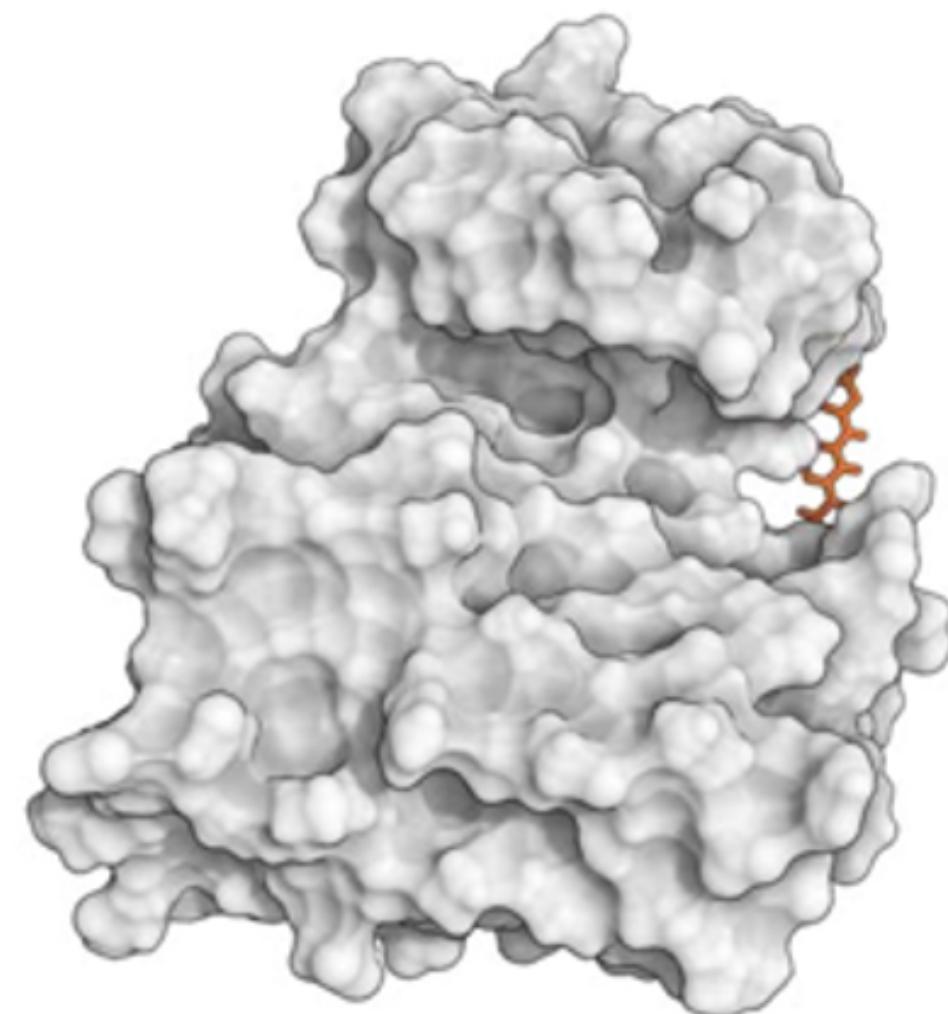
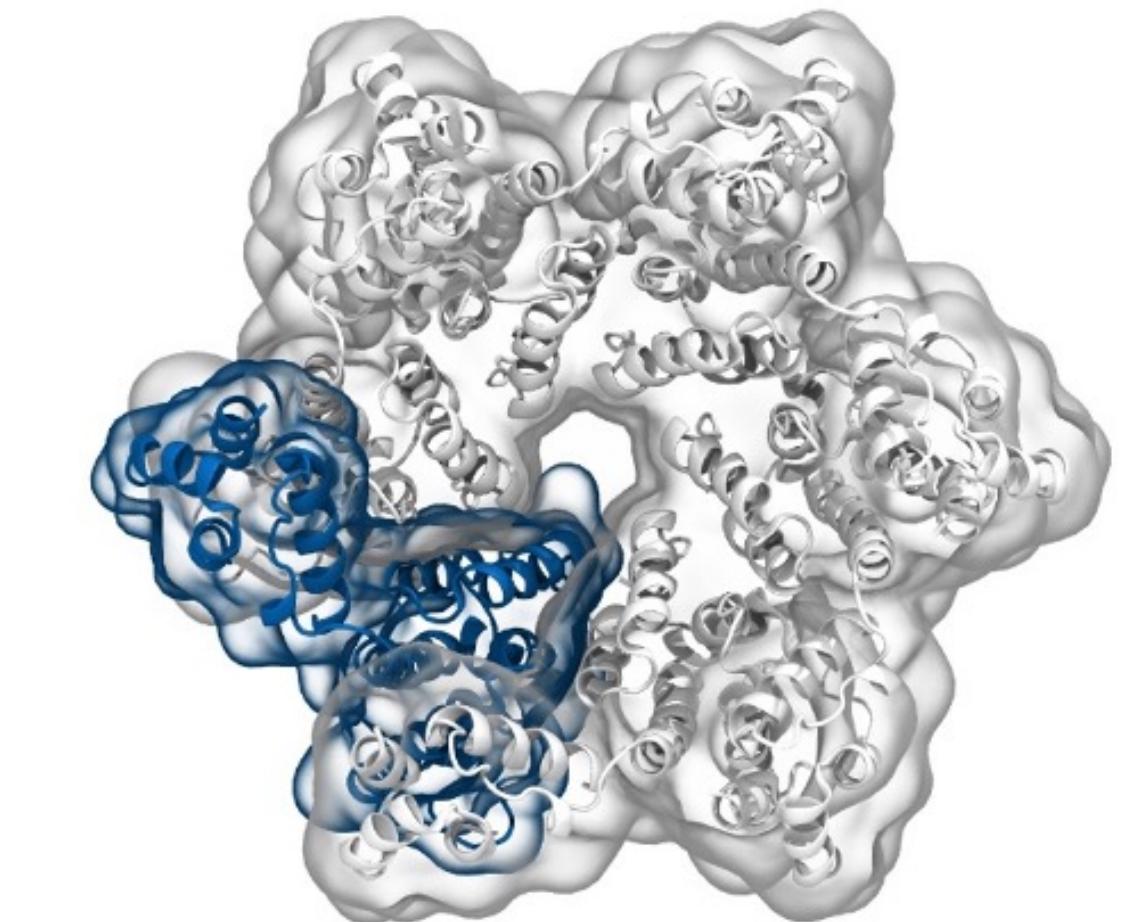


Simulation of Biomolecules



Lecture 4: Setting up a protein simulation

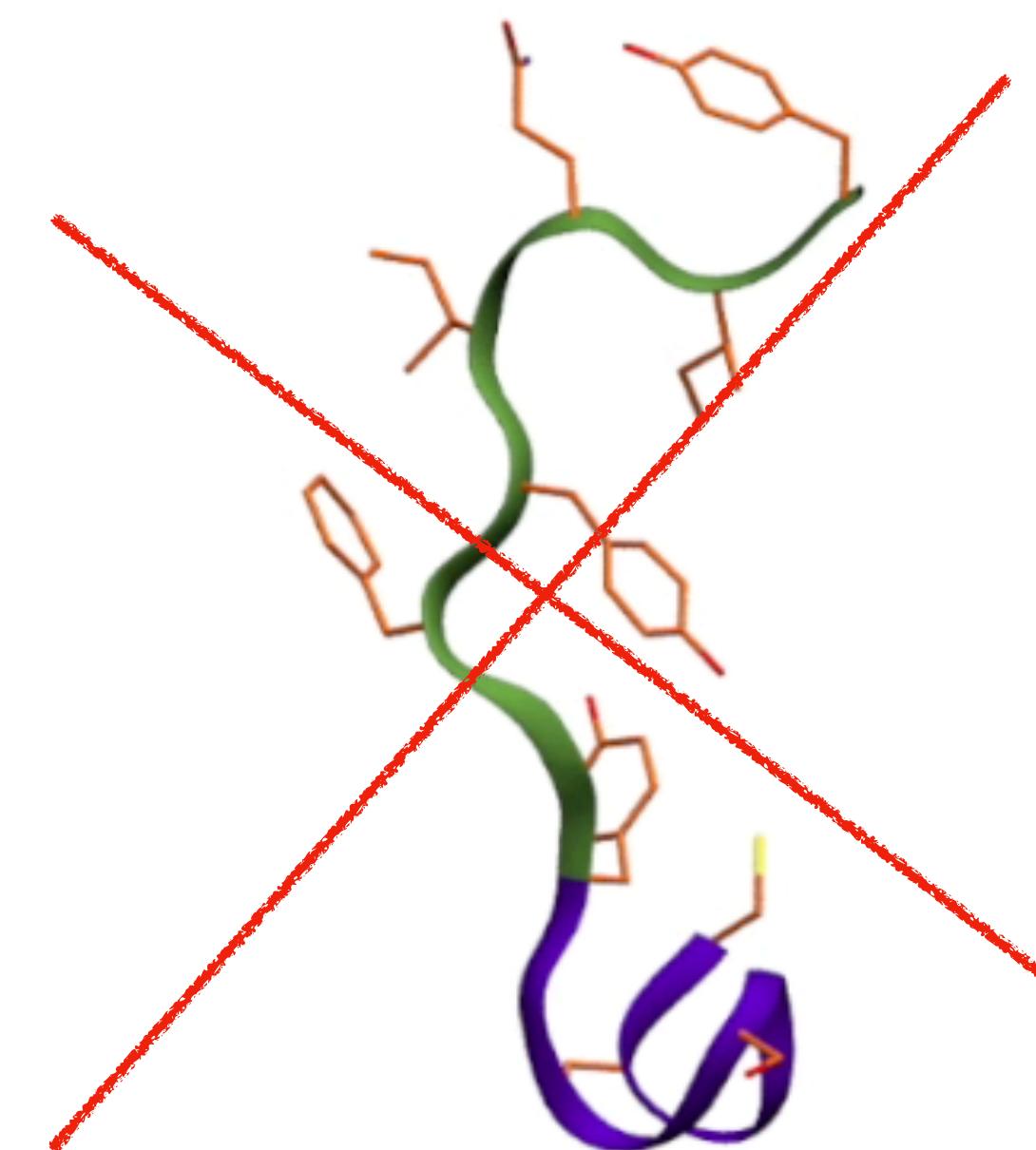
2023 CCP5 Summer School



Dr Matteo Degiacomi
Durham University
matteo.t.degiacomi@durham.ac.uk

Dr Antonia Mey
University of Edinburgh
antonia.mey@ed.ac.uk

You will not run any MD simulation today, but you will learn how to set one up and how to analyse it



Part I: Steps and considerations needed when setting up a simulation

Useful resources to learn running simulations

Best Practices for Foundations in Molecular Simulations [Article v1.0]

Efrem Braun¹, Justin Gilmer², Heather B. Mayes³, David L. Mobley⁴, Jacob I. Monroe⁵, Samarjeet Prasad⁶, Daniel M. Zuckerman⁷

Amber Tutorials

A suite of tutorials for the BioSimSpace framework for interoperable biomolecular simulation [Article v1.0]

Lester O. Hedges^{1,2*}, Sofia Bariani^{3†}, Matthew Burman², Finlay Clark³, Benjamin P. Cossins⁴, Adele Hardie³, Anna M. Herz³, Dominykas Lukauskis⁵, Antonia S.J.S. Mey³, Julien Michel^{2,3*}, Jenke Scheen^{3‡}, Miroslav Suruzhon⁴, Christopher J. Woods¹, Zhiyi Wu⁴

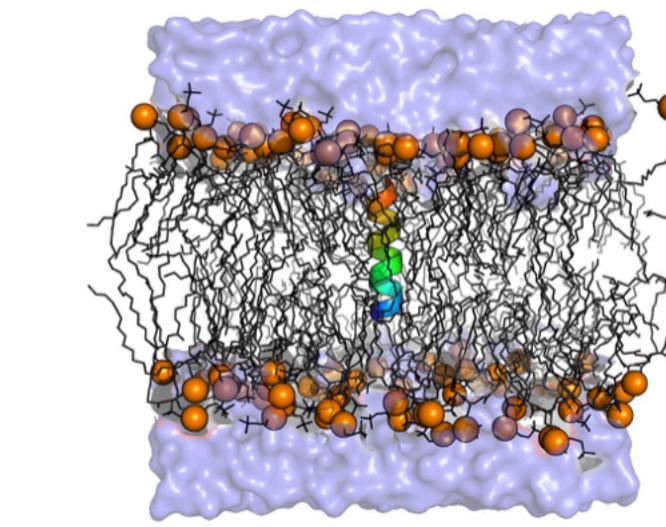
From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]

Justin A. Lemkul

Department of Biochemistry, Virginia Polytechnic Institute and State University
<https://orcid.org/0000-0001-6661-8653>

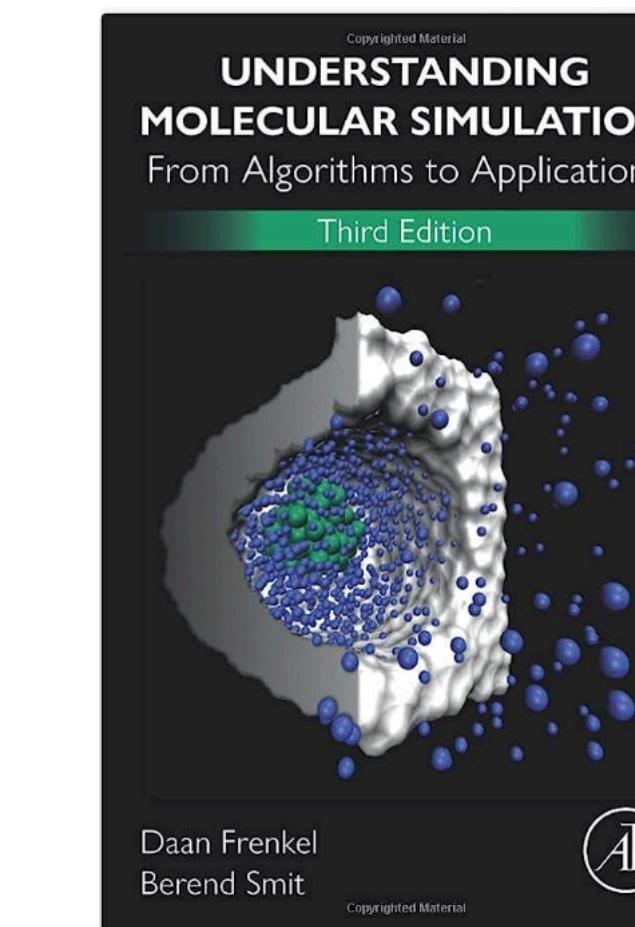
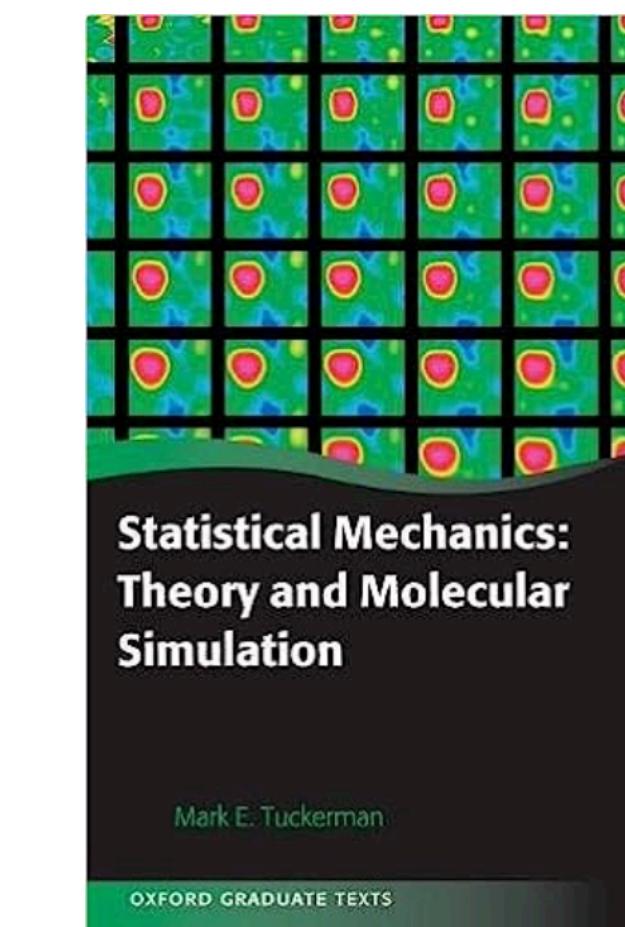
DOI: <https://doi.org/10.33011/livecoms.1.1.5068>

Keywords: tutorials, gromacs, molecular dynamics simulation, computational chemistry



PDF

ARTICLE CODE REPOSITORY



Starting point for setting up a simulation

Getting your protein structure

AlphaFold Protein Structure Database



Getting ligands/co-factors

ZINC20



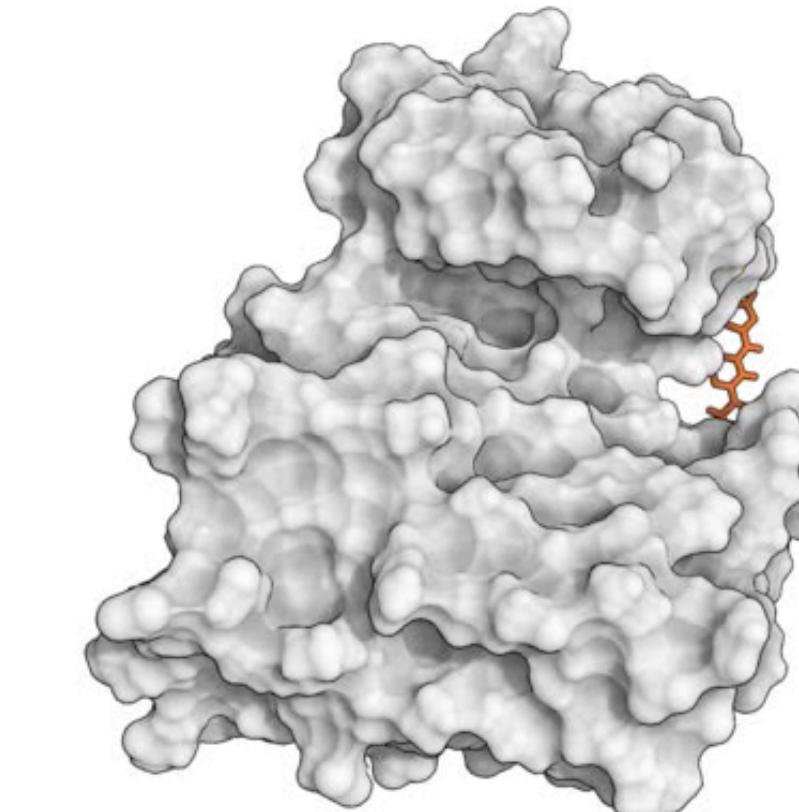
Prepping your protein and/or ligand(s)

SCHRÖDINGER
Maestro

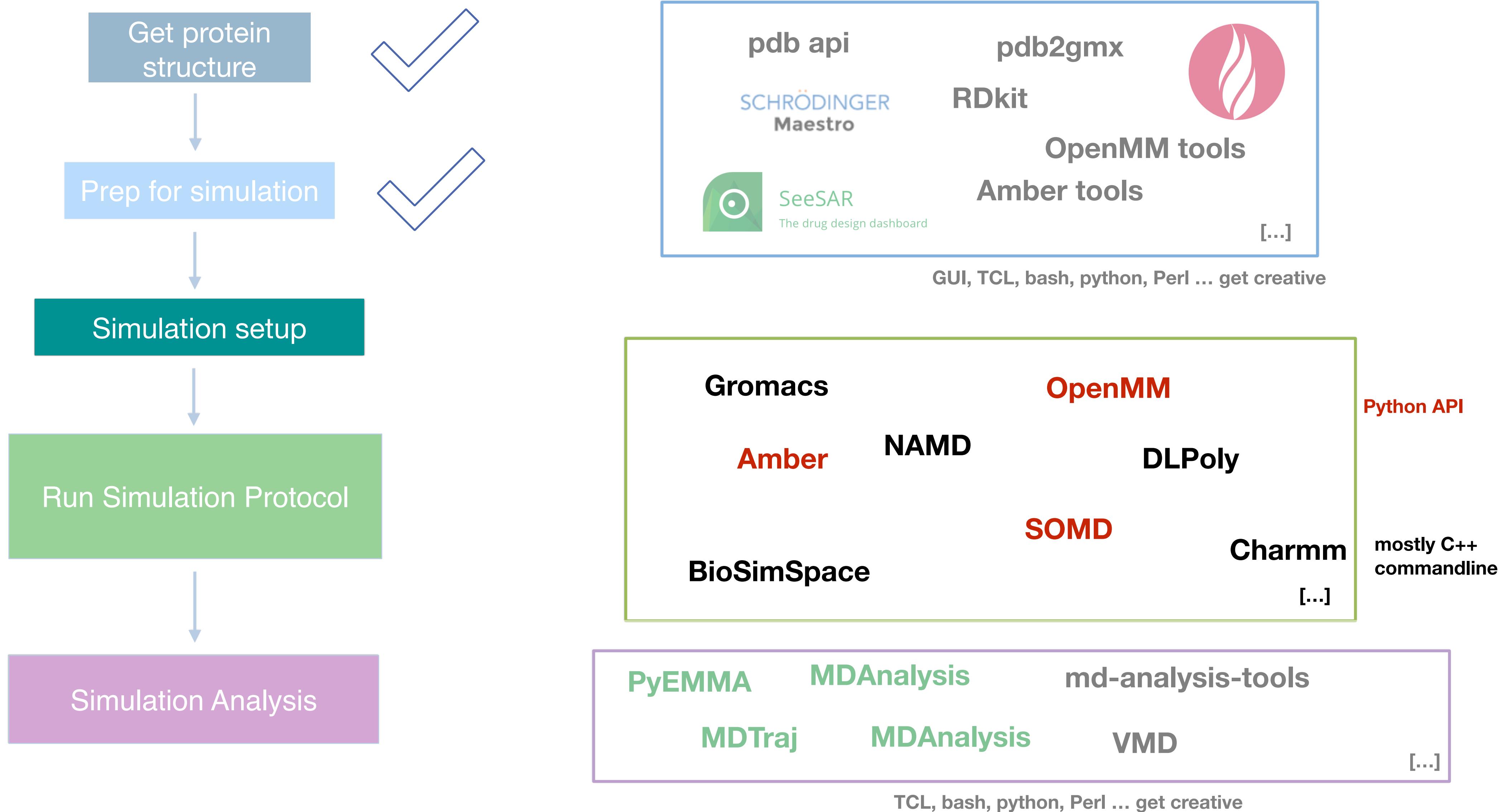


SeeSAR
The drug design dashboard

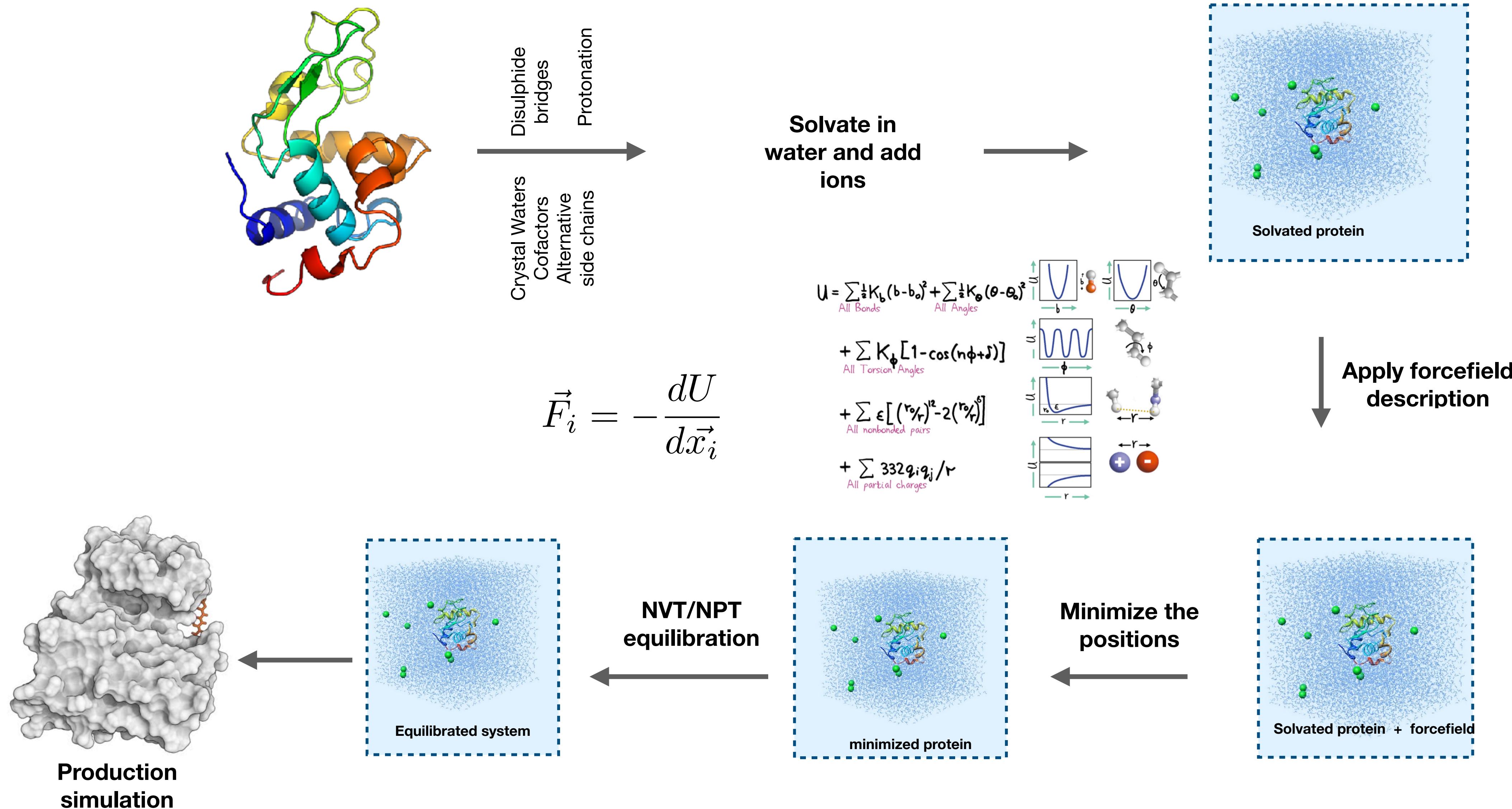
Getting ready for your simulation...



A typical workflow for molecular dynamics

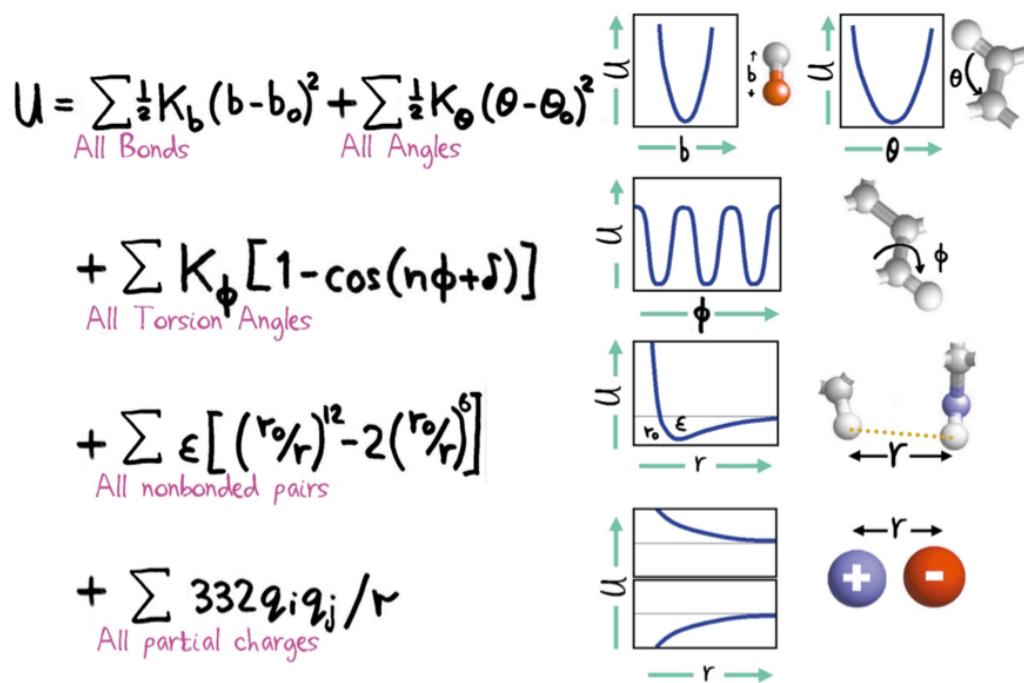


Molecular dynamics require multiple steps for the setup of simulations



There are many different choices for force fields to be made

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

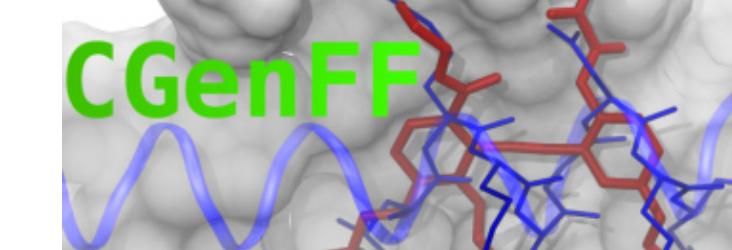


Proteins and other constituents need different force fields

- **Amber** (Peter Kollmann, UCSF)
– Glycam parameters cover most sugars (Robert J. Woods, University of Georgia)
- **CHARMM** (Martin Karplus, Harvard)
– POPC, POPE, DPPC lipids
- **OPLS** (William Jorgensen, Yale)
- **GROMOS** (Wilfried van Gunsteren, ETHZ)

There is no “best force field”!

Small molecule force fields

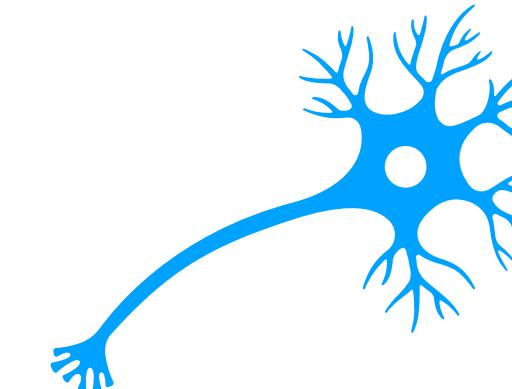


$$E_{\text{pair}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 +$$

$$\sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i,j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

GAFF

Machine learned force fields

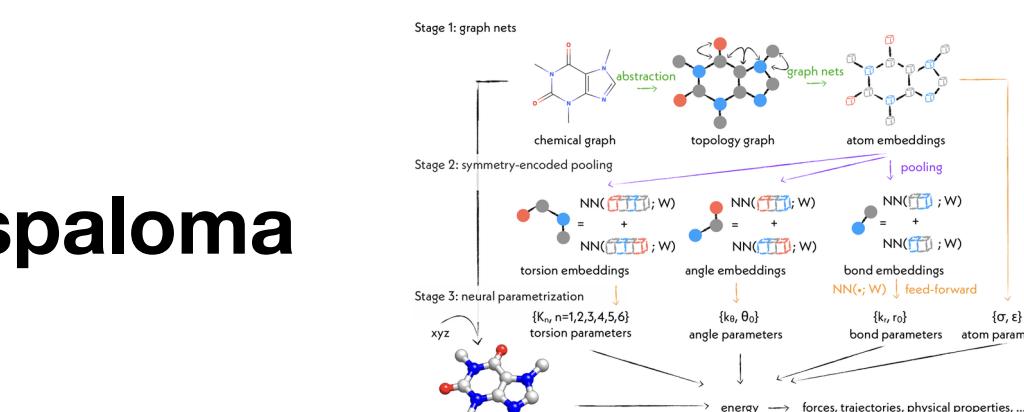


Ani-2x

SchNet
PhysNet

Espaloma

BandNN



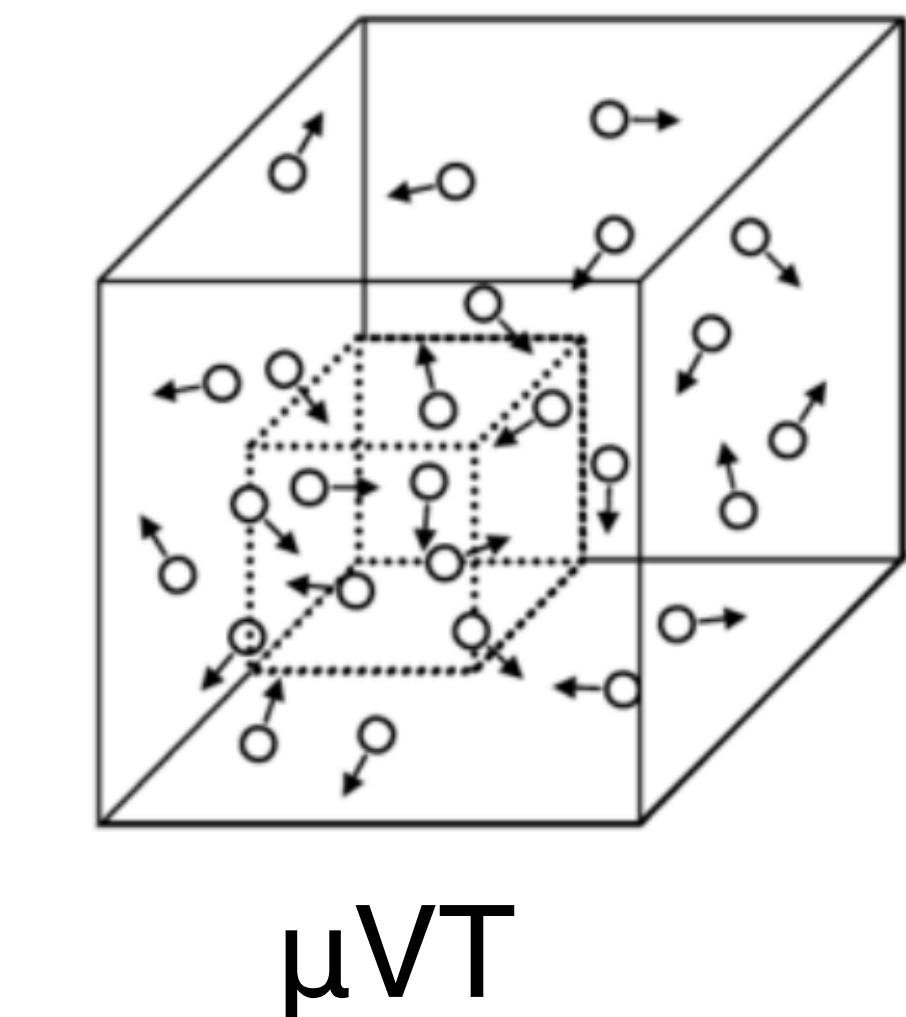
Coarse-grained force fields....

Choosing your thermodynamic ensemble

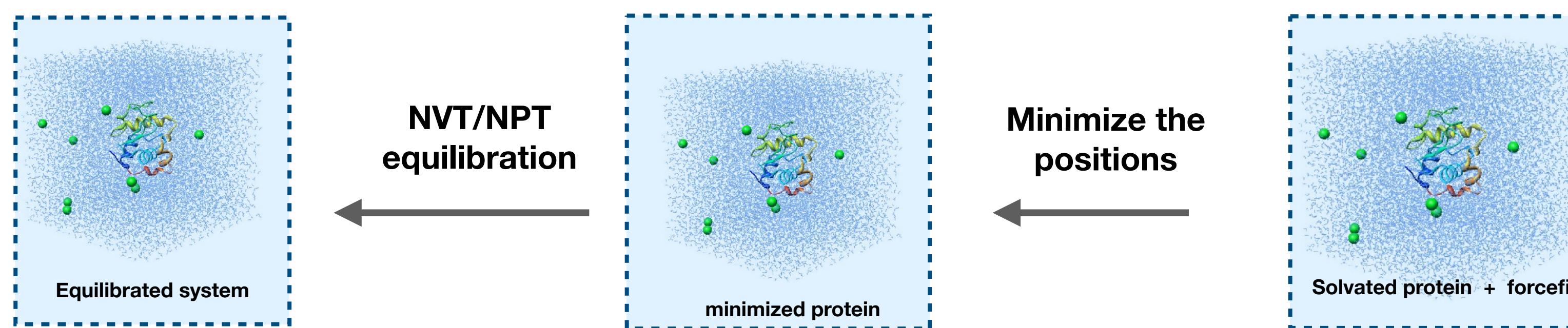
Simulations replicate a specific *thermodynamic ensemble* (typically NVT or NPT), or even grand canonical (μ VT)

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

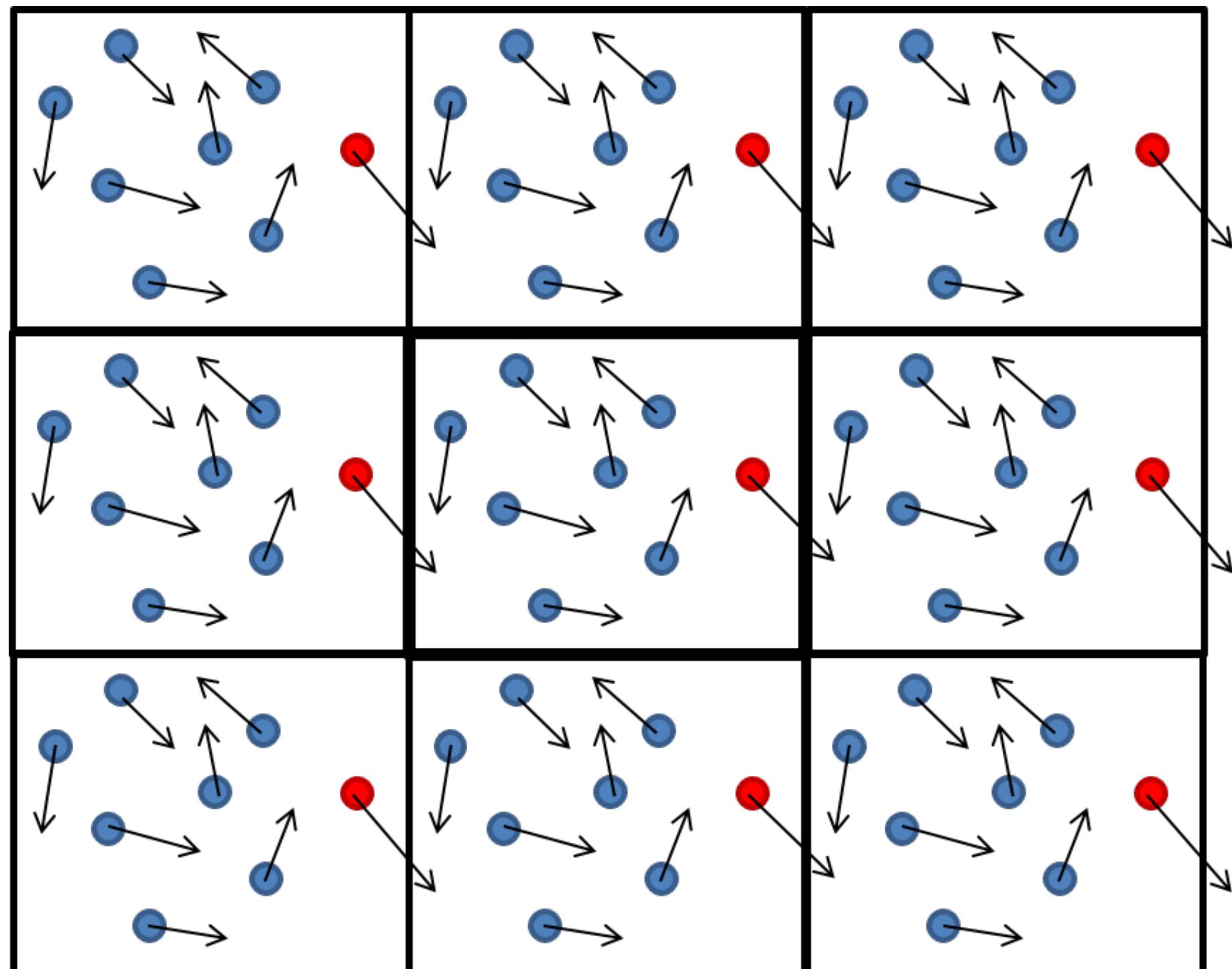
- Nose-Hoover
- Berendsen
- Parrinello-Rahman
- Langevin piston
- ...



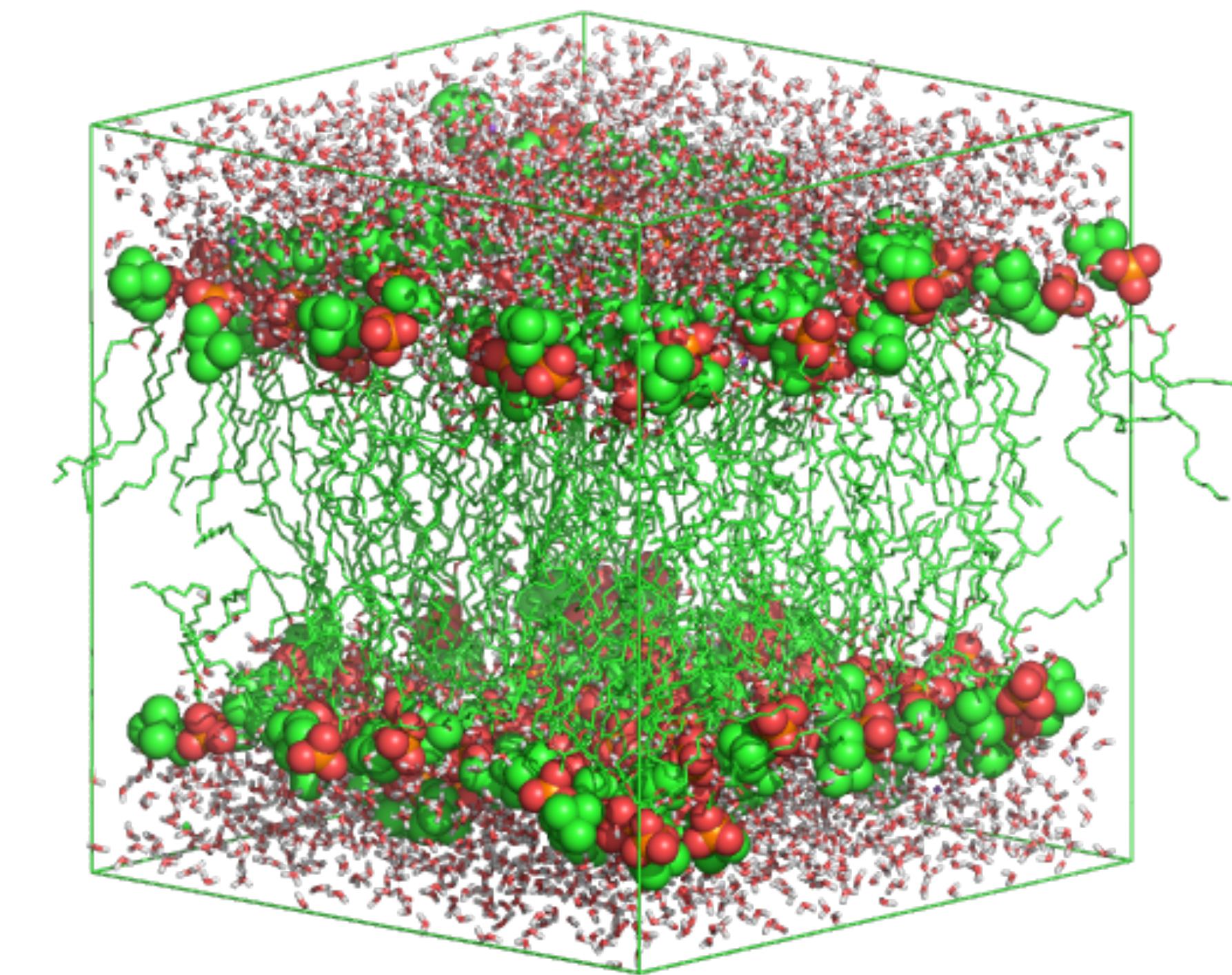
μ VT



Periodic boundary conditions and pressure coupling



Typically you chose periodic boundary conditions in x-y-z.



If you want to simulate membrane systems you want to chose semi-isotropic pressure coupling!

Sampling timescales for protein systems

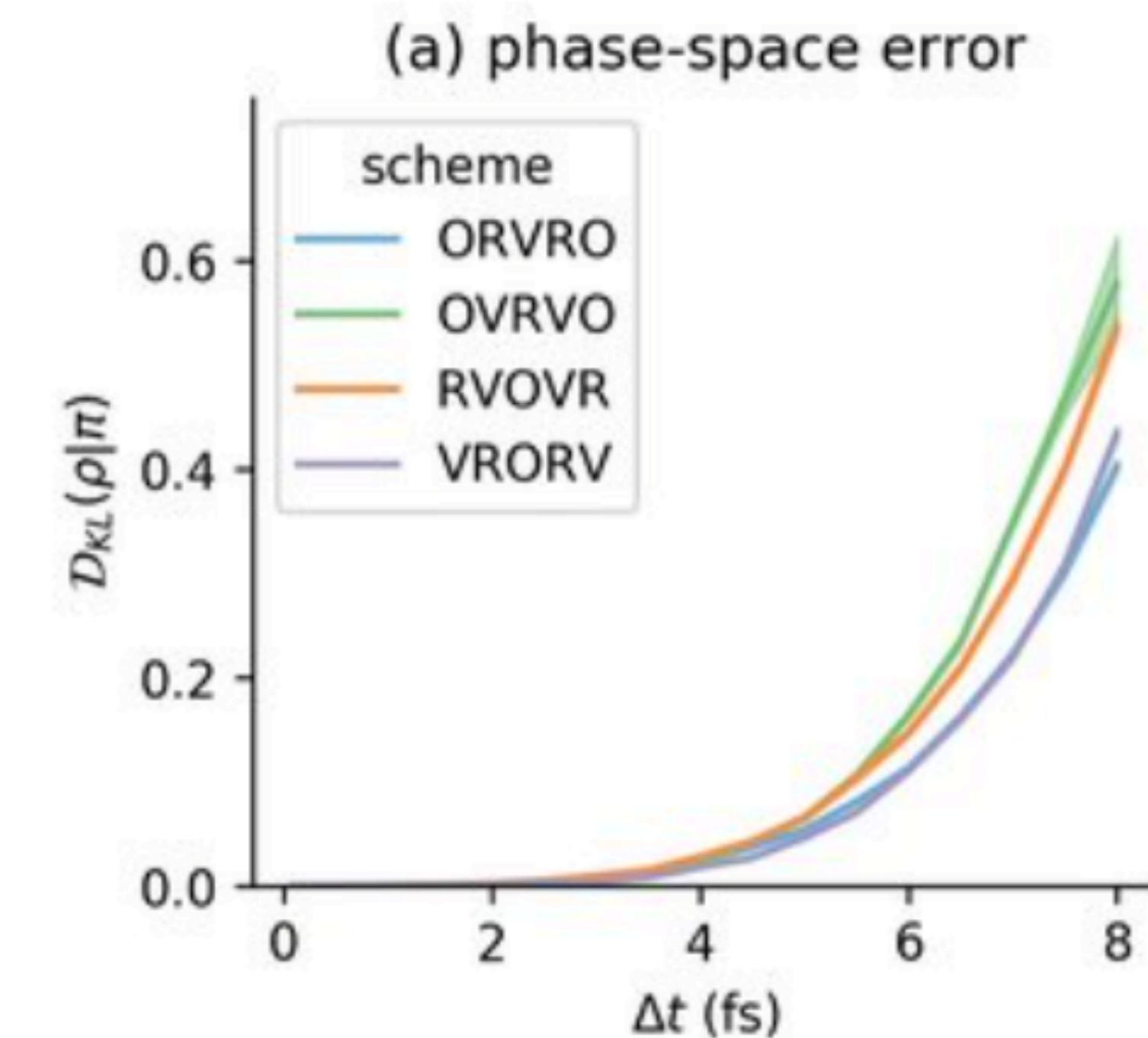
The steepest gradient determines the smallest timestep:

Timestep size is imposed by the fastest phenomenon we want to observe :

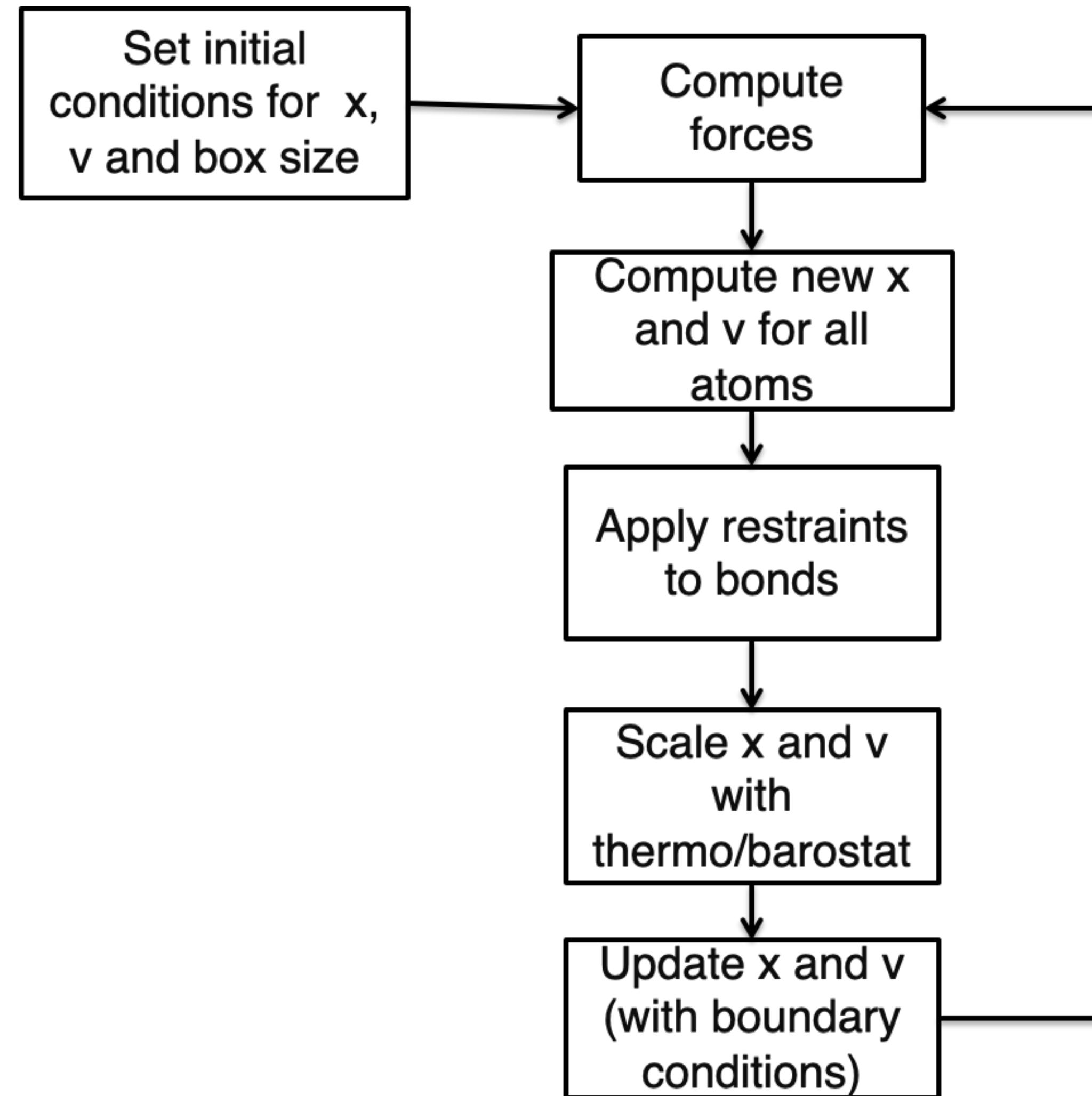
- Covalent bond hydrogen-heavy atom (10^{14} Hz): 0.5 fs
- Covalent bond heavy atom-heavy atom: 1 fs
- Angles fluctuations: 2 fs

Restraining covalent bond distances
allows to use 1-2 fs timesteps
(restraining methods: SHAKE, RATTLE,
LINCS,...)

Hydrogen Mass repartitioning: 4 fs
Other integrators: 4 fs- 6 fs.



An actual MD timestep



An example MD protocol

- **AIM:** equilibrate your system
- **YOU WANT:** constant volume, pressure and temperature, stable Root Mean Square Deviation, healthy Ramachandran plot, no exotic chemistry, bulk water (if used), stabilization of whichever other quantity you are interested about (e.g. Rgyr, ...)

Example:

1. Minimize energy, 1000 steepest descent
2. Heat system from 0 to 300 K in 500 ps, NPT, Berendsen barostat 1 atm. α -carbon restrained with 10 Kcal/mol harmonic potential. 2 fs timestep, SHAKE all bonds,
3. 1 ns nVT equilibration with Langevin dynamics, no atom constrained.
4. **Production:** 200 ns NPT, Nose-Hoover barostat, PME for electrostatics

DETERMINE HANDLING OF CUTOFFS

- As a general rule, electrostatics are long-range enough that either the cutoff needs to be larger than the system size (for finite systems) or periodicity is needed along with full treatment of long-range electrostatics (Section 3.4)
- Nonpolar interactions can often be safely treated with cutoffs of 1-1.5 nm as long as the system size is at least twice that, but long-range dispersion corrections may be needed (Section 4.1)

CHOOSE APPROPRIATE SETTINGS FOR THE DESIRED ENSEMBLE

- Pick a thermostat that gives the correct distribution of temperatures, not just the correct average temperature; if you have a small system or a system with weakly interacting component choose one which works well even in the small-system limit.
- Pick a barostat that gives the correct distribution of pressures
- Consider the known shortcomings and limitations of certain integrators and thermostats/barostats and whether your choices will impact the properties you are calculating

CHOOSE AN APPROPRIATE Timestep FOR STABILITY AND AVOIDING ENERGY DRIFT

- Determine the highest-frequency motion in the system (typically bond vibrations unless bond lengths are constrained)
- As a first guess, set the timestep to approximately one tenth of the highest-frequency motion's characteristic period
- Test this choice by running a simulation in the microcanonical ensemble, and ensure that energy is conserved

Let's try setting up a simulation with BioSimSpace

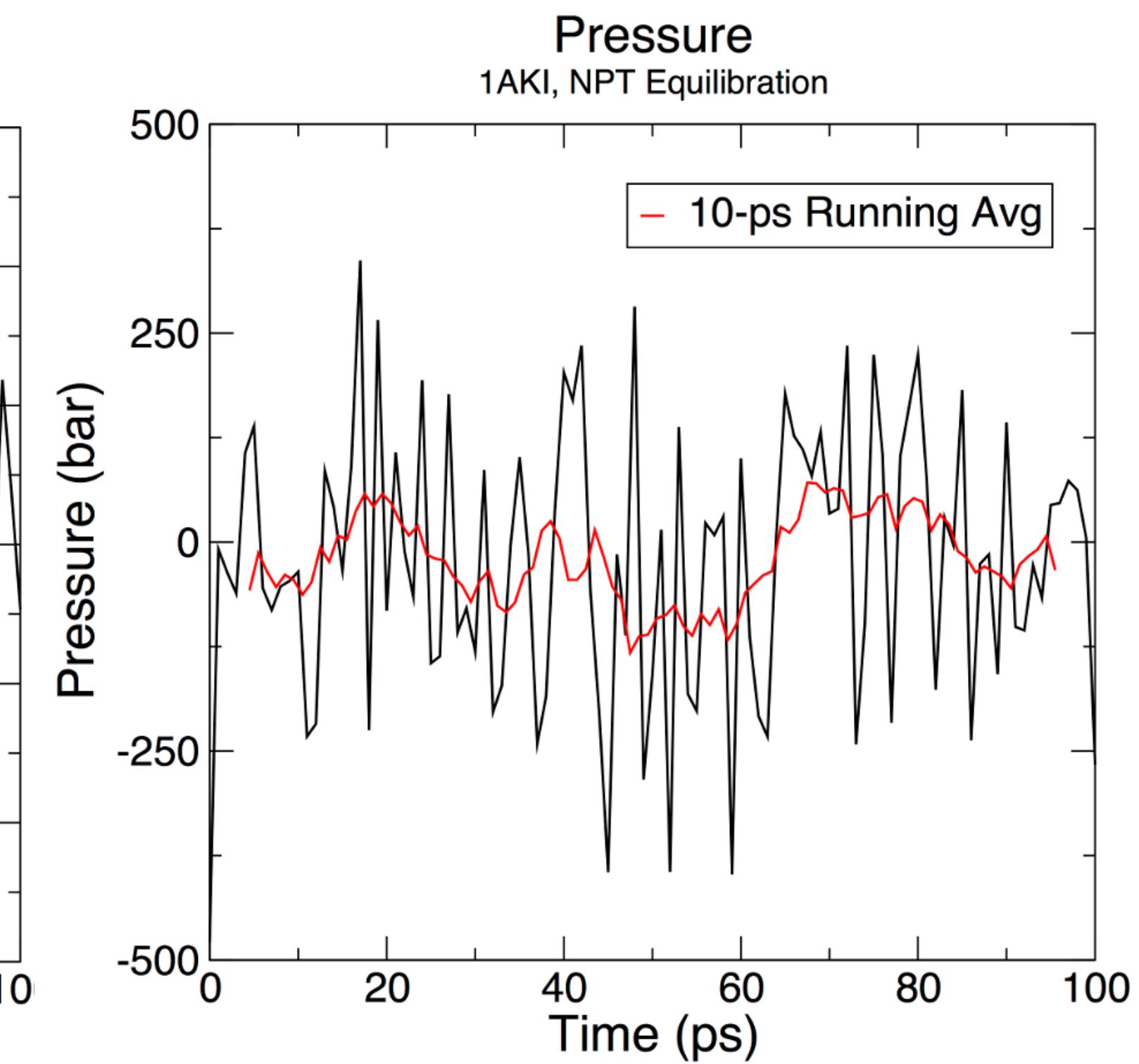
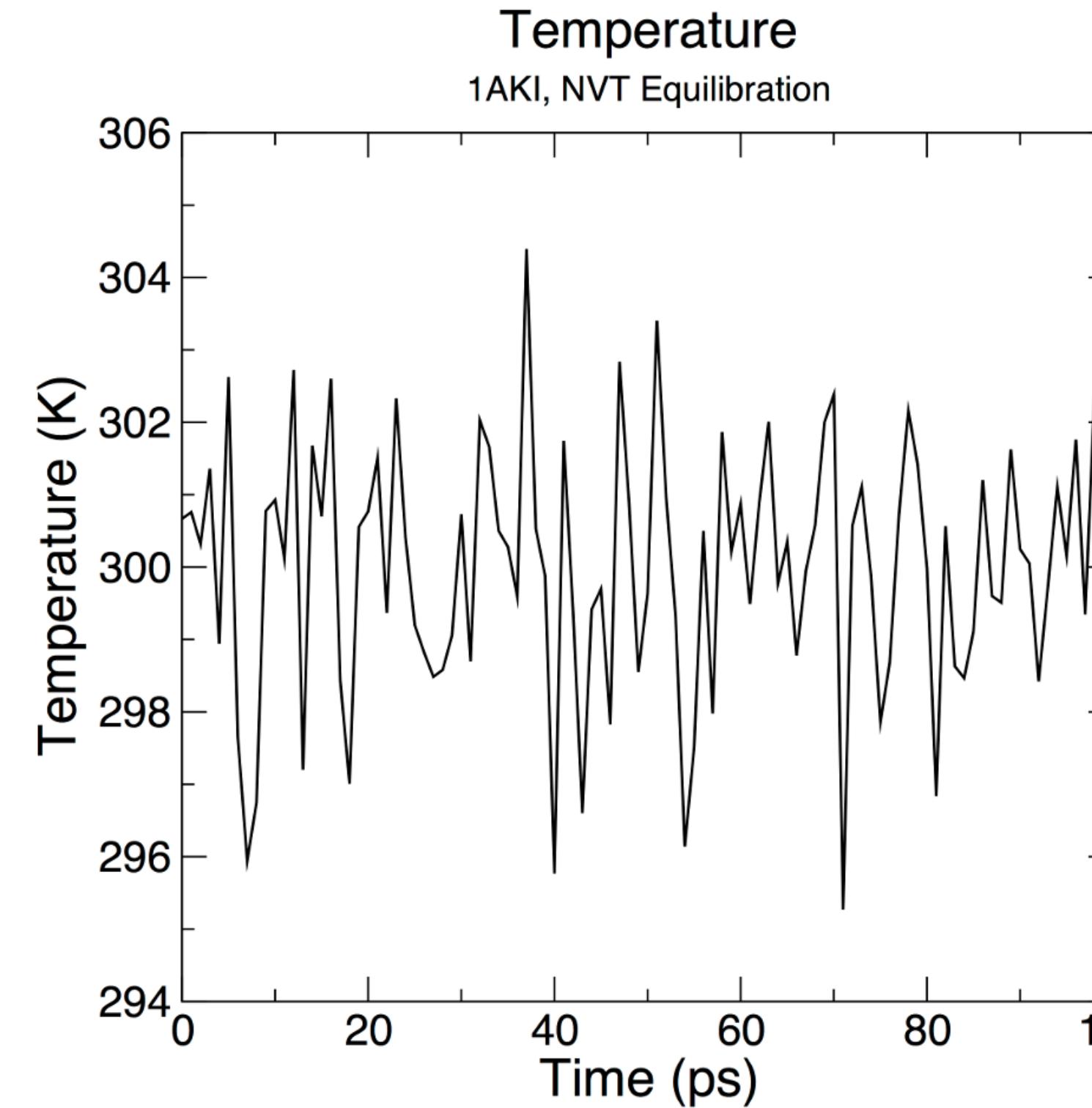
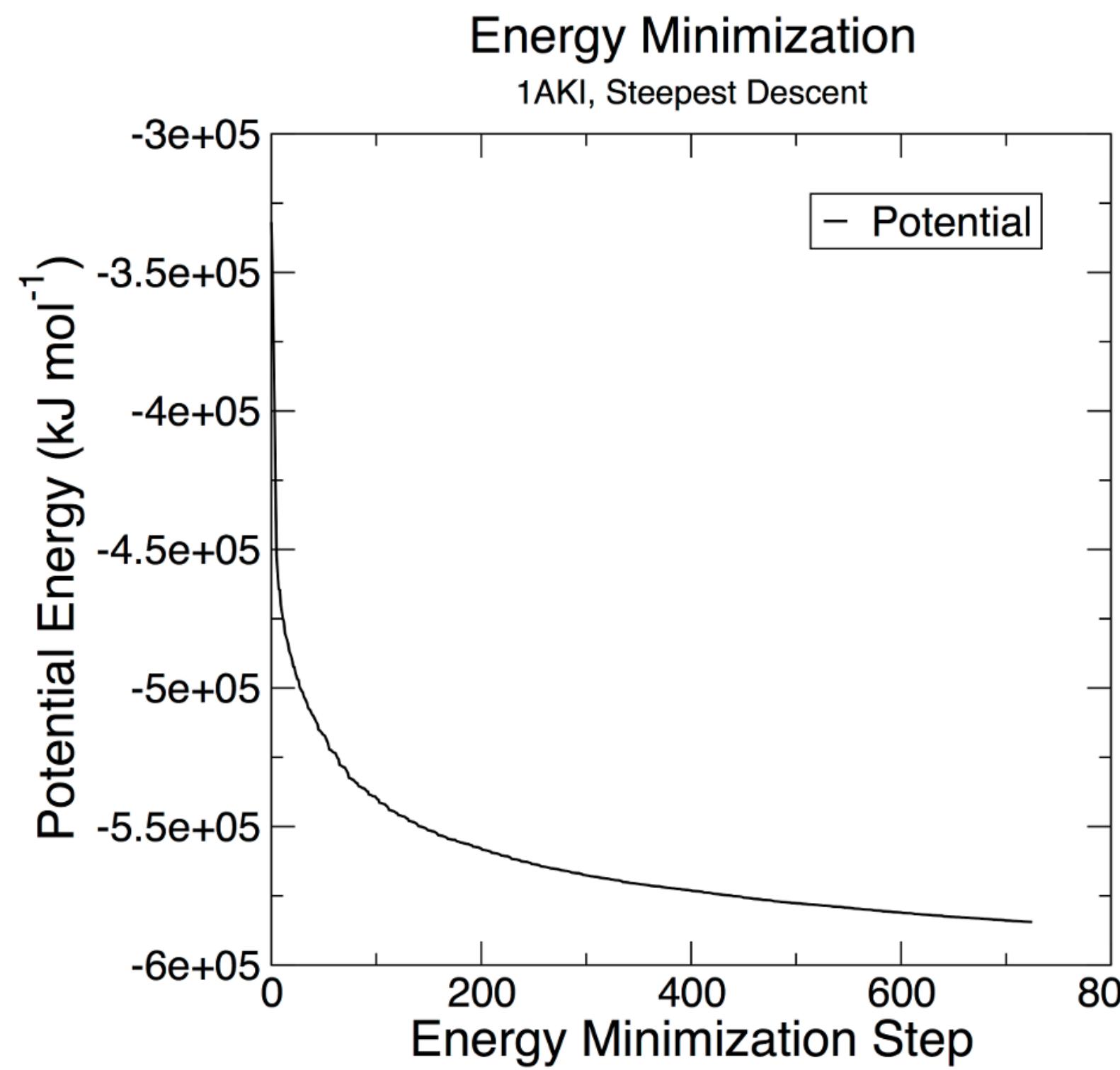
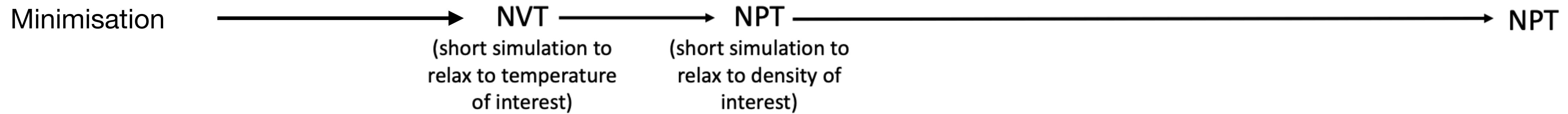
Part II: Steps and considerations needed when setting up a simulation



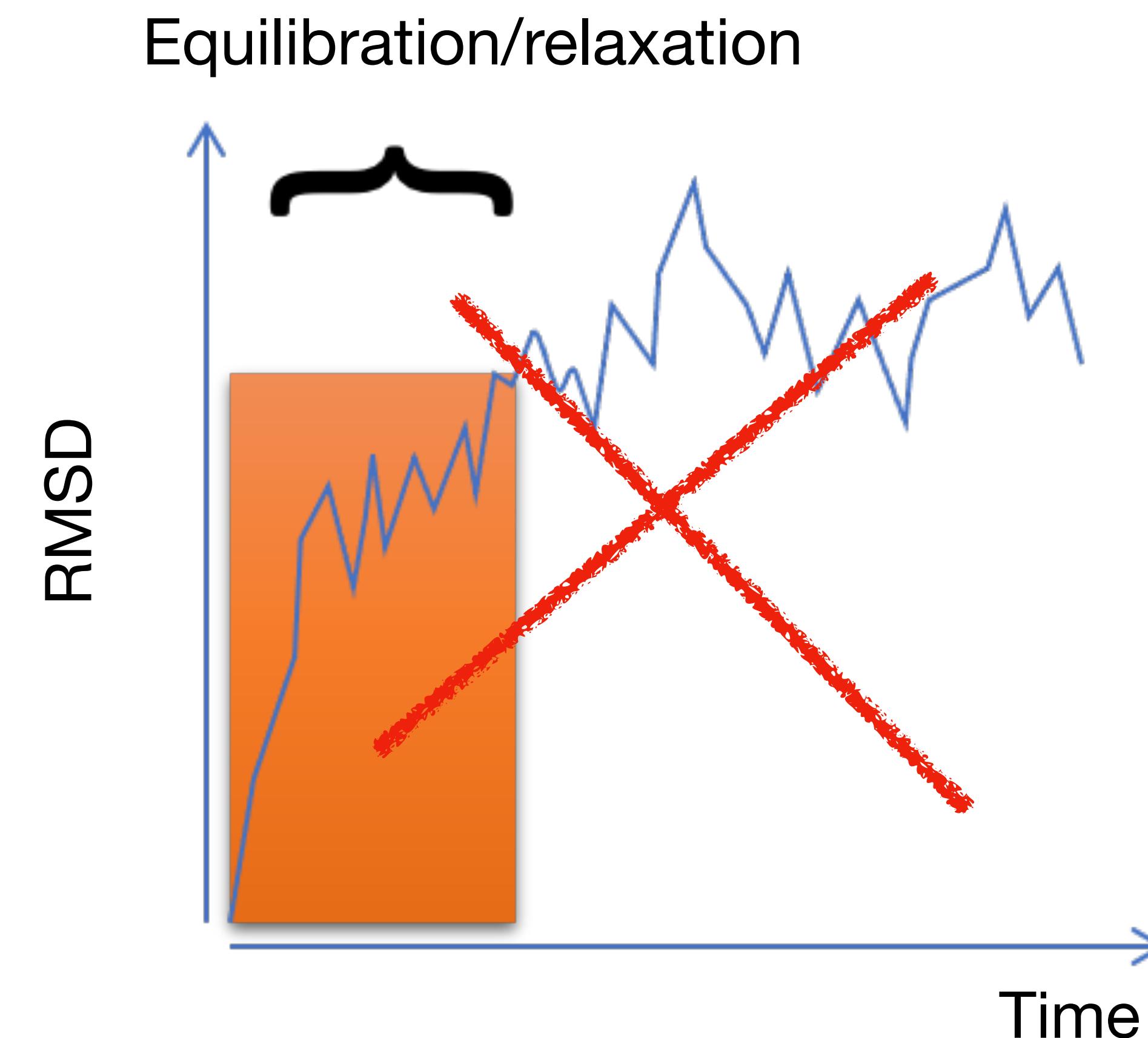
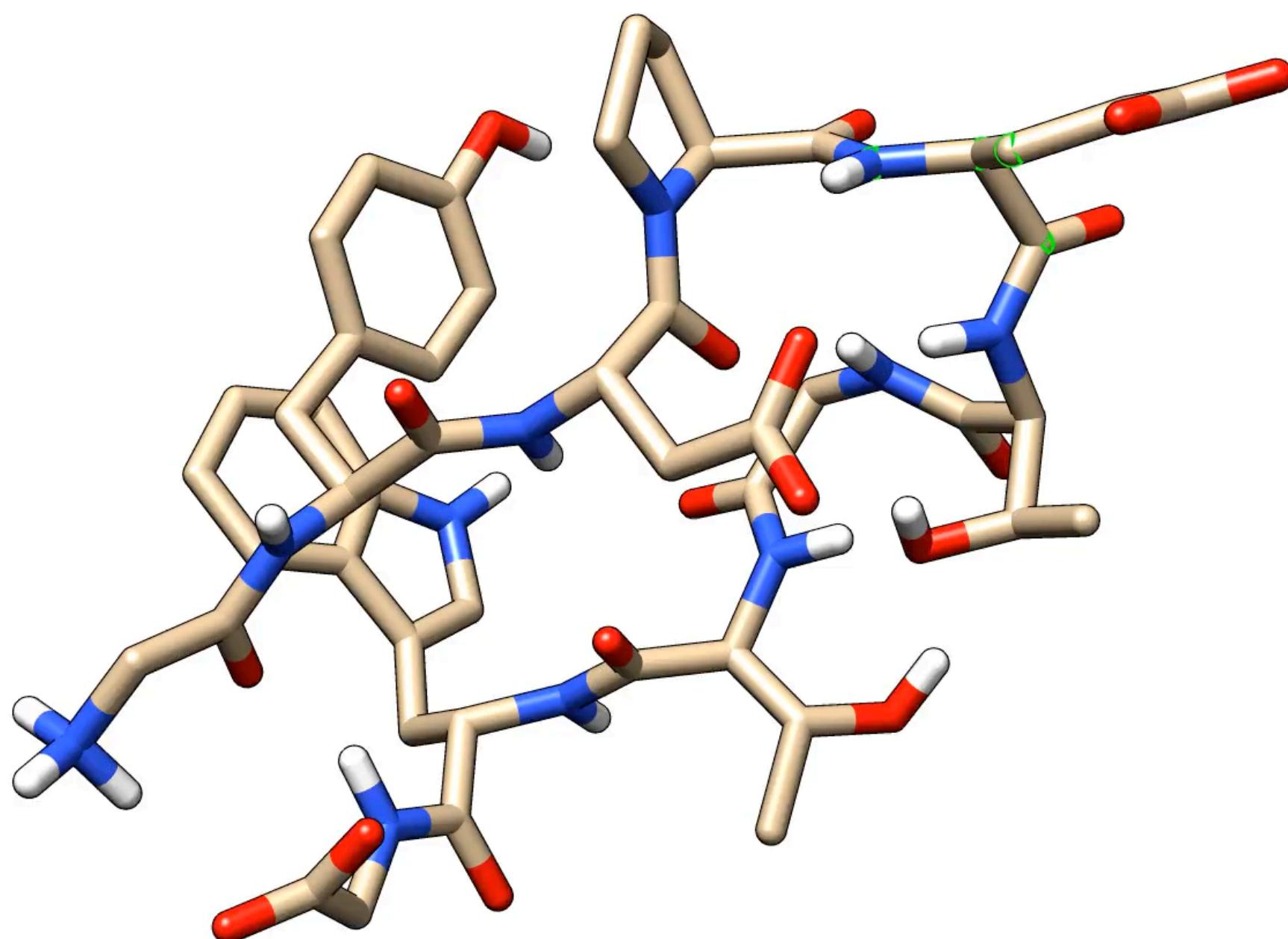
Some of the material presented here is adapted from Prof. Charlie Laughton

Volume and pressure equilibration

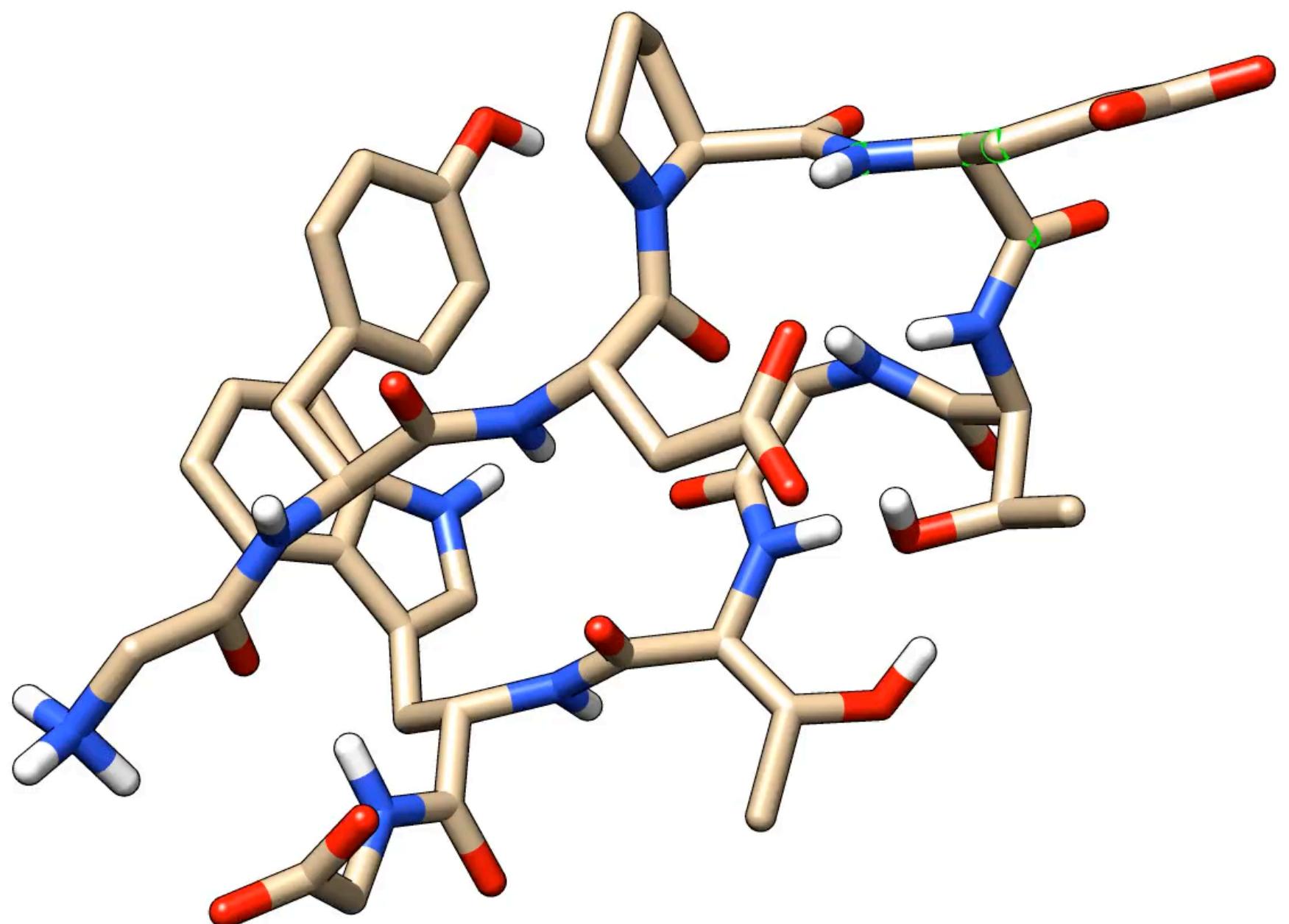
Steps until production:



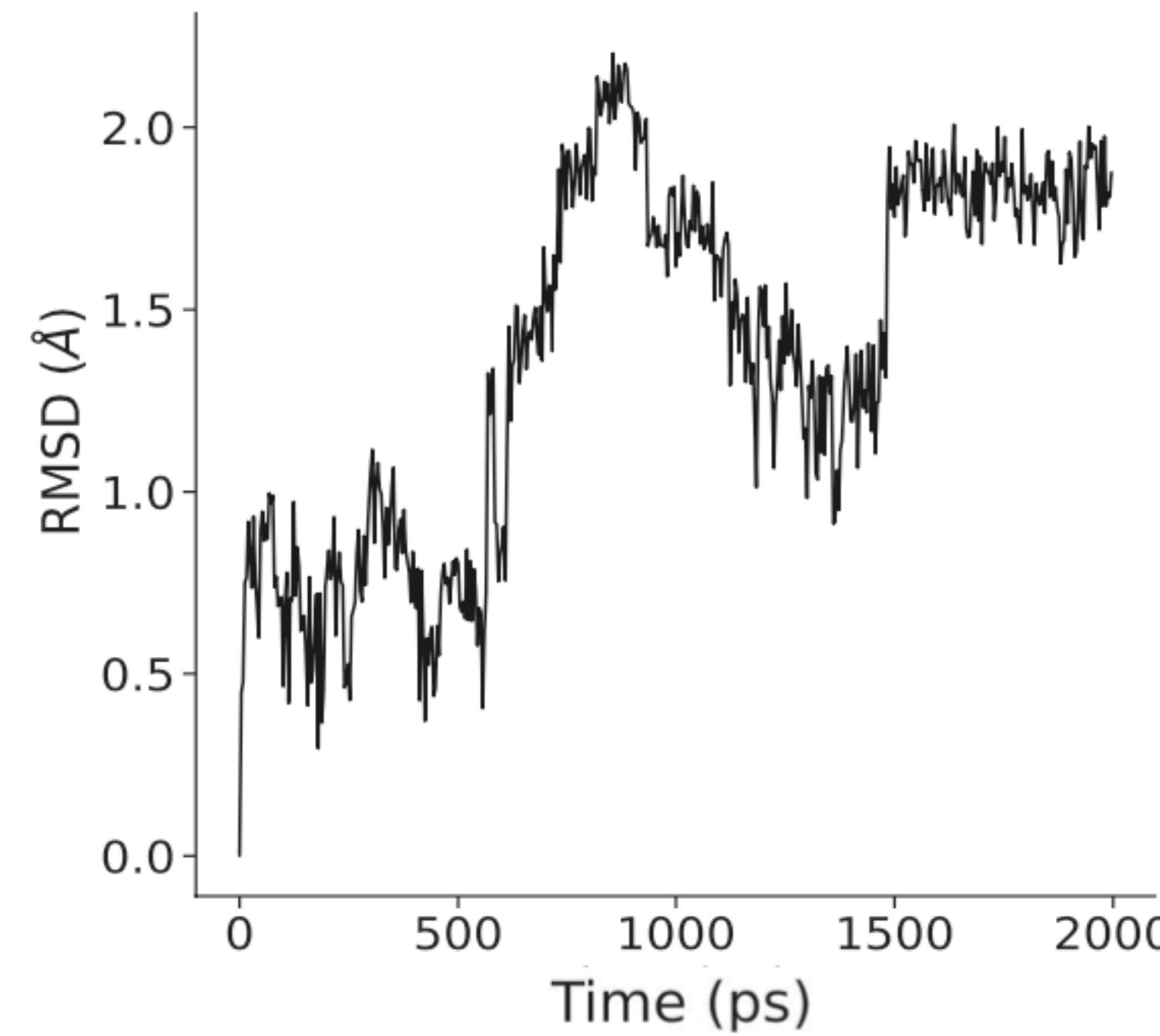
What do we mean by equilibration?



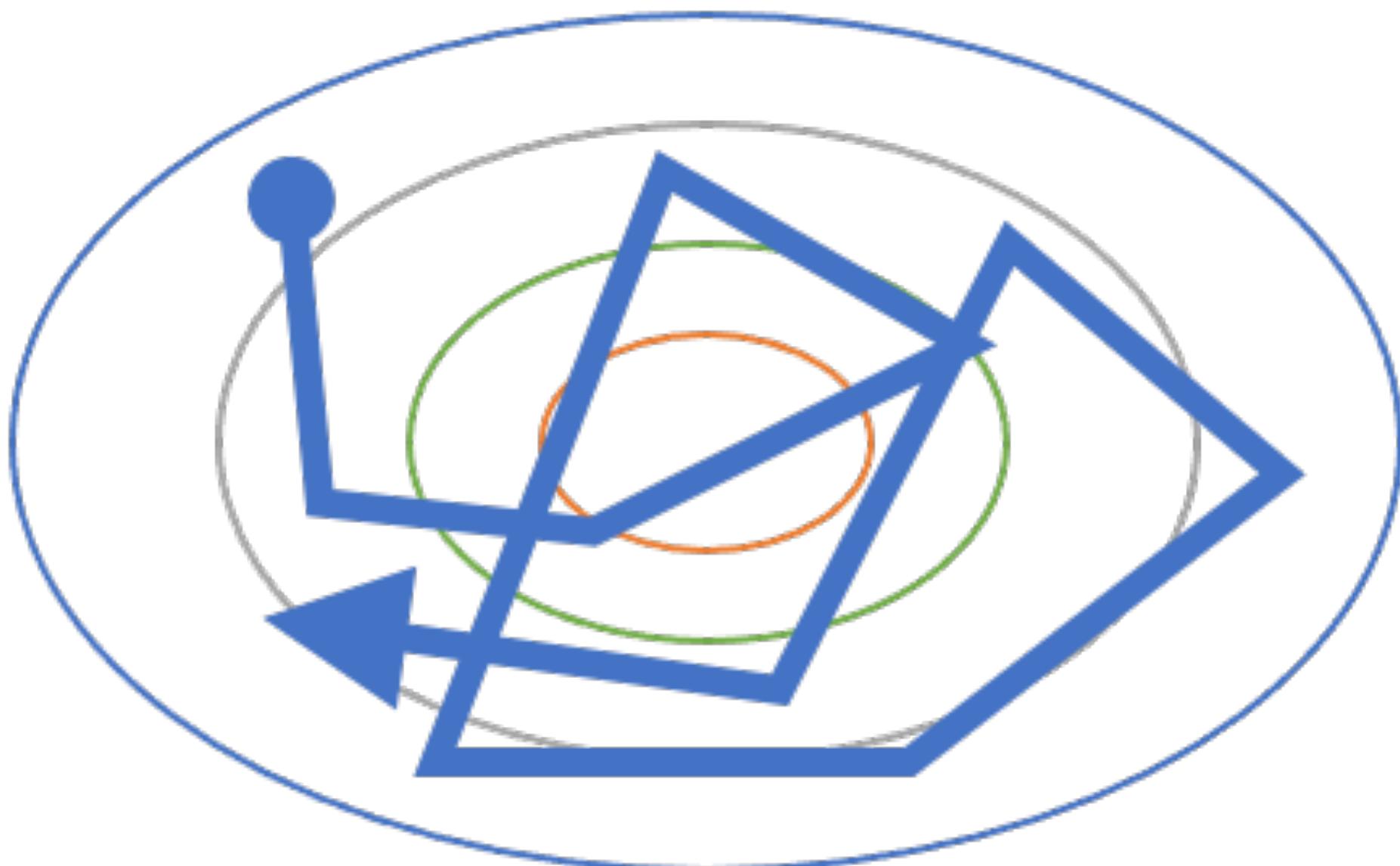
What is RMSD?



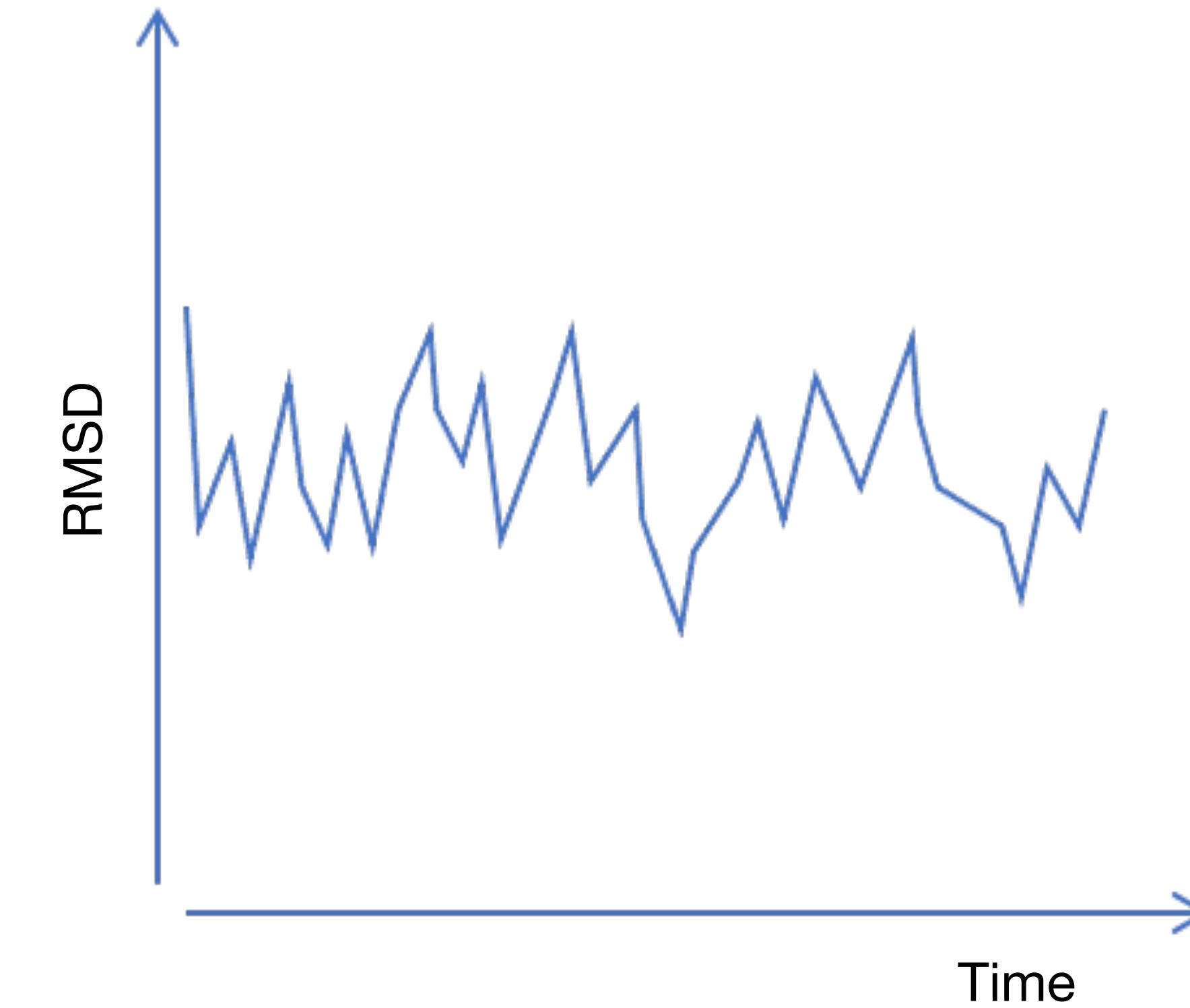
$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=0}^N (v_i - w_i)^2}$$



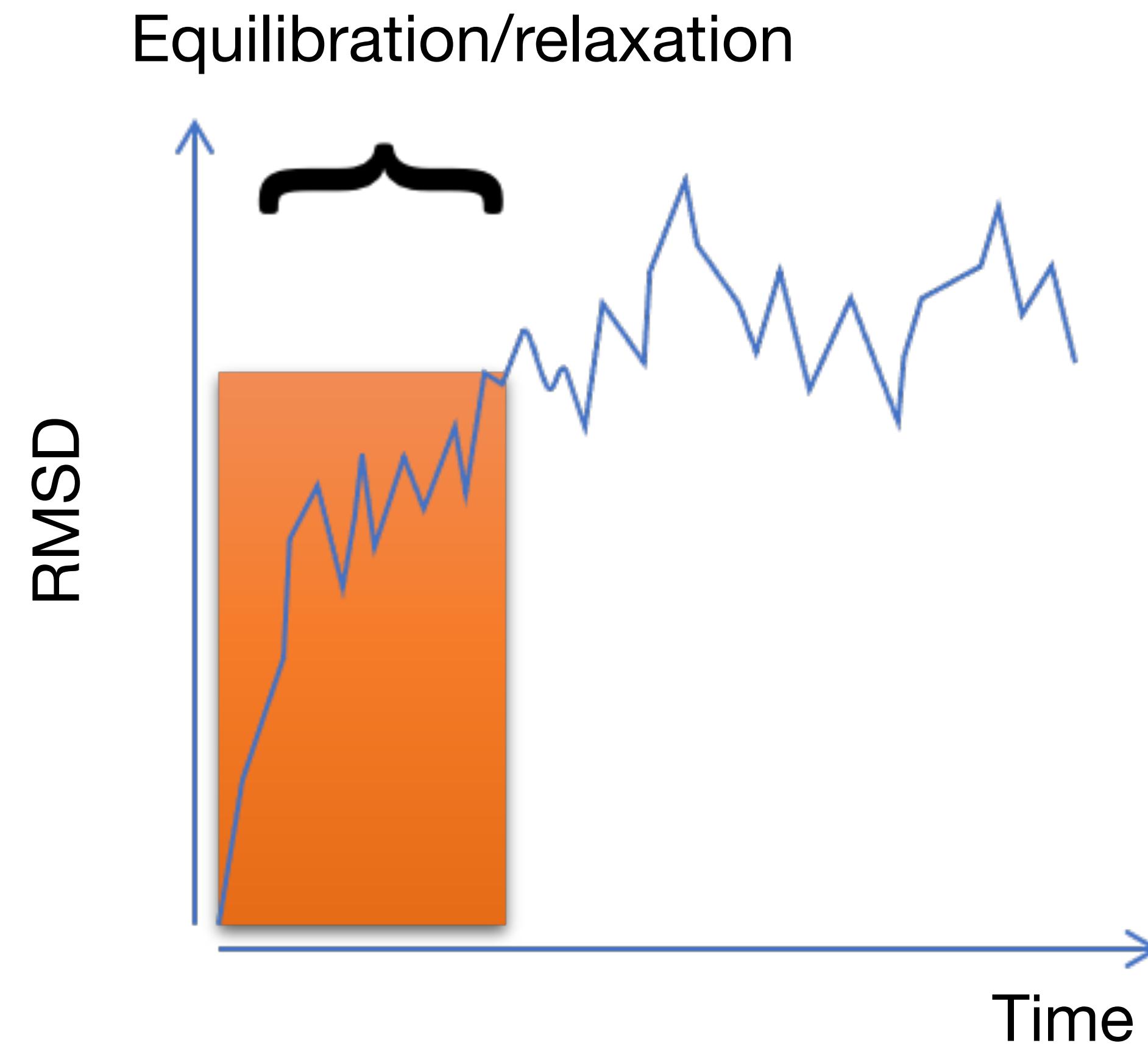
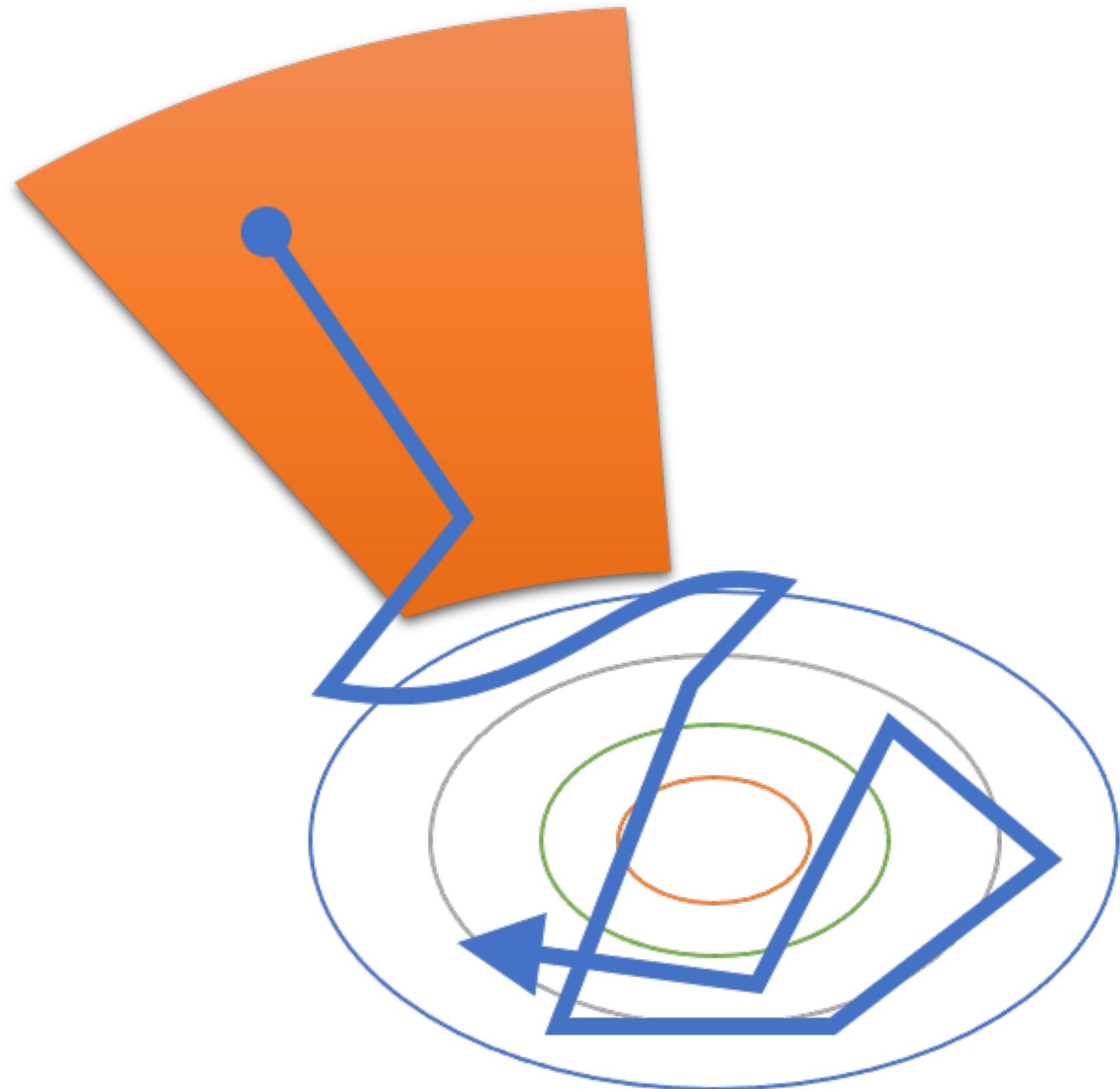
An MD walker over a potential energy surface



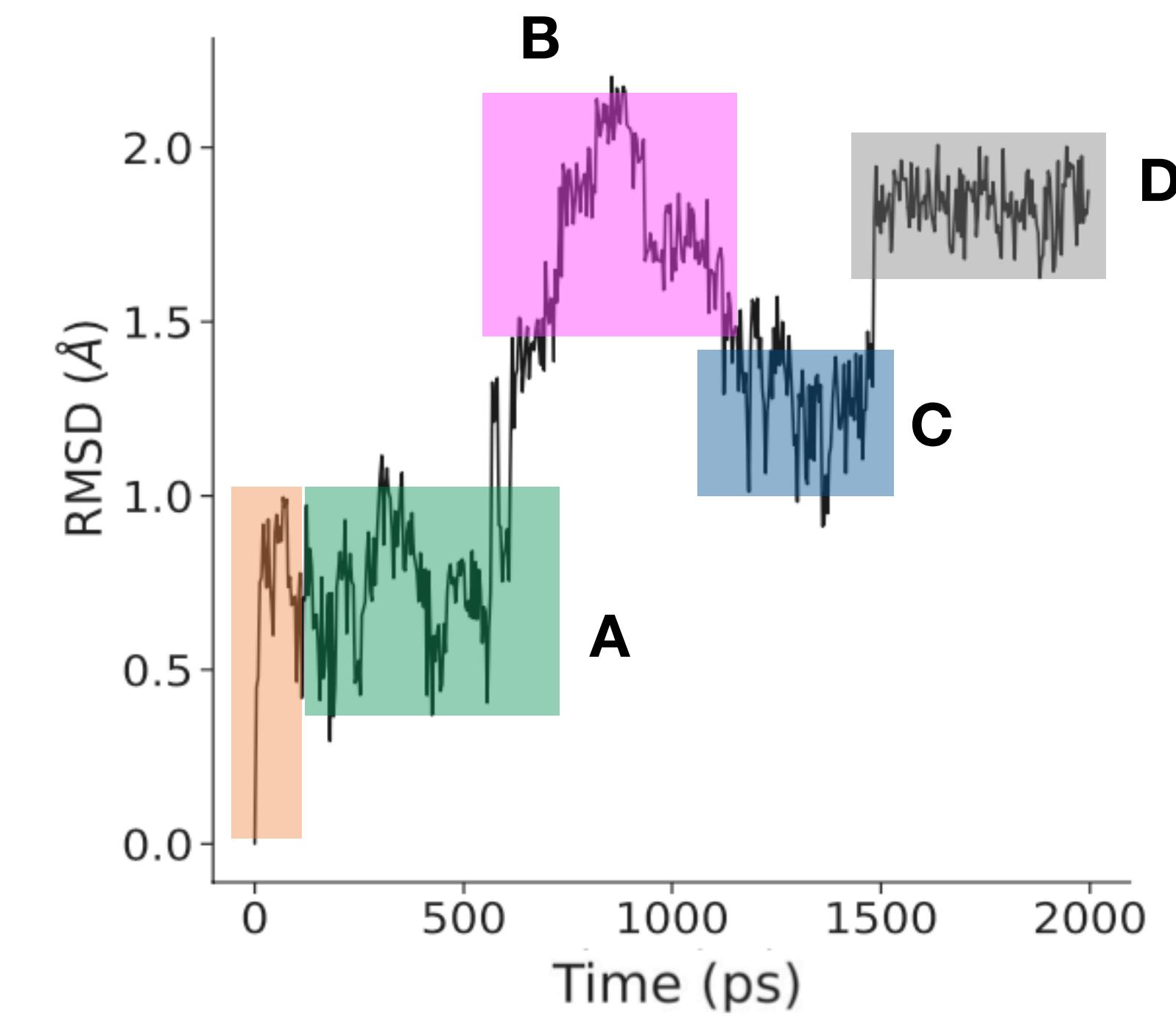
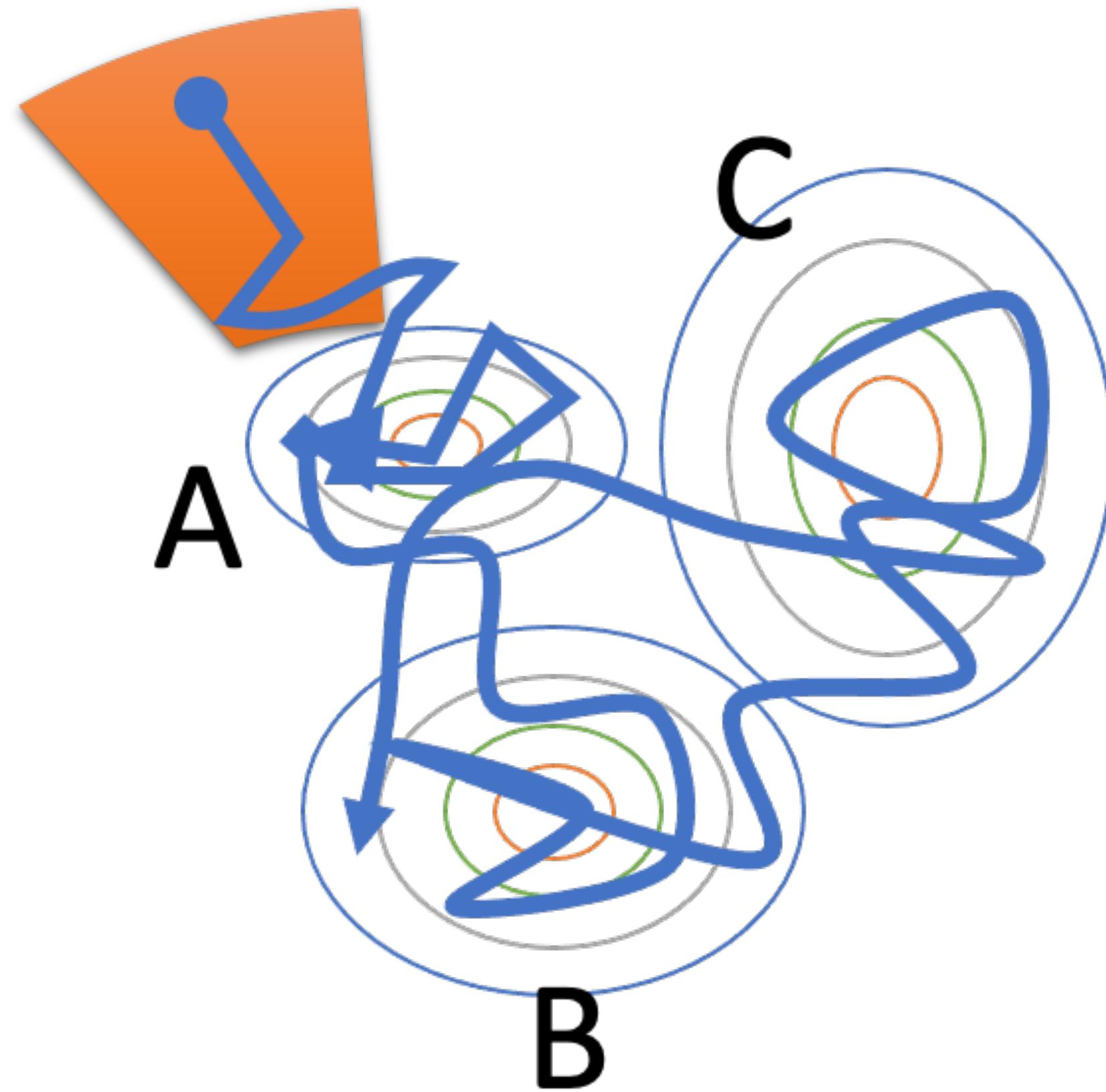
$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=0}^N (v_i - w_i)^2}$$



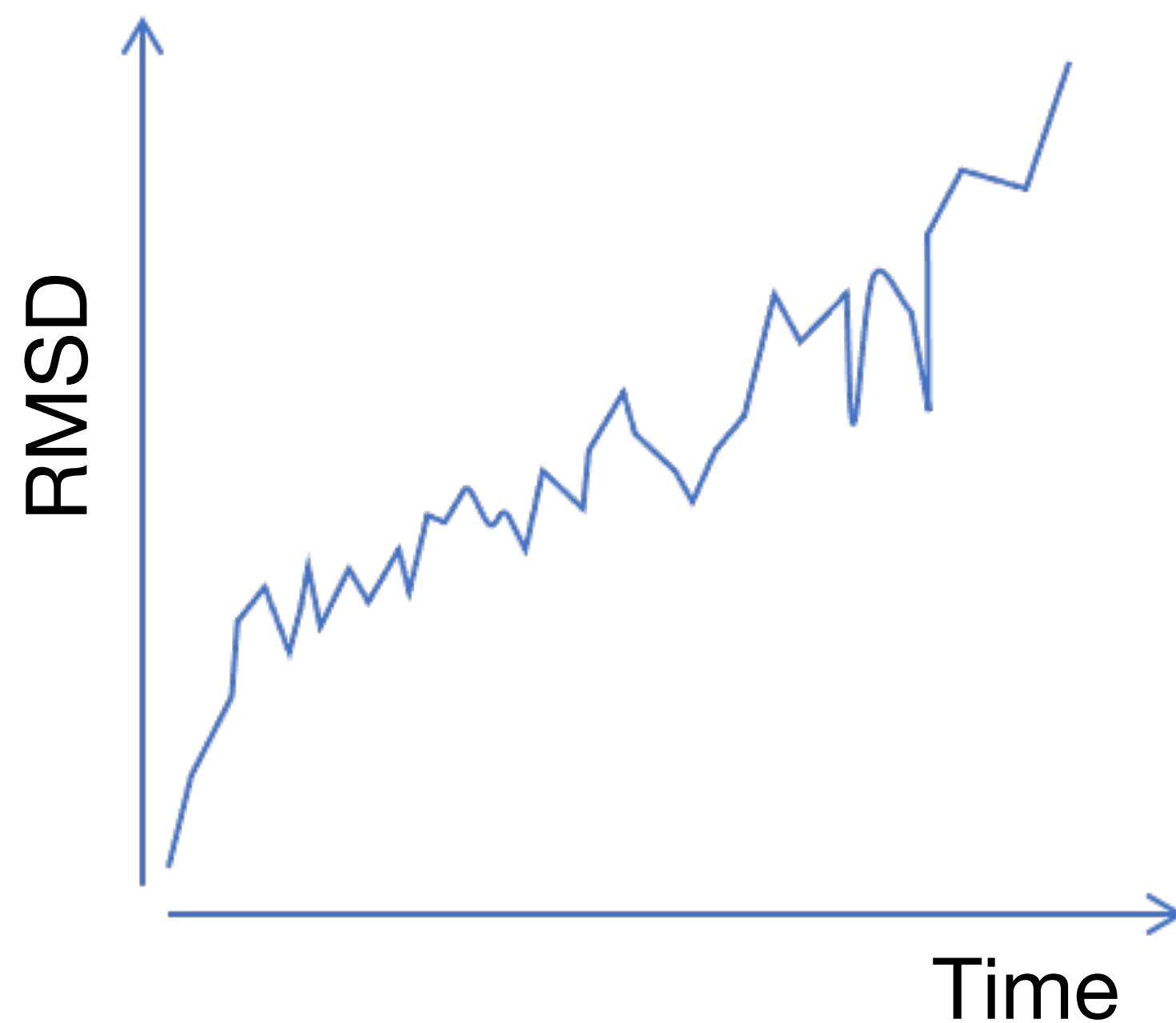
An MD walker over a potential energy surface



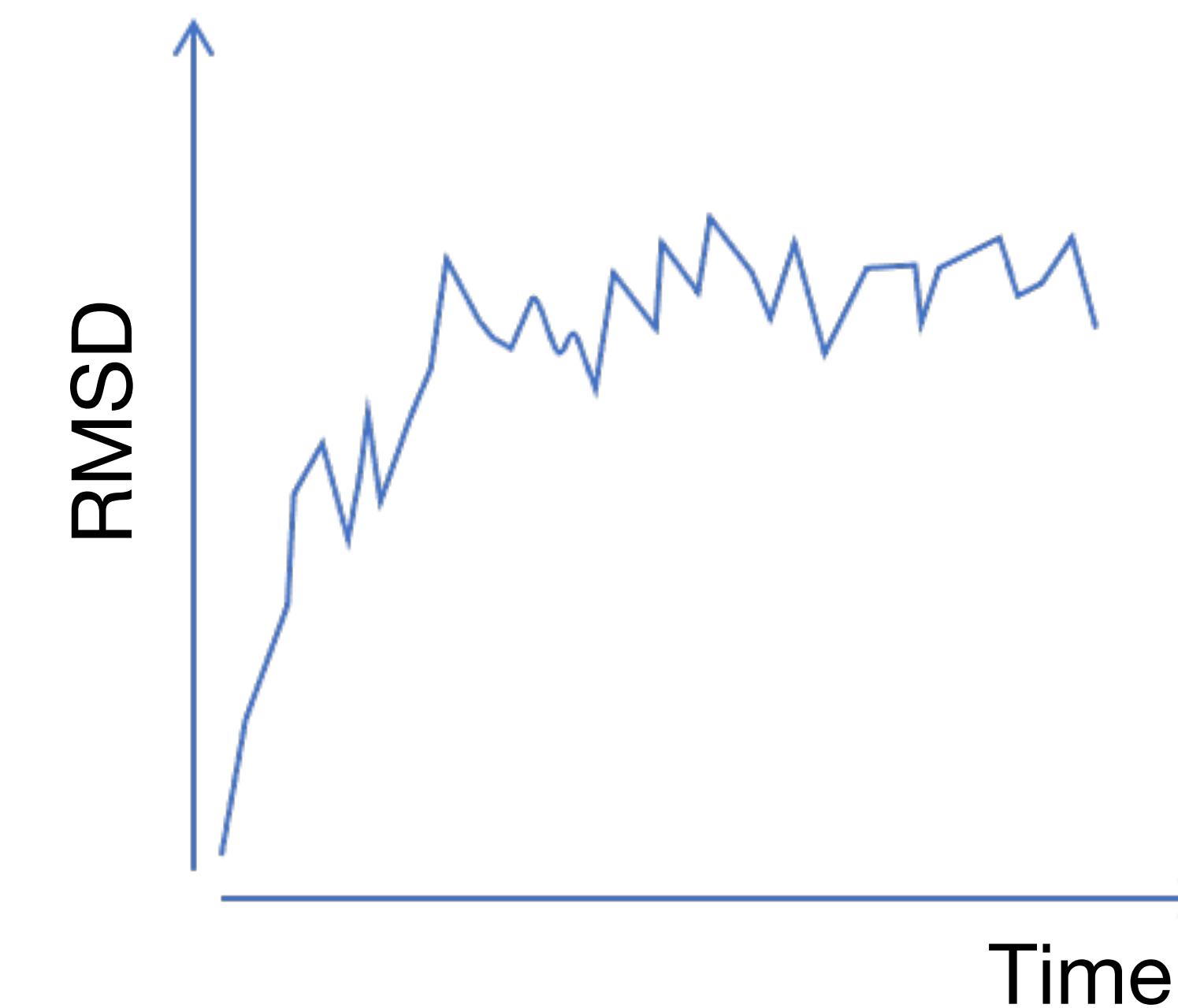
An MD walker over a potential energy surface



RMSD plots are of limited value



This is definitely bad



This is not necessarily good

How to best assess equilibration then?

Equilibration phase:
is the system in a
“relaxed” state?

Production phase: do
we have good sampling
and convergence?

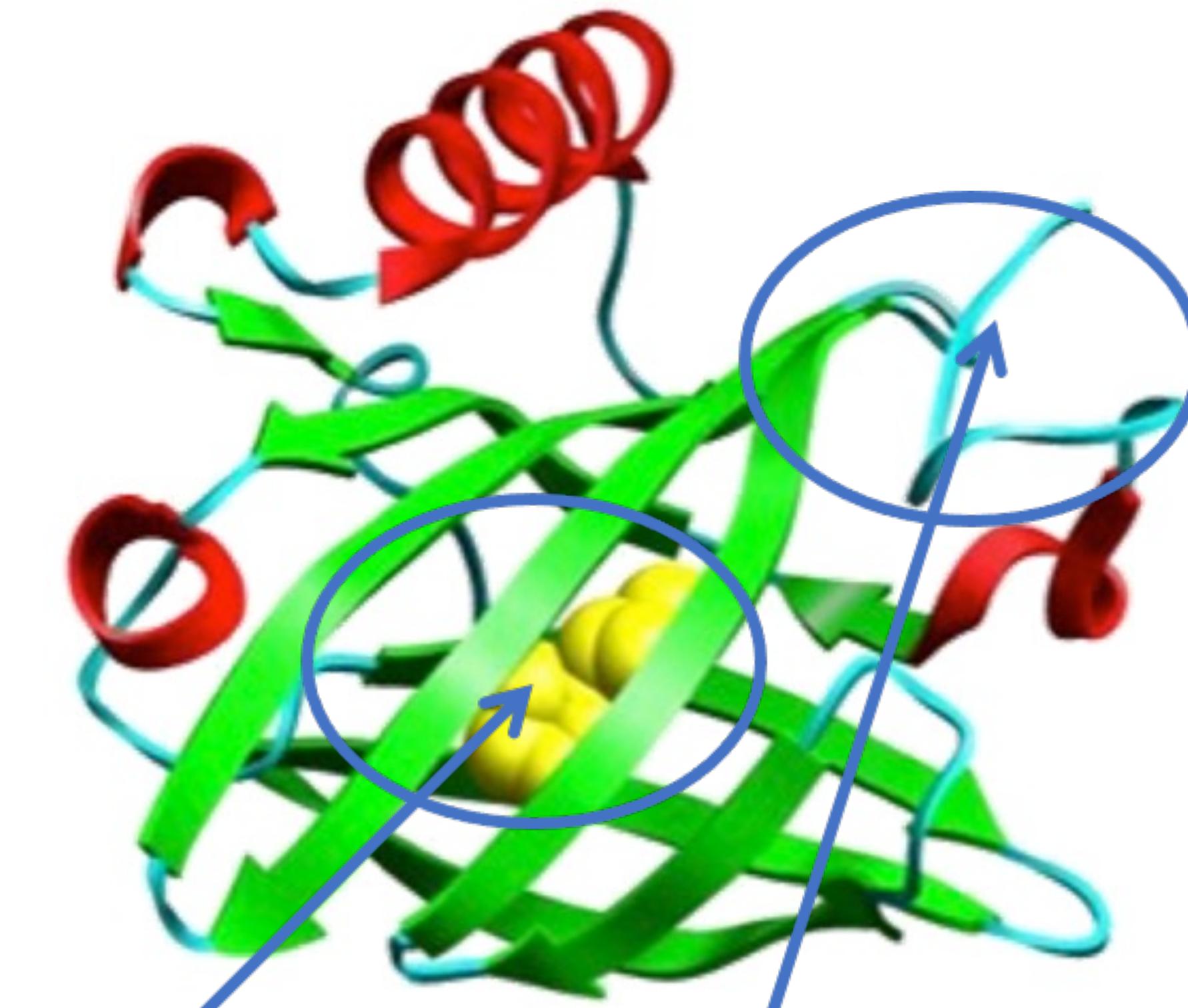


Thinking about the problem holistically: an integrated framework for the analysis of equilibration, sampling, and convergence.

Not all parts of a protein relax in the same way

Typically:

- The core relaxes faster than the surface
- The main chain relaxes faster than side chains
- Helices and sheets relax faster than unstructured elements



If you are interested in this

You *may* not need to worry so much about this

Use metrics of interest to assess simulation

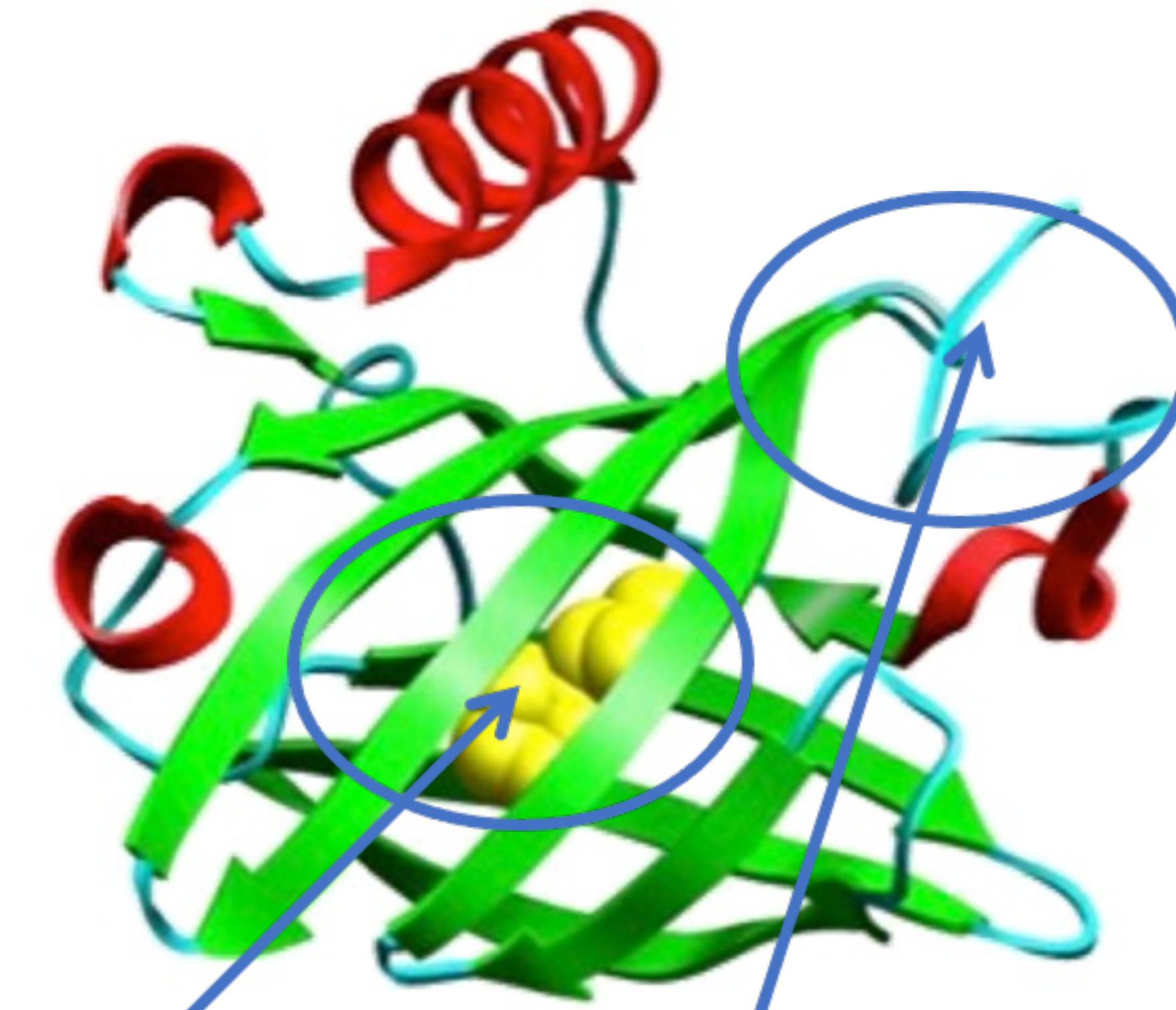
$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=0}^N (v_i - w_i)^2}$$

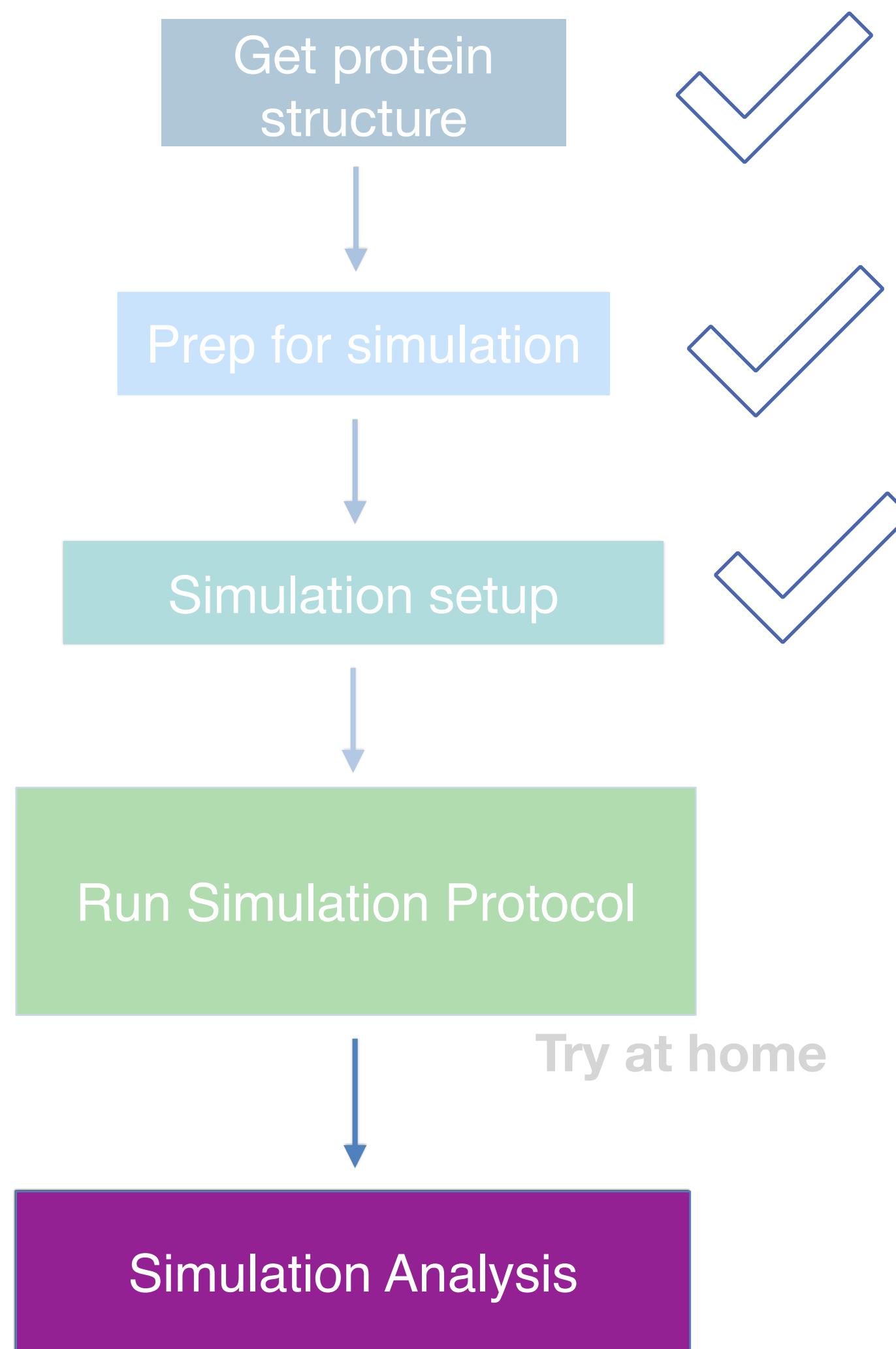
$$R_g(x) = \sqrt{\frac{\sum_i m_i r_i(y)^2 + m_i r_i(z)^2}{\sum_i m_i}}$$

Average distance between atoms...

If you are interested in this

You *may* not need to worry so much about this





Part III: Introduction to MDAnalysis for simulation analysis