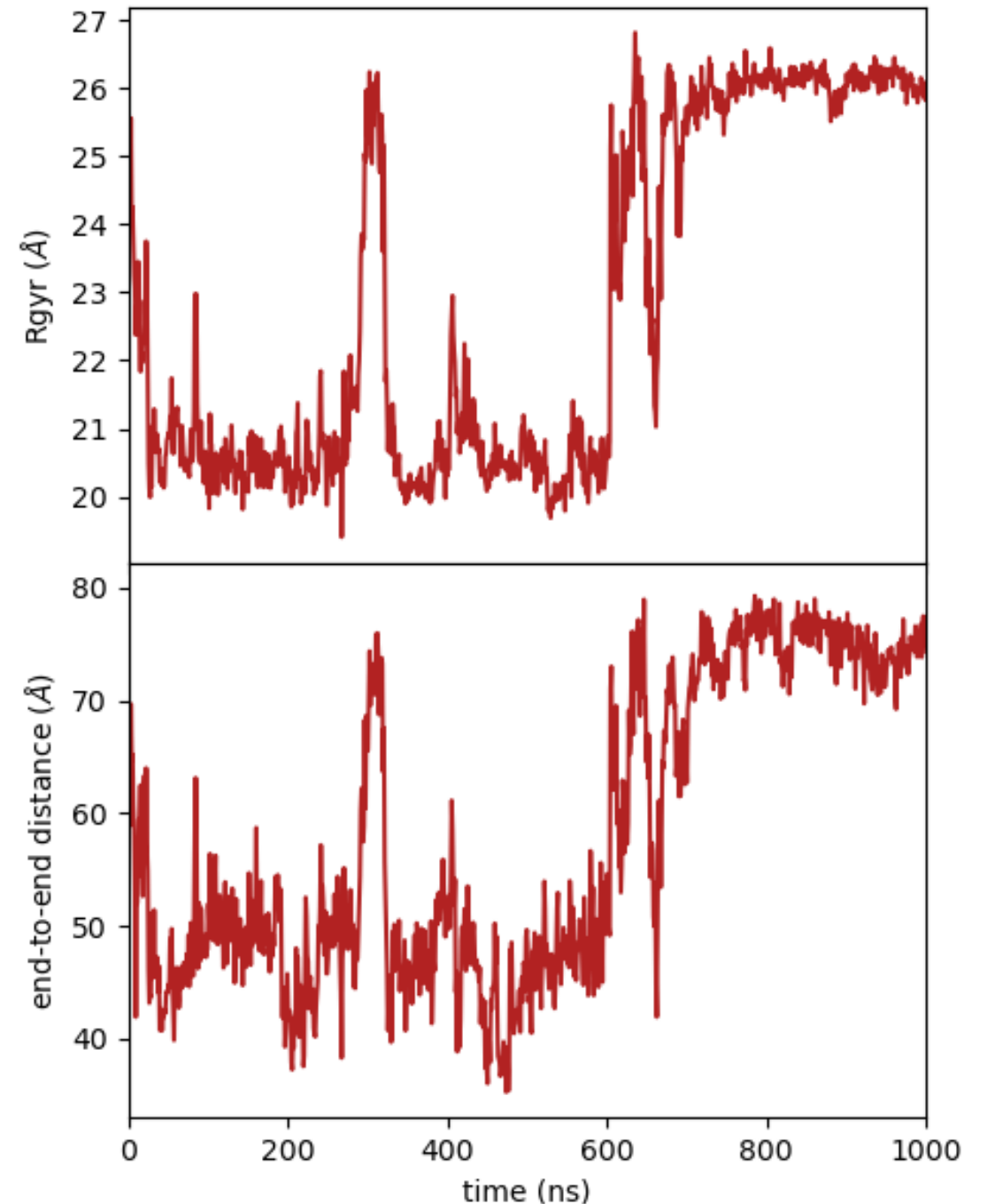
# end-to-end distance and Radius of Gyration (Rg)

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

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



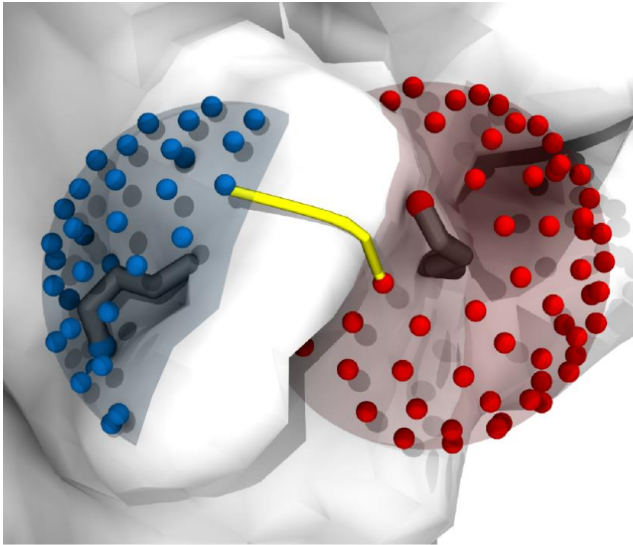
Internal properties do *not* depend on alignment



# Some quantities are not *directly* measurable

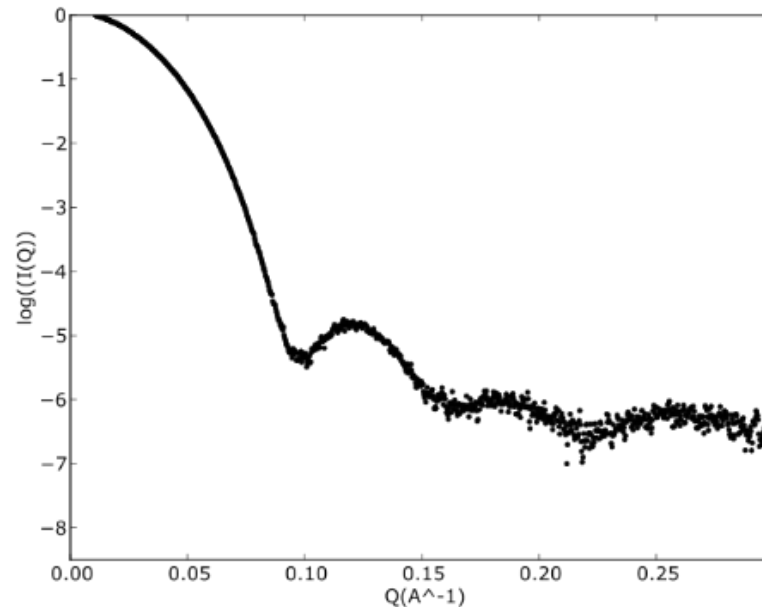
Submit MD conformers to external software simulating experimental data, e.g.:

## Chemical cross-linking



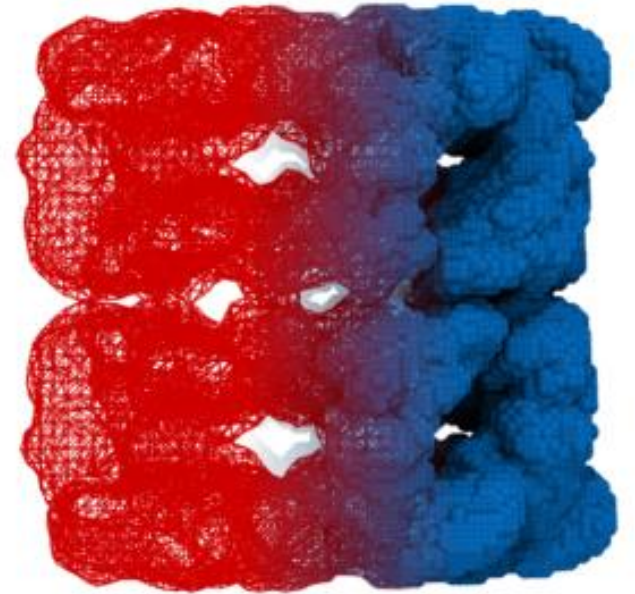
DynamXL, Xwalk, ...

## SAXS



CRY SOL, FoXS, ...

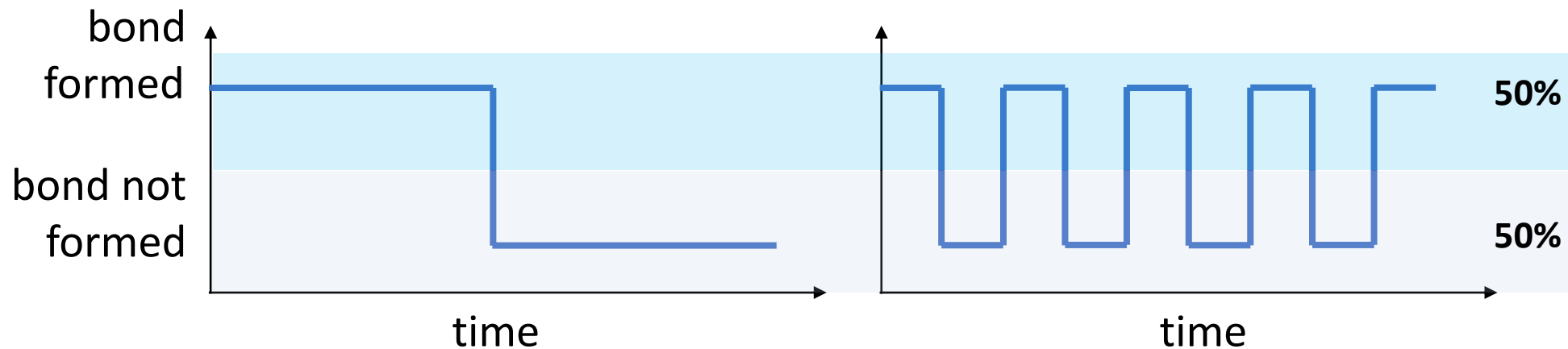
## Collision cross-section



IMPACT, MobCal,  
CollisionCode, ...

# Warning: time averaging may hide processes

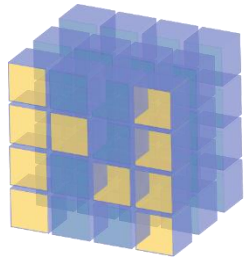
**Thought experiment:** typically hydrogen bond is considered established if donor-acceptor distance  $< 2.5 \text{ \AA}$ , and donor-acceptor-hydrogen angle  $< 20^\circ$ .



Reporting % time a bond is established in simulation can be misleading!



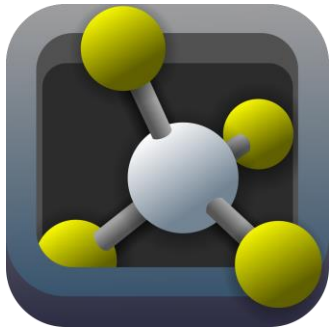
# Large ecosystem of software and packages for data analysis



NumPy



**SciPy**



Open-Source Cheminformatics  
and Machine Learning

VMD

Visual Molecular Dynamics



**MDTRAJ**



**PyEMMA**

And many more...

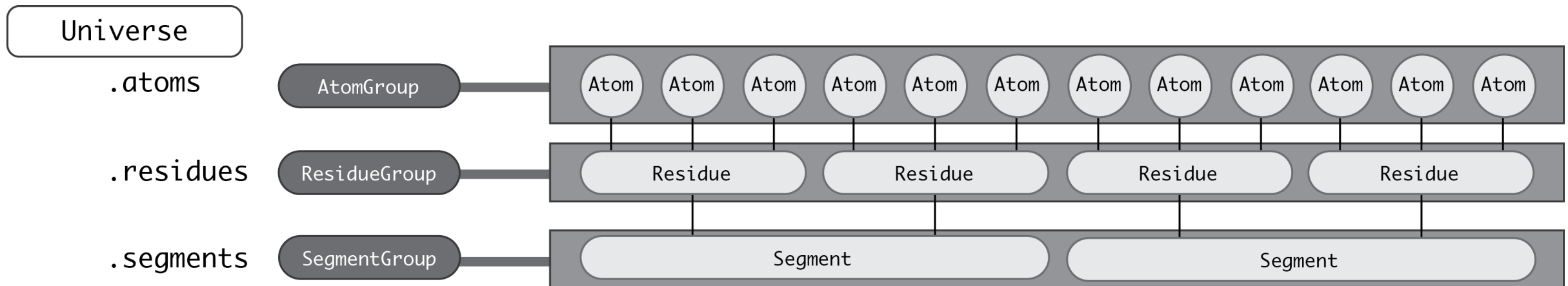
# MDAnalysis

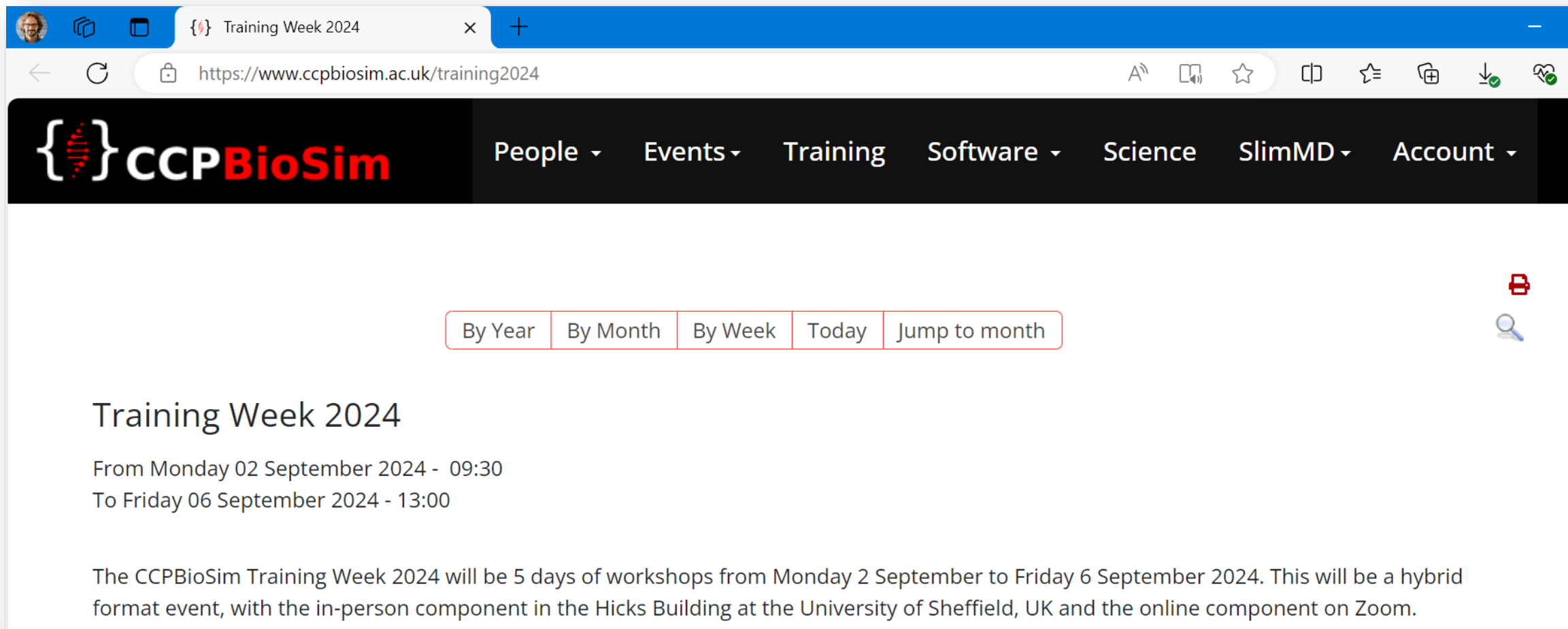


The **Universe** contains everything about a MD system

- *Static information*: atoms and their connectivities
- *Dynamic information*: The trajectory

Data accessible via a hierarchy of containers





The screenshot shows a web browser window with a blue header bar. The address bar displays the URL <https://www.ccpbiosim.ac.uk/training2024>. The website's navigation bar is black with white text for the CCPBioSim logo and menu items: People, Events, Training, Software, Science, SlimMD, and Account. The main content area has a white background. At the top right, there are icons for a calendar, a magnifying glass, and a red printer icon. Below these is a filter bar with buttons: By Year, By Month, By Week, Today, and Jump to month. The main heading is "Training Week 2024". Below it, the dates are listed: "From Monday 02 September 2024 - 09:30" and "To Friday 06 September 2024 - 13:00". A paragraph of text describes the event: "The CCPBioSim Training Week 2024 will be 5 days of workshops from Monday 2 September to Friday 6 September 2024. This will be a hybrid format event, with the in-person component in the Hicks Building at the University of Sheffield, UK and the online component on Zoom."

CCPBioSim

People ▾ Events ▾ Training Software ▾ Science SlimMD ▾ Account ▾

By Year By Month By Week Today Jump to month

## Training Week 2024

From Monday 02 September 2024 - 09:30  
To Friday 06 September 2024 - 13:00

The CCPBioSim Training Week 2024 will be 5 days of workshops from Monday 2 September to Friday 6 September 2024. This will be a hybrid format event, with the in-person component in the Hicks Building at the University of Sheffield, UK and the online component on Zoom.

Thursday 03 September 2024, 9:30am – 5pm : **Intermediate and Advanced MDAnalysis**

This workshop will build on basic MDAnalysis skills to learn about building more complex analysis scripts. Please feel free to bring your own systems and problems to get help from the experts.