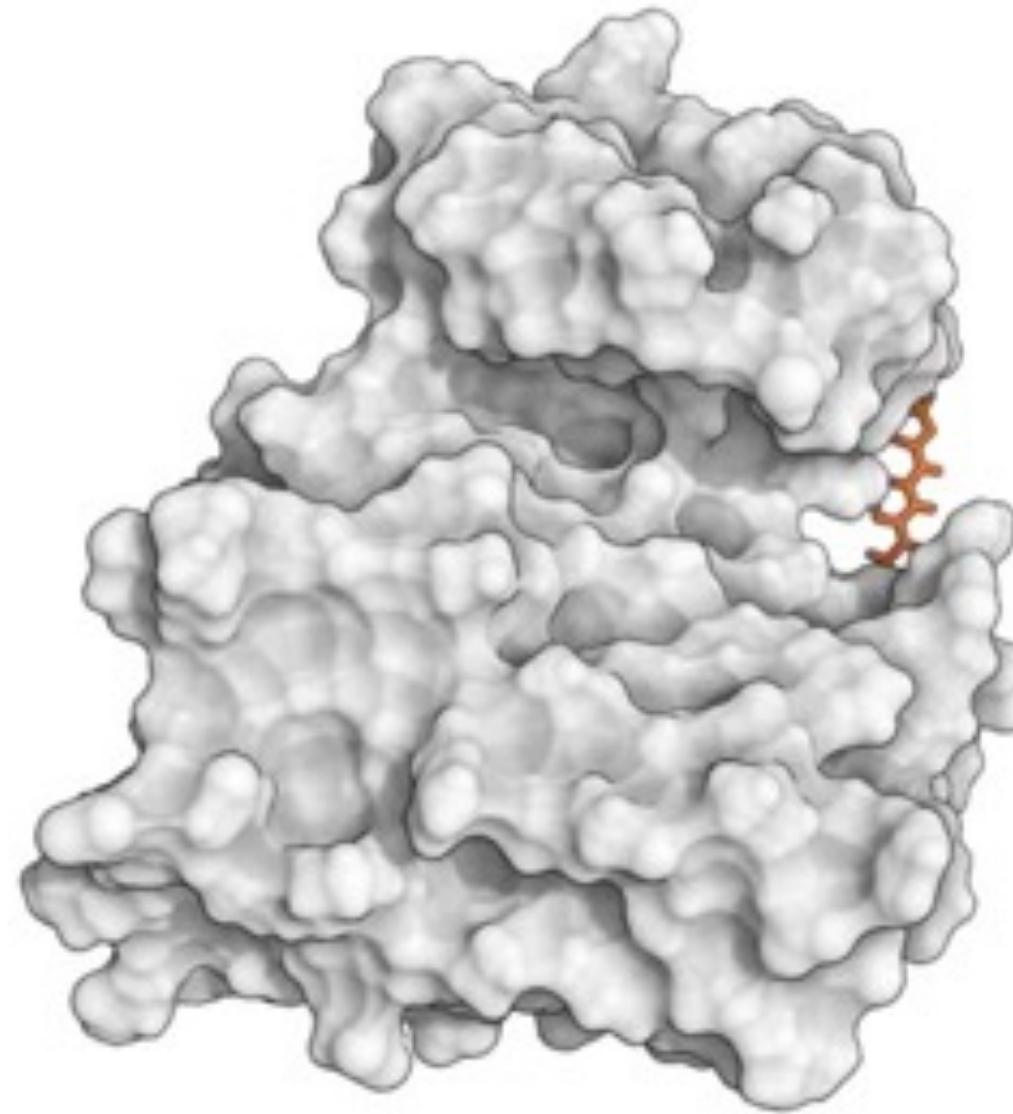


# Simulation of Biomolecules

## Setting up a protein simulation



Dr Matteo Degiacomi

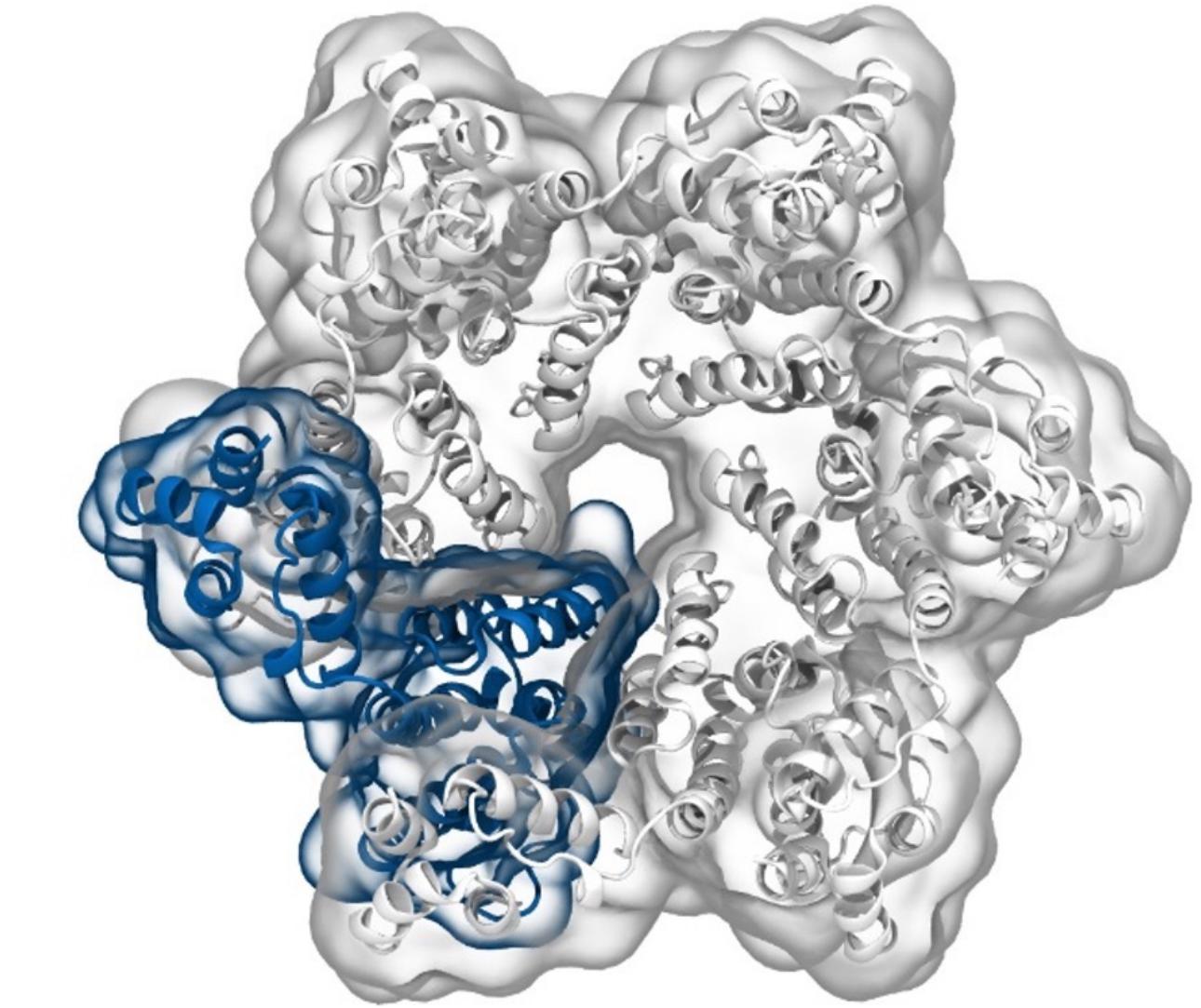
Durham University

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

Dr Antonia Mey

University of Edinburgh

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



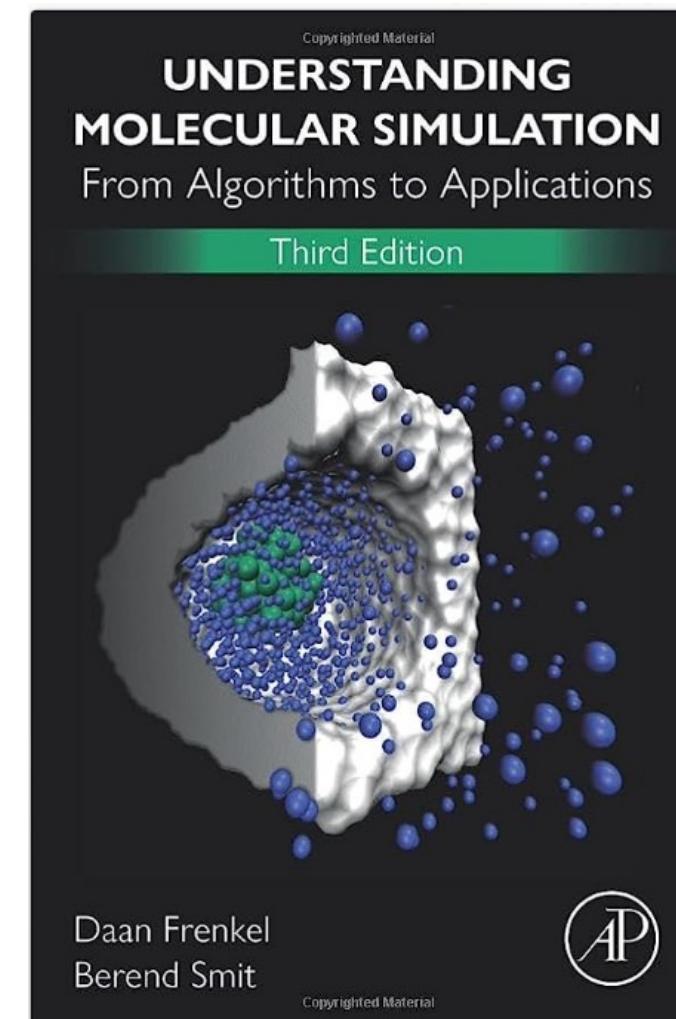
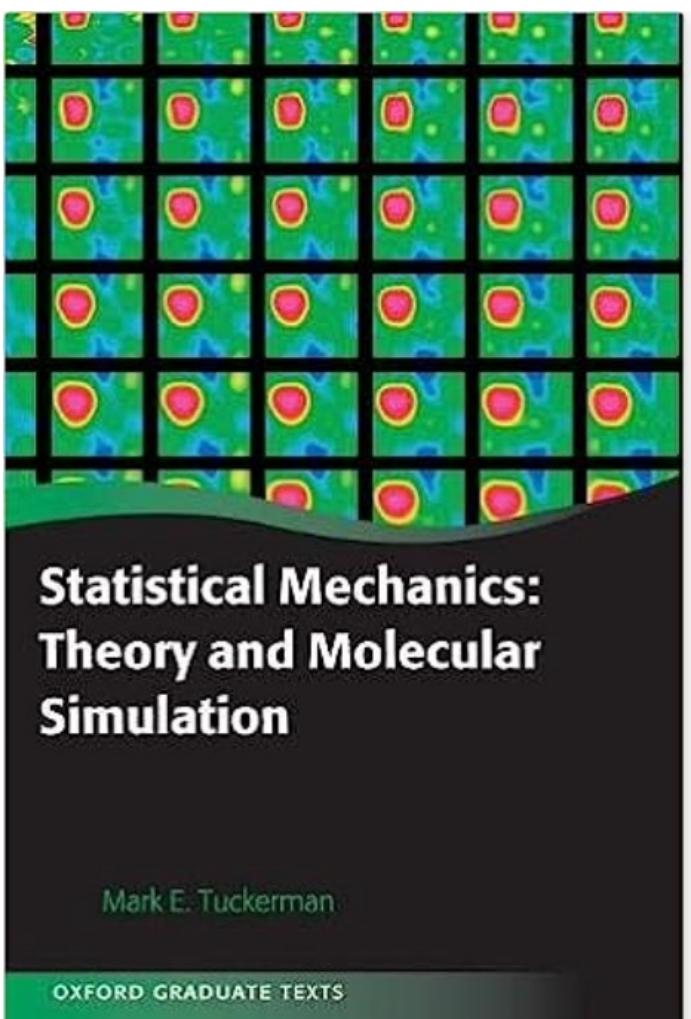
# Useful resources to learn running simulations

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

Efrem Braun<sup>1</sup>, Justin Gilmer<sup>2</sup>, Heather B. Mayes<sup>3</sup>, David L. Mobley<sup>4</sup>, Jacob I. Monroe<sup>5</sup>, Samarjeet Prasad<sup>6</sup>, Daniel M. Zuckerman<sup>7</sup>

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

Lester O. Hedges<sup>1,2\*</sup>, Sofia Bariami<sup>3†</sup>, Matthew Burman<sup>2</sup>, Finlay Clark<sup>3</sup>, Benjamin P. Cossins<sup>4</sup>, Adele Hardie<sup>3</sup>, Anna M. Herz<sup>3</sup>, Dominykas Lukauskis<sup>5</sup>, Antonia S.J.S. Mey<sup>3</sup>, Julien Michel<sup>2,3\*</sup>, Jenke Scheen<sup>3‡</sup>, Miroslav Suruzhon<sup>4</sup>, Christopher J. Woods<sup>1</sup>, Zhiyi Wu<sup>4</sup>



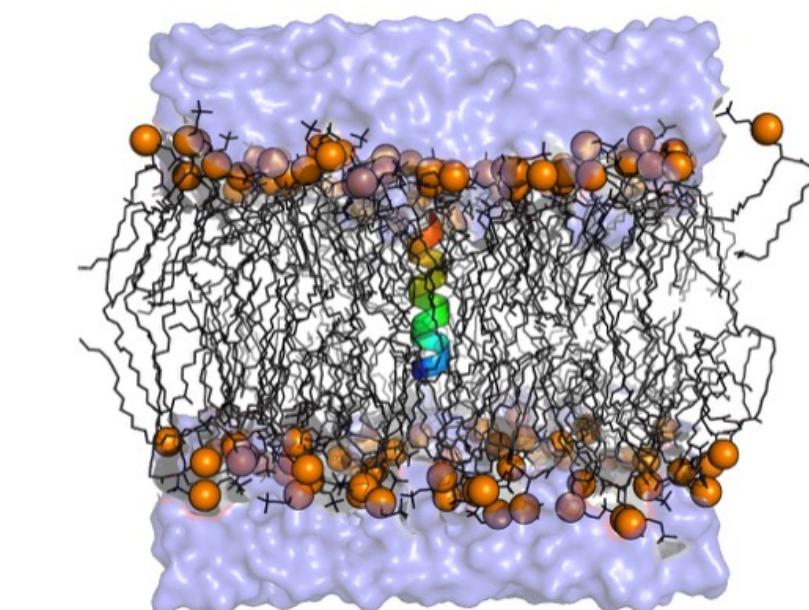
## From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]

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

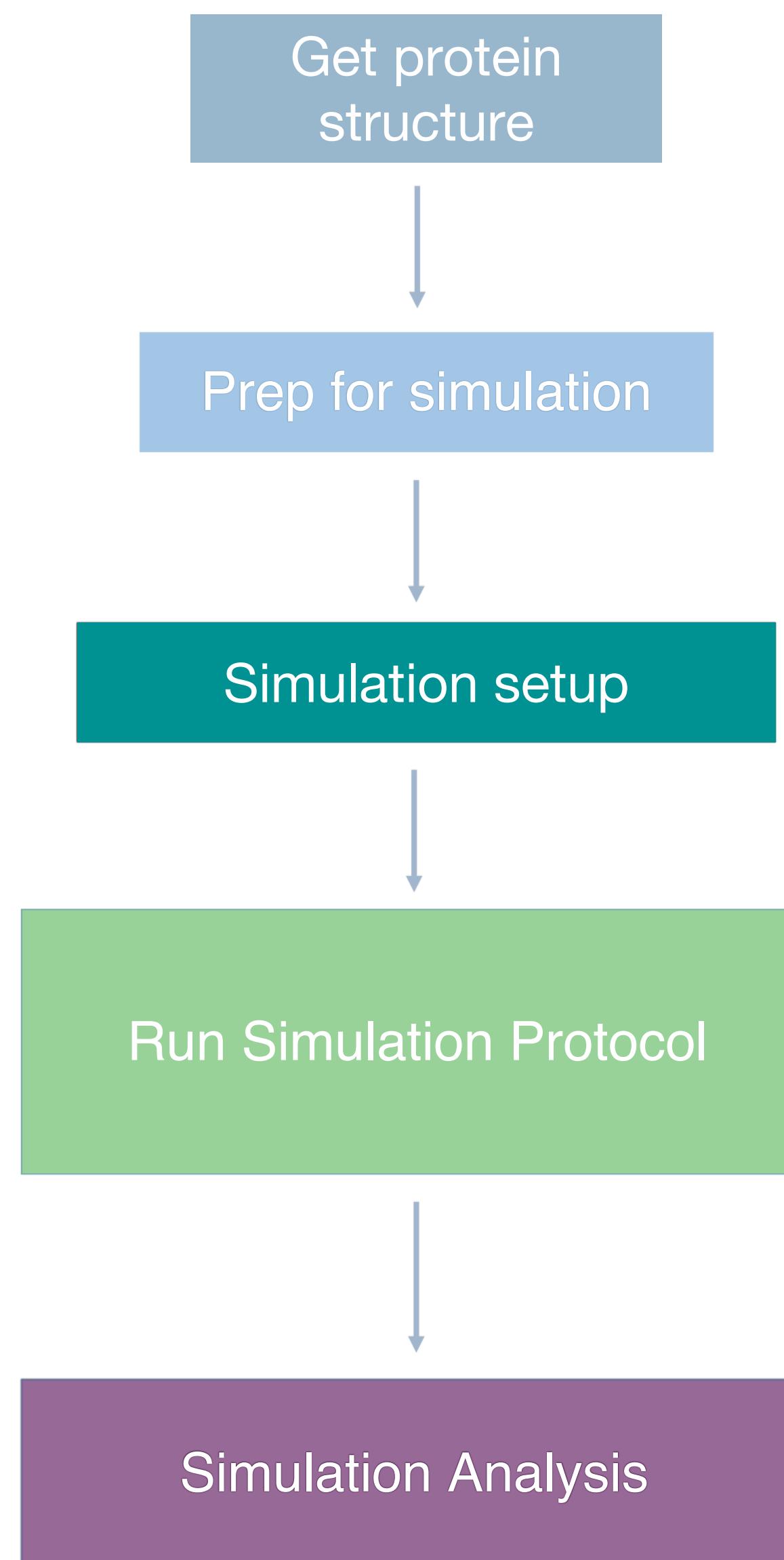
ARTICLE CODE REPOSITORY

GROMACS: [tutorials.gromacs.org](https://tutorials.gromacs.org)

Amber: [ambermd.org/tutorials](https://ambermd.org/tutorials)

OpenMM: [docs.openmm.org/latest/userguide/library/03\\_tutorials.html](https://docs.openmm.org/latest/userguide/library/03_tutorials.html)

# A typical workflow for molecular dynamics



## Getting your protein structure

AlphaFold Protein Structure Database

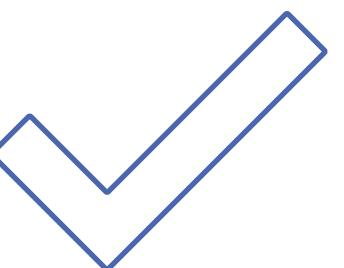
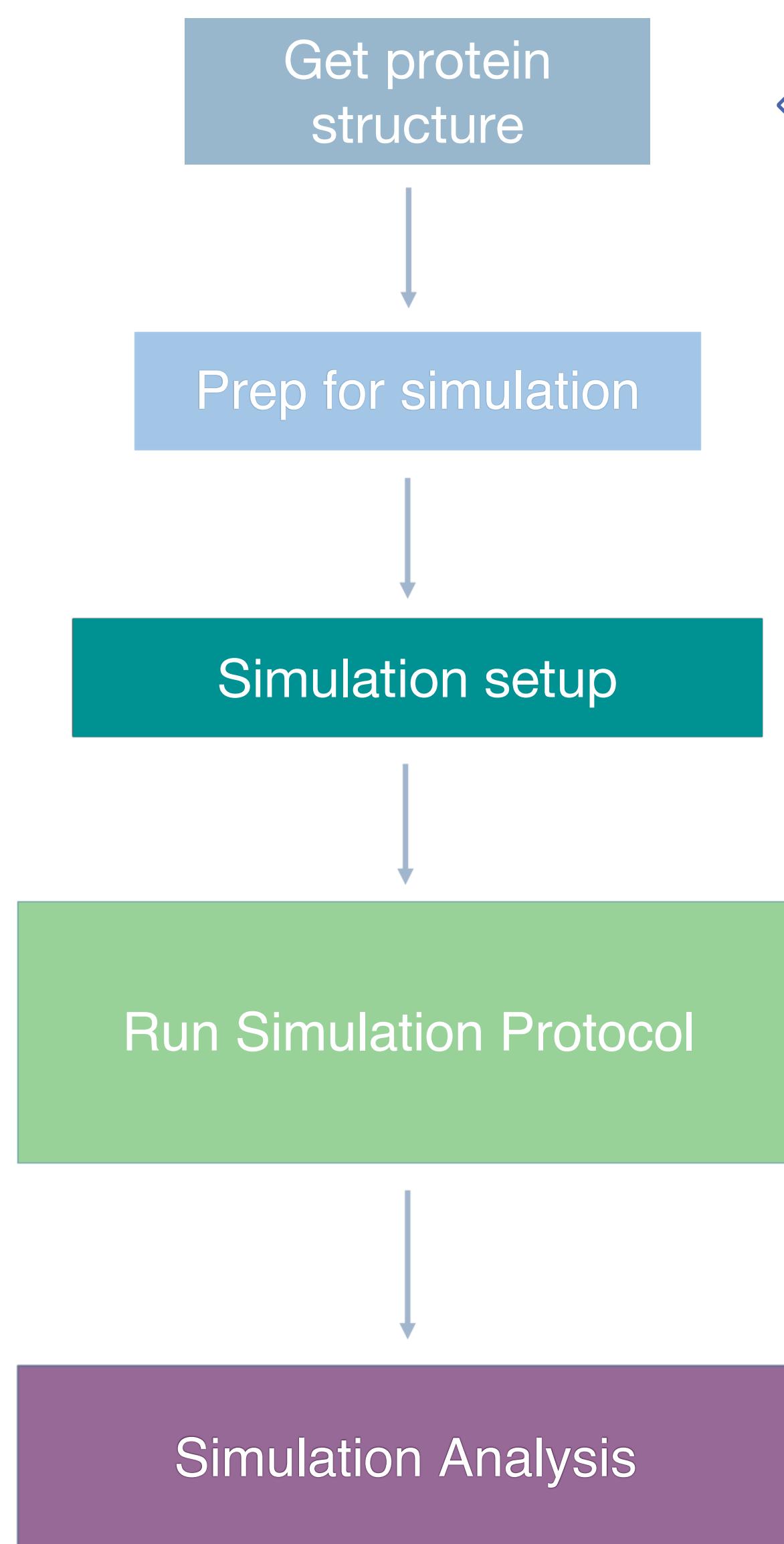


## Getting ligands/co-factors

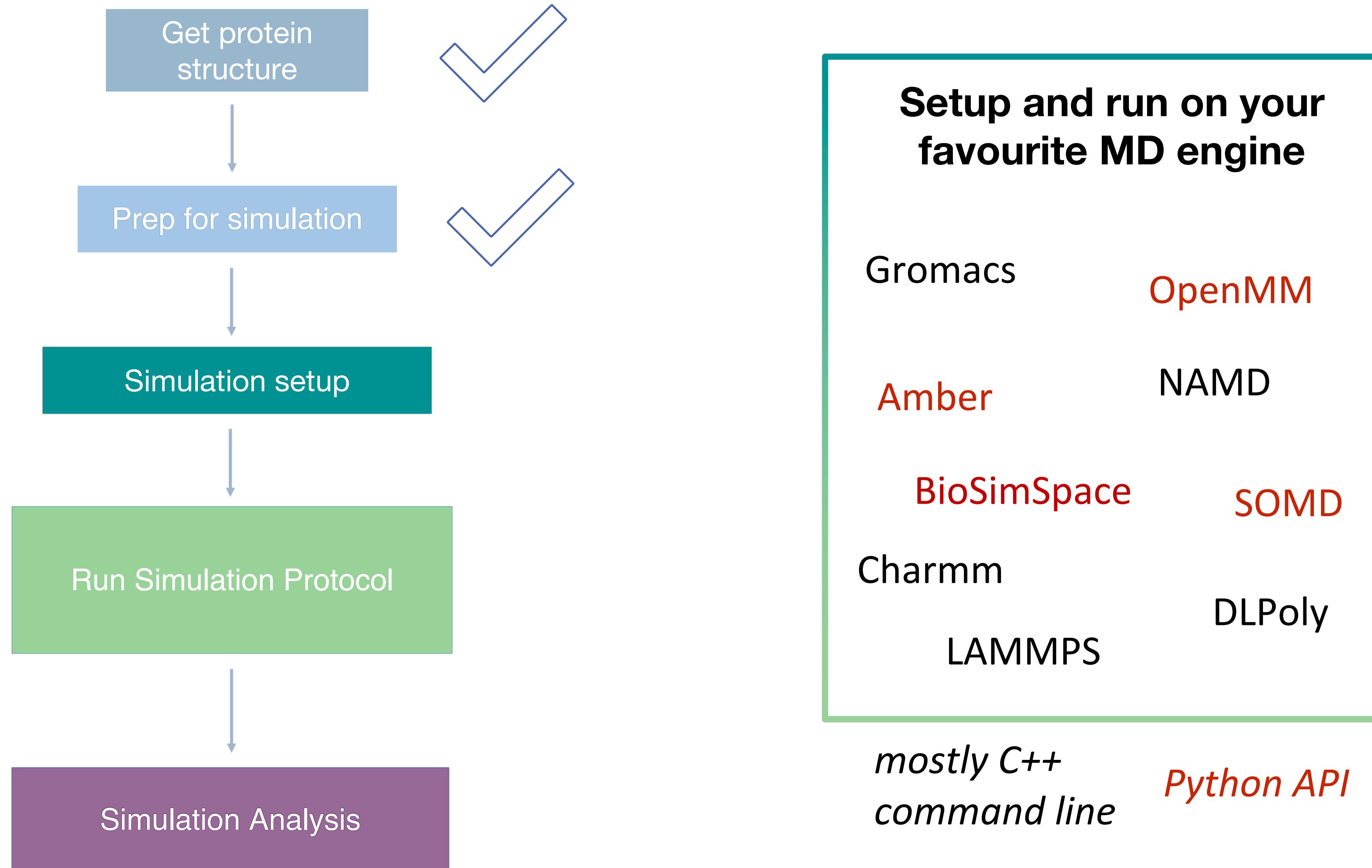
ZINC20



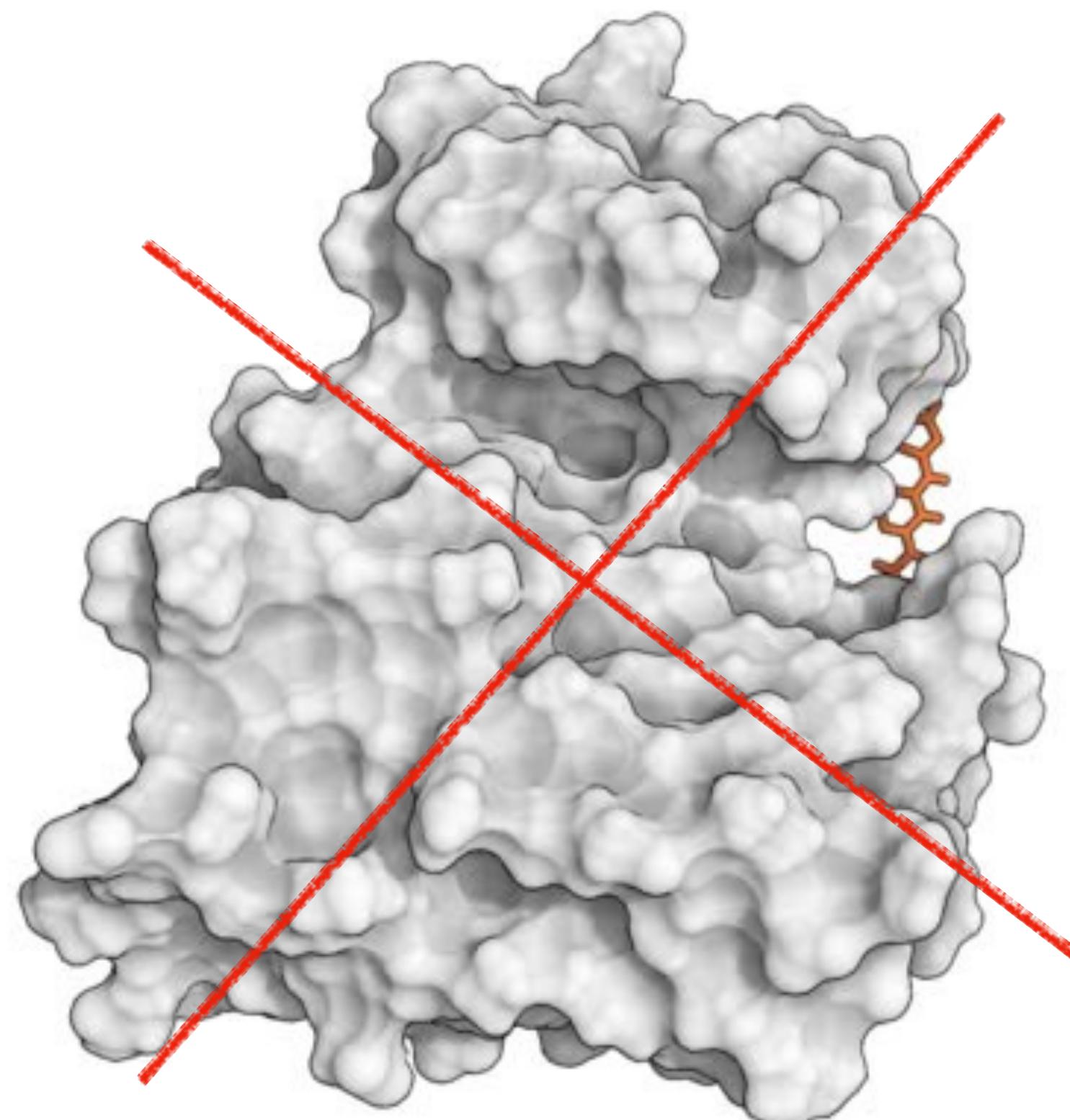
# A typical workflow for molecular dynamics



# A typical workflow for molecular dynamics



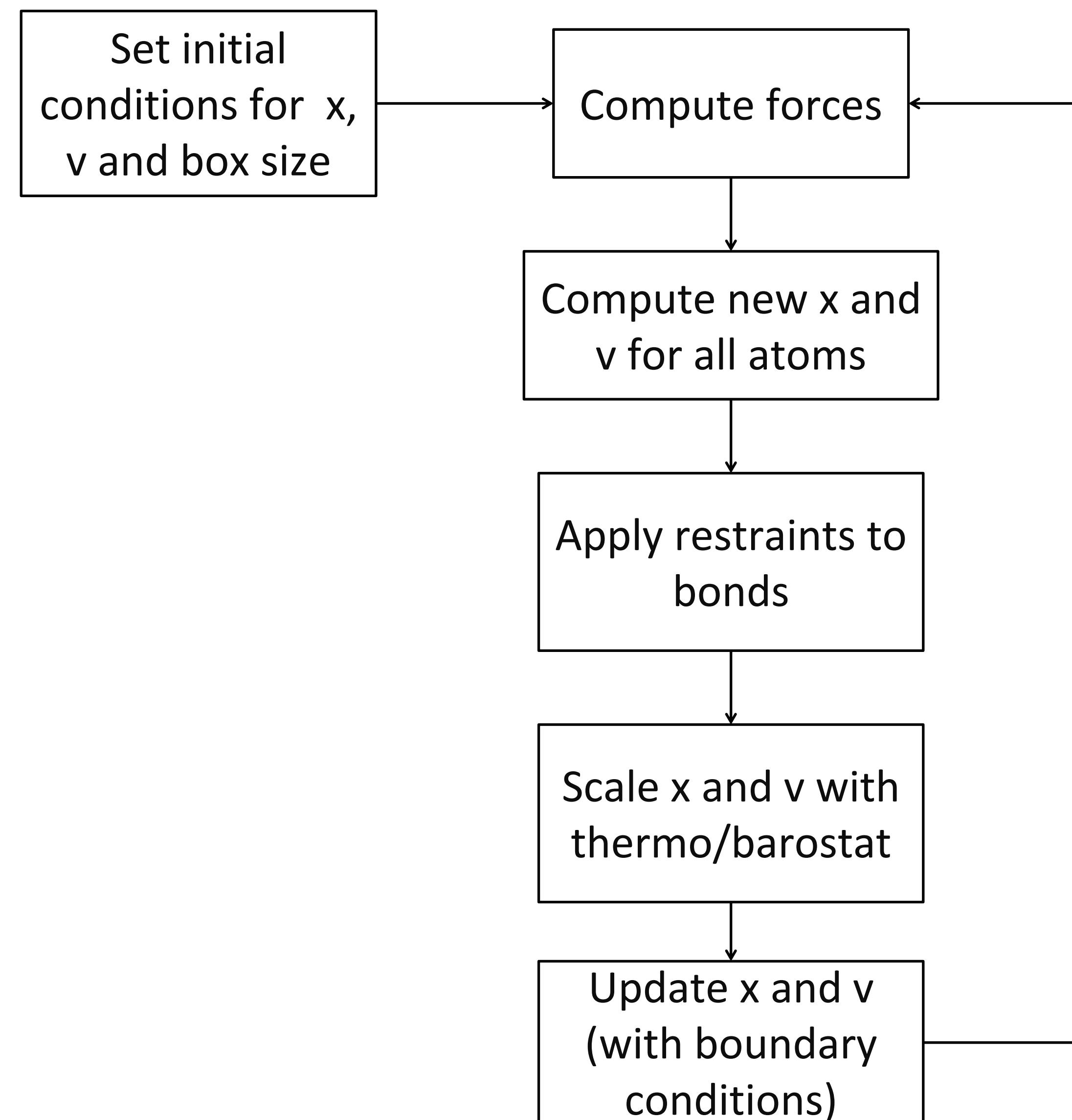
# Disclaimer!



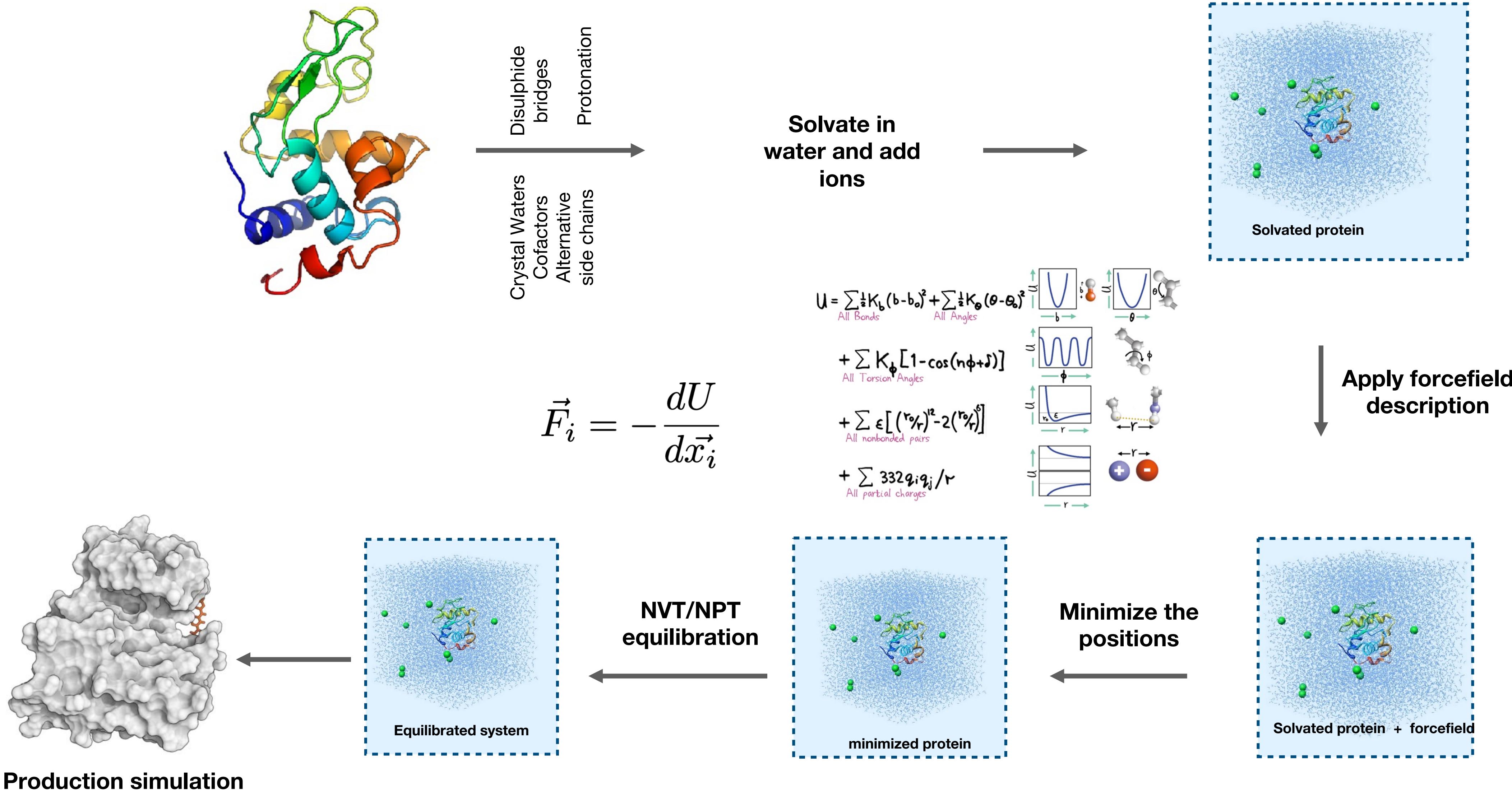
Running biomolecular MD can take days on specialised hardware.

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

# A Molecular Dynamics timestep

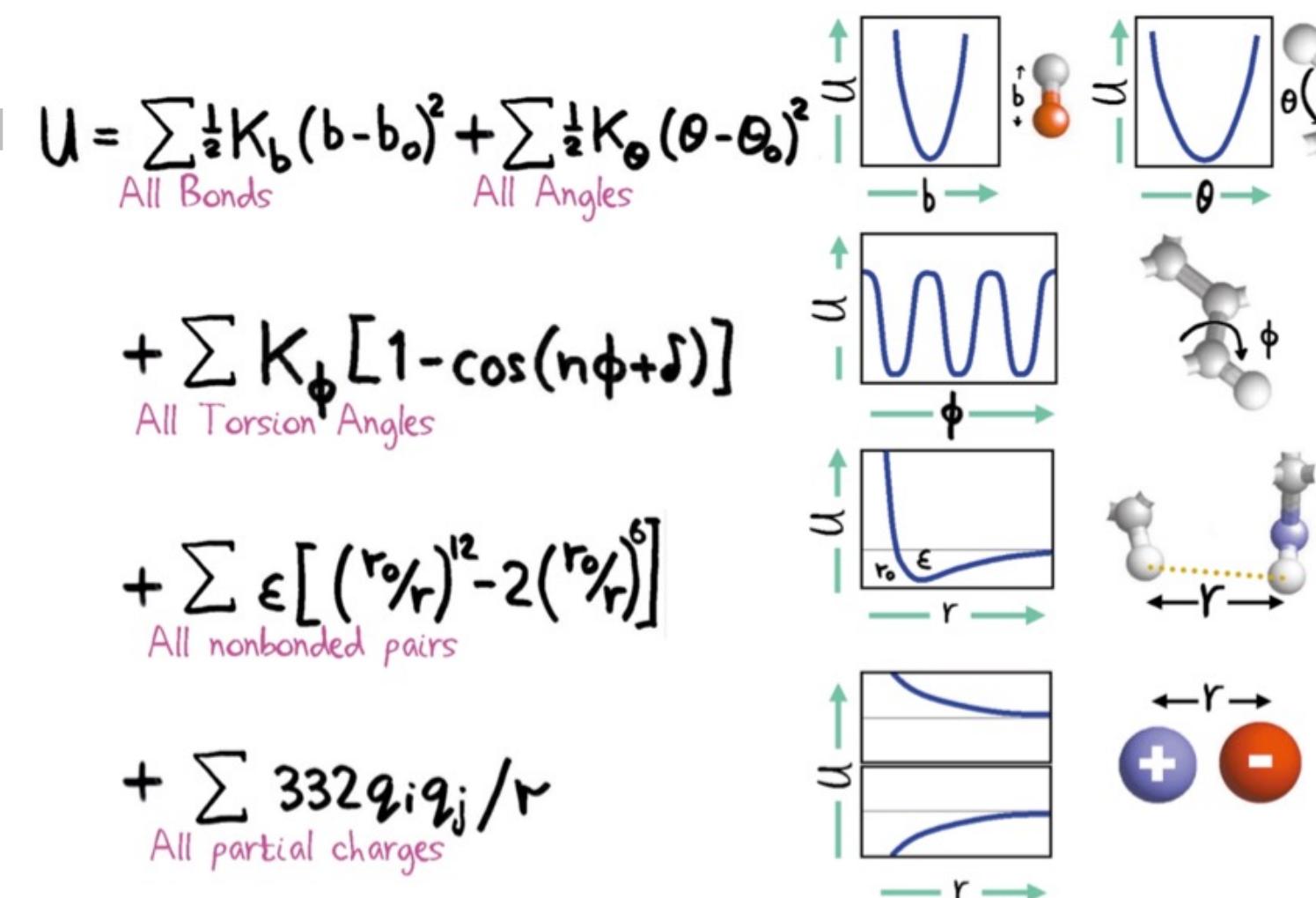


# Molecular dynamics require multiple steps for the setup of simulations



# There are many different choices for force fields to be made

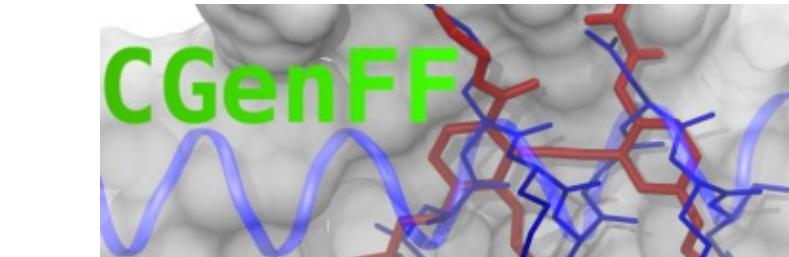
$$\vec{F}_i = -\frac{dU}{d\vec{x}_i}$$



- **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)

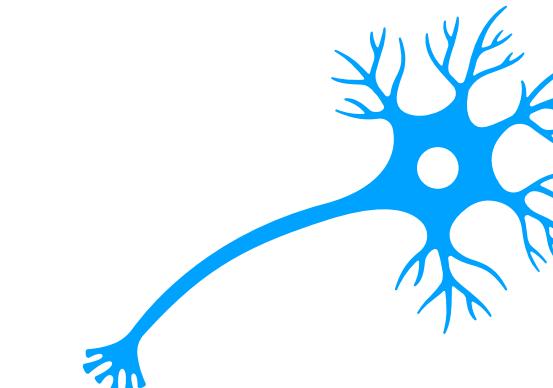
**There is no “best force field”!**

## Small molecule force fields



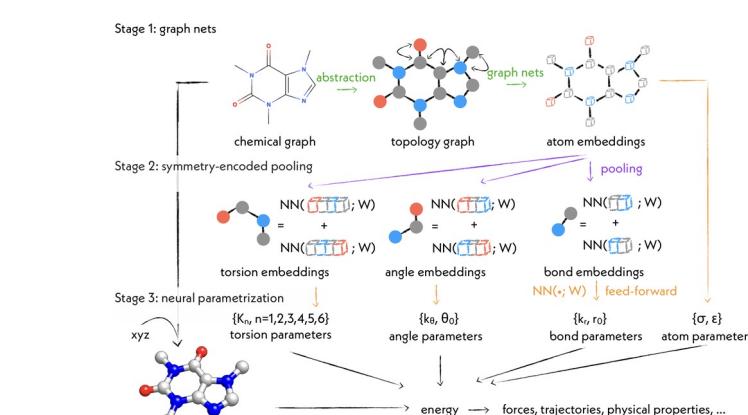
$$E_{\text{pair}} = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

## Machine learned force fields



Espaloma

Ani-2x



SchNet

PhysNet

BandNN

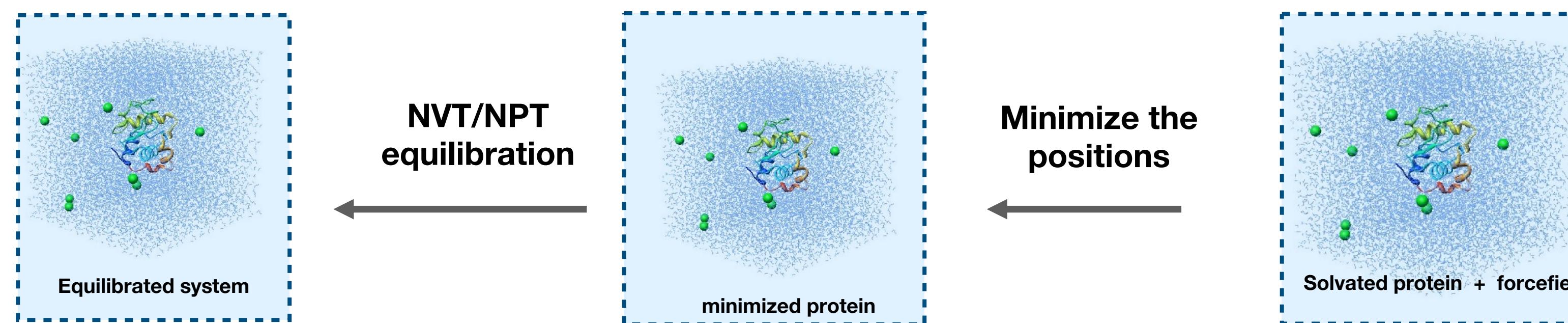
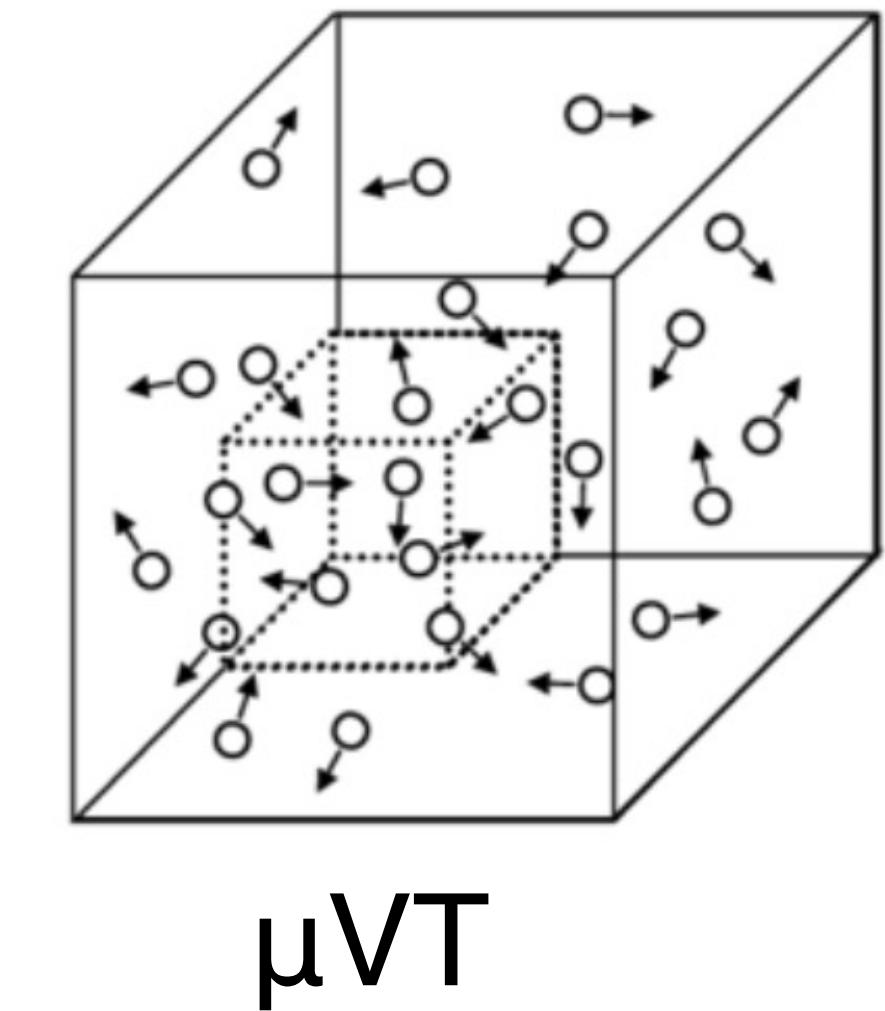
**Coarse-grained force fields...**

# Choosing your thermodynamic ensemble

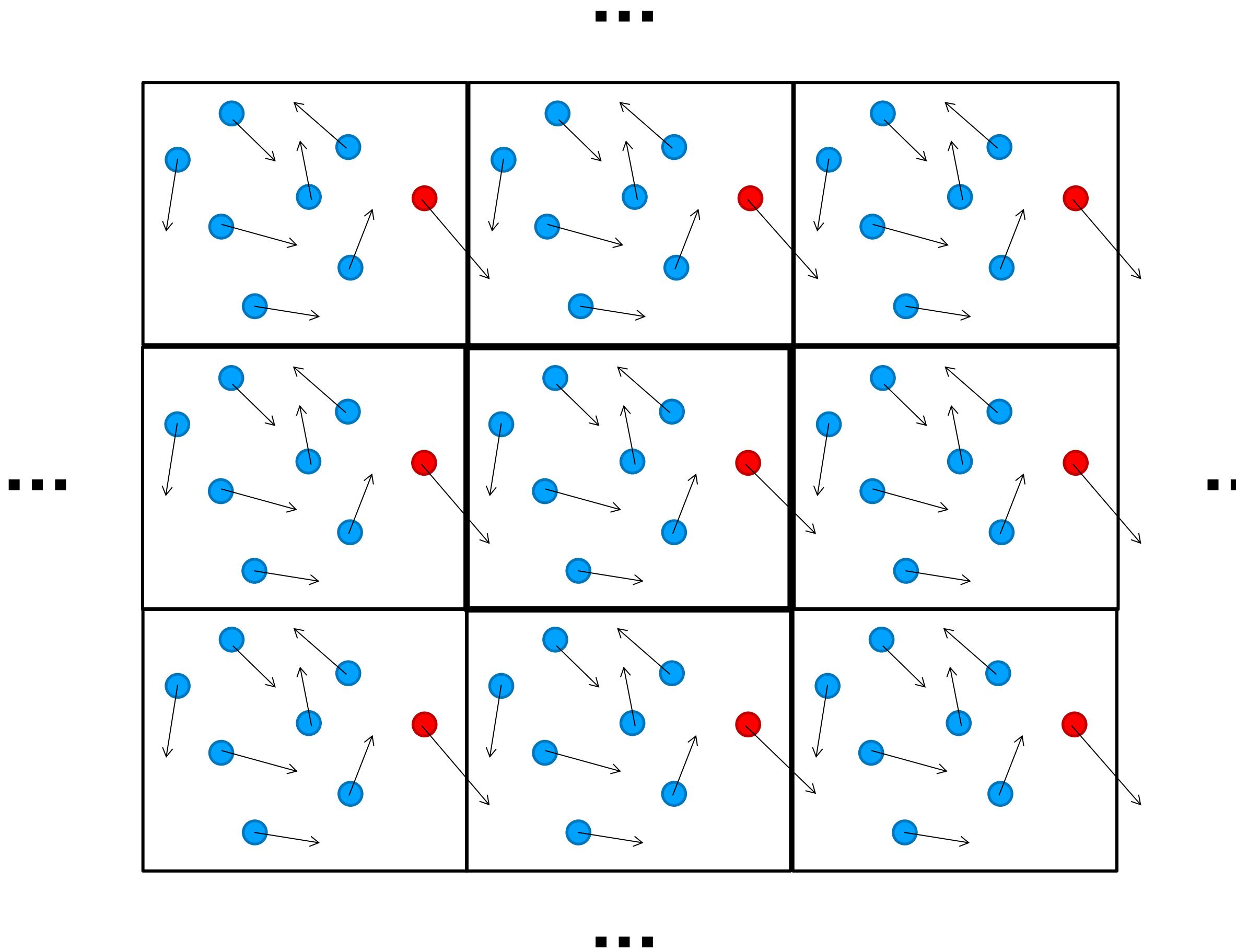
Simulations replicate a specific *thermodynamic ensemble* (typically NVT or NPT), or even grand canonical ( $\mu$ VT)

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

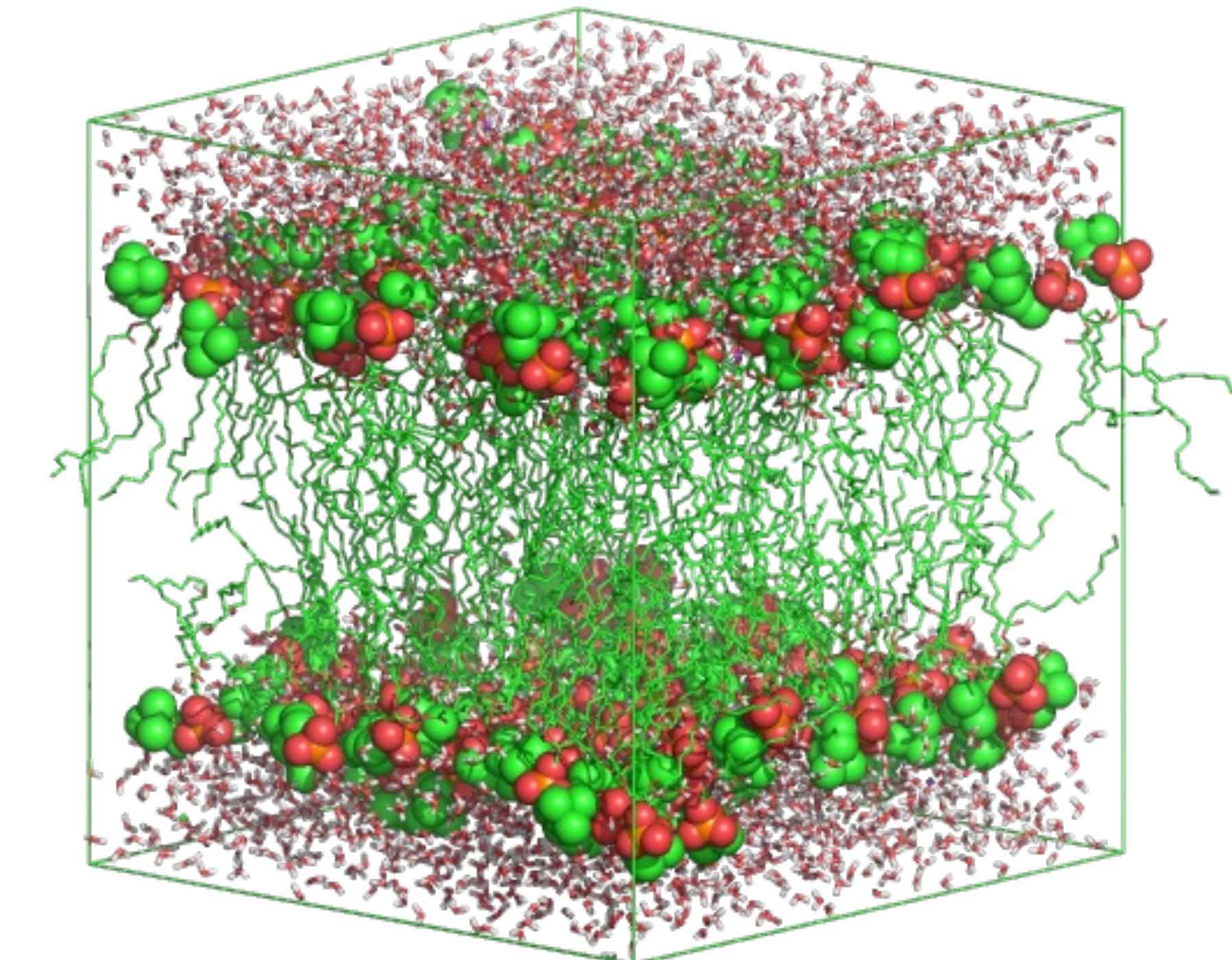
- Nose-Hoover
- Berendsen
- Parrinello-Rahman
- Langevin piston
- ...



# Periodic boundary conditions (PBC) and pressure coupling



Typically, PBC 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:

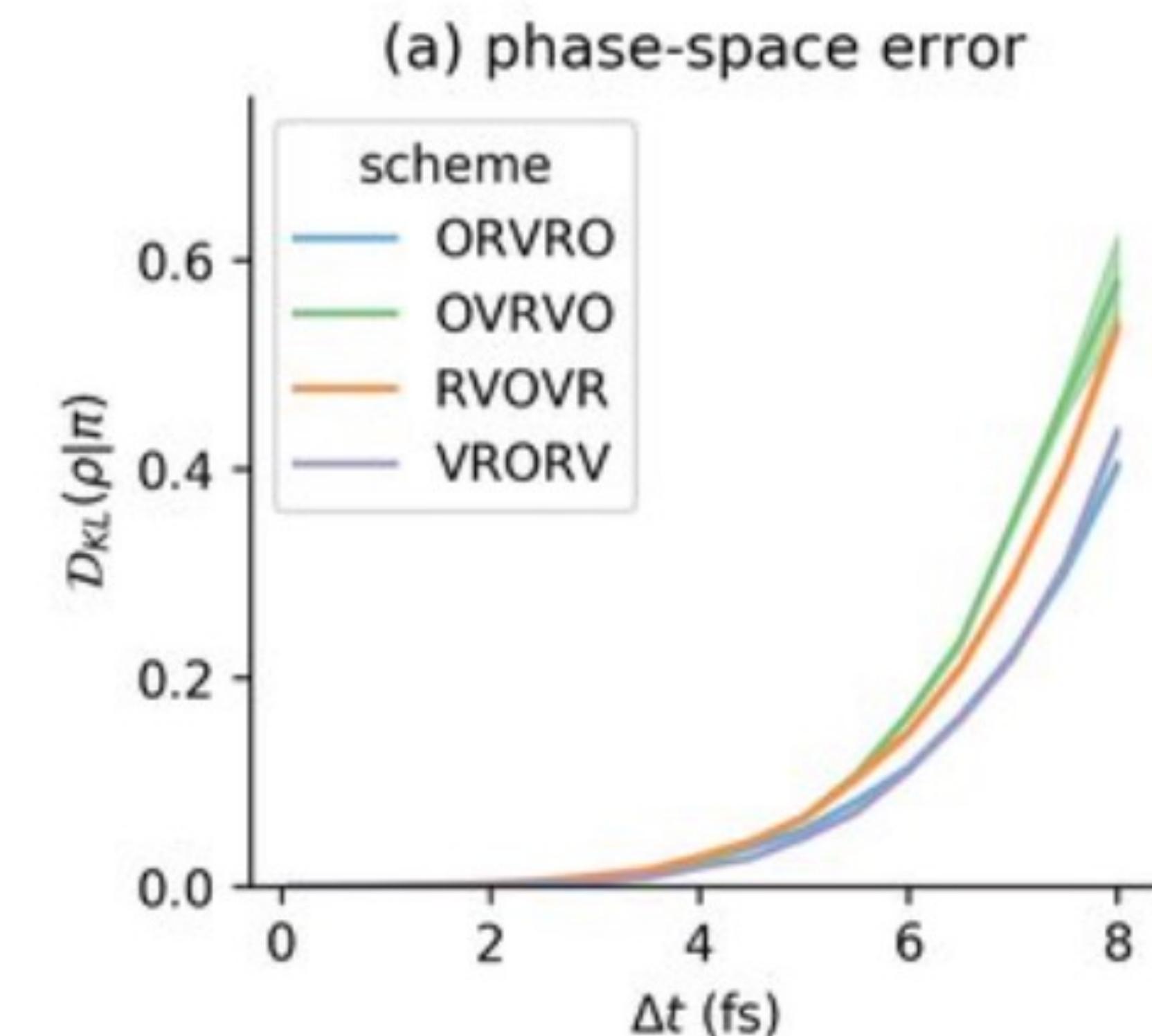
Timestep size is imposed by the fastest phenomenon we want to observe :

- 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 (e.g., Langevin): 4 fs - 6 fs.



# “equilibration” and “convergence”, what do they mean?

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

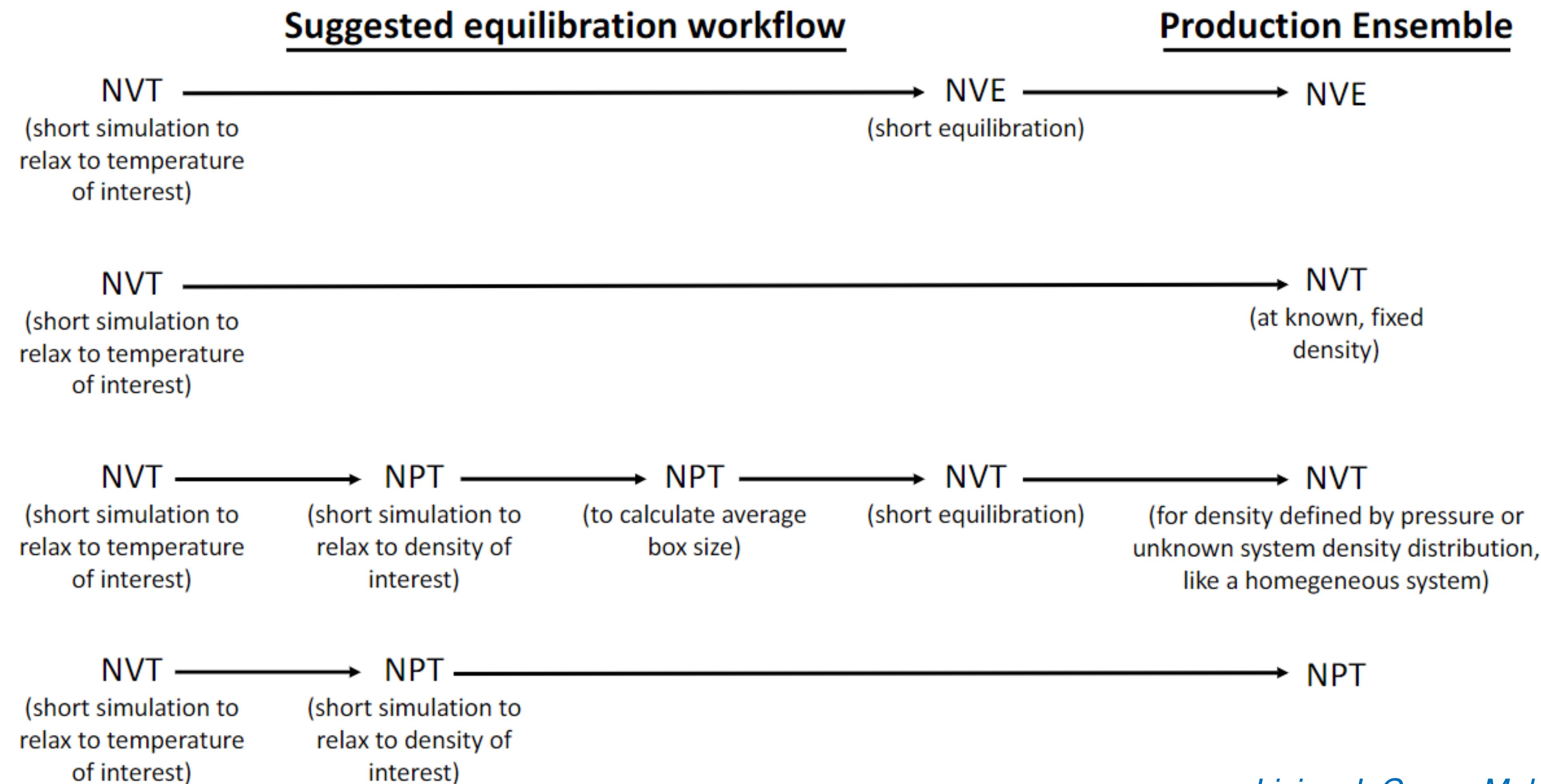
Production phase: do  
we have good sampling  
and convergence?



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

# Example equilibration protocols

**YOU WANT:** constant volume, pressure, and temperature, healthy Ramachandran plot, no exotic chemistry, bulk water (if used)



# An example simulation protocol

**YOU WANT:** constant volume, pressure, and temperature, healthy Ramachandran plot, no exotic chemistry, bulk water (if used)

## Equilibration:

1. Minimize energy, 1000 steepest descent
2. Heat system from 0 to 300 K in 500 ps, NPT,  
Berendsen barostat 1 atm. a-carbon  
restrained with 10 kcal/mol harmonic potential.  
2 fs timestep, LINCS all bonds
3. 1 ns nVT equilibration with Langevin  
dynamics, no atom constrained.

## Production:

4. 1  $\mu$ s 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

# Which MD engine should I use?

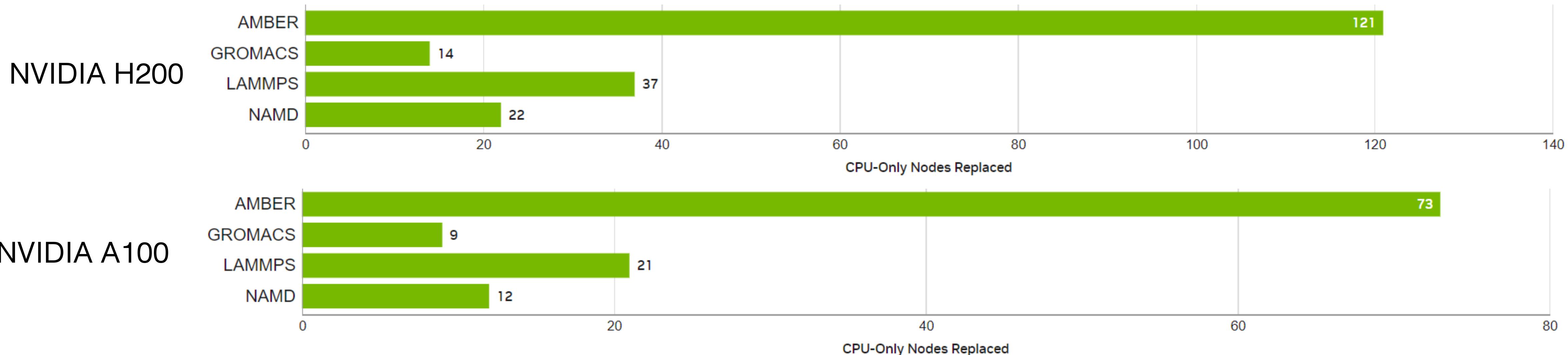
Consider:

- Support for force field of choice
- Enables running desired simulation protocol
- Performance for available hardware
- Ease of use

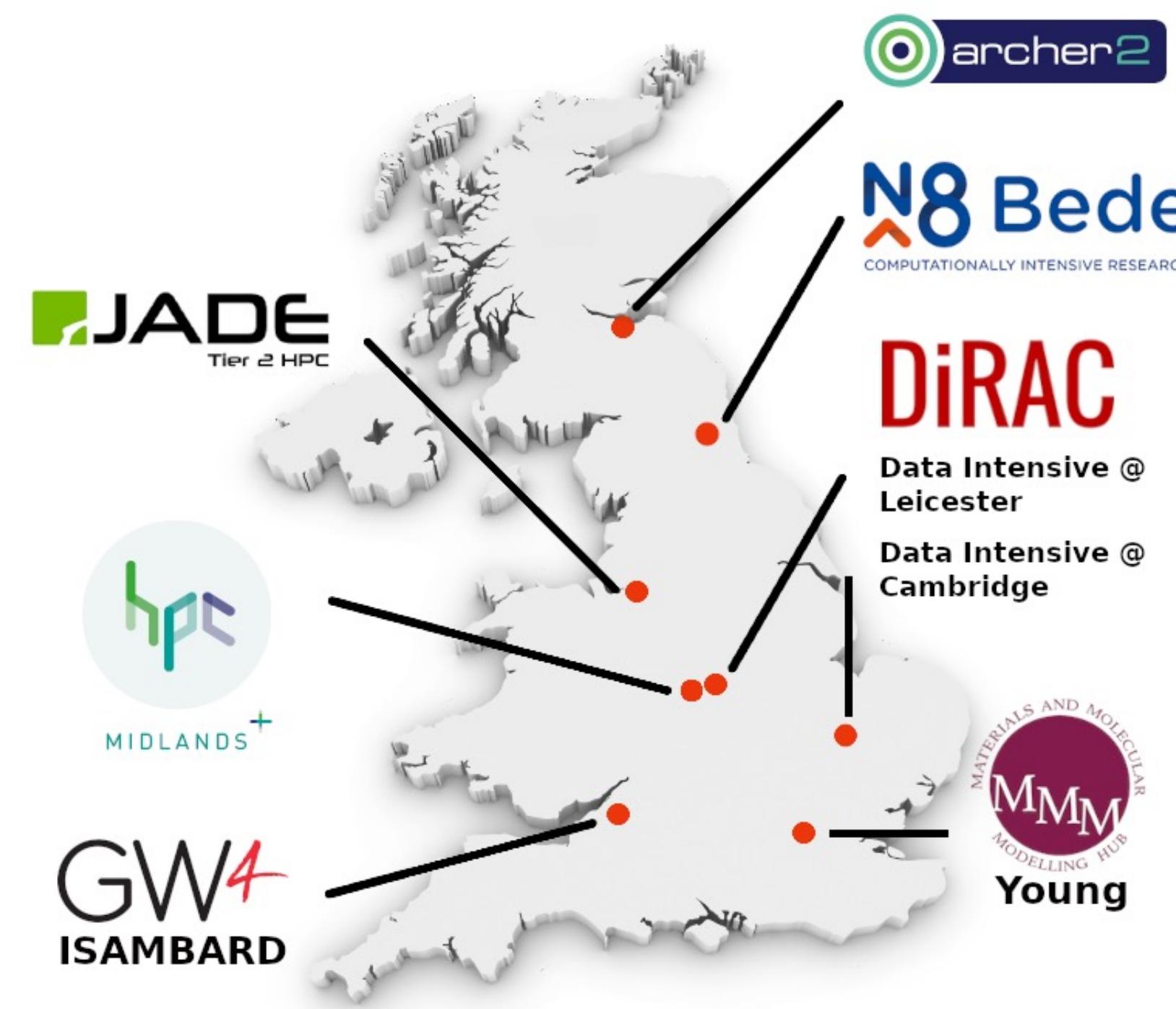
*depends on number of atoms,  
hardware, simulation protocol,  
MD engine*

Graphical Processing Units (GPUs) are especially effective for MD

From: [developer.nvidia.com/hpc-application-performance](https://developer.nvidia.com/hpc-application-performance)



# Calculating runtimes: example on UK Tier 2 systems



The screenshot shows a web browser displaying the HECBioSim HPC calculator at <https://www.hecbiosim.ac.uk/access-hpc/hpc-calculator>. The page features a banner with the HECBioSim logo and a molecular simulation image. The navigation menu includes Home, Contact Us, Covid-19, Events, Access HPC (selected), Highlights, Publications, Downloads, and About Us.

## The HECTime Calculator

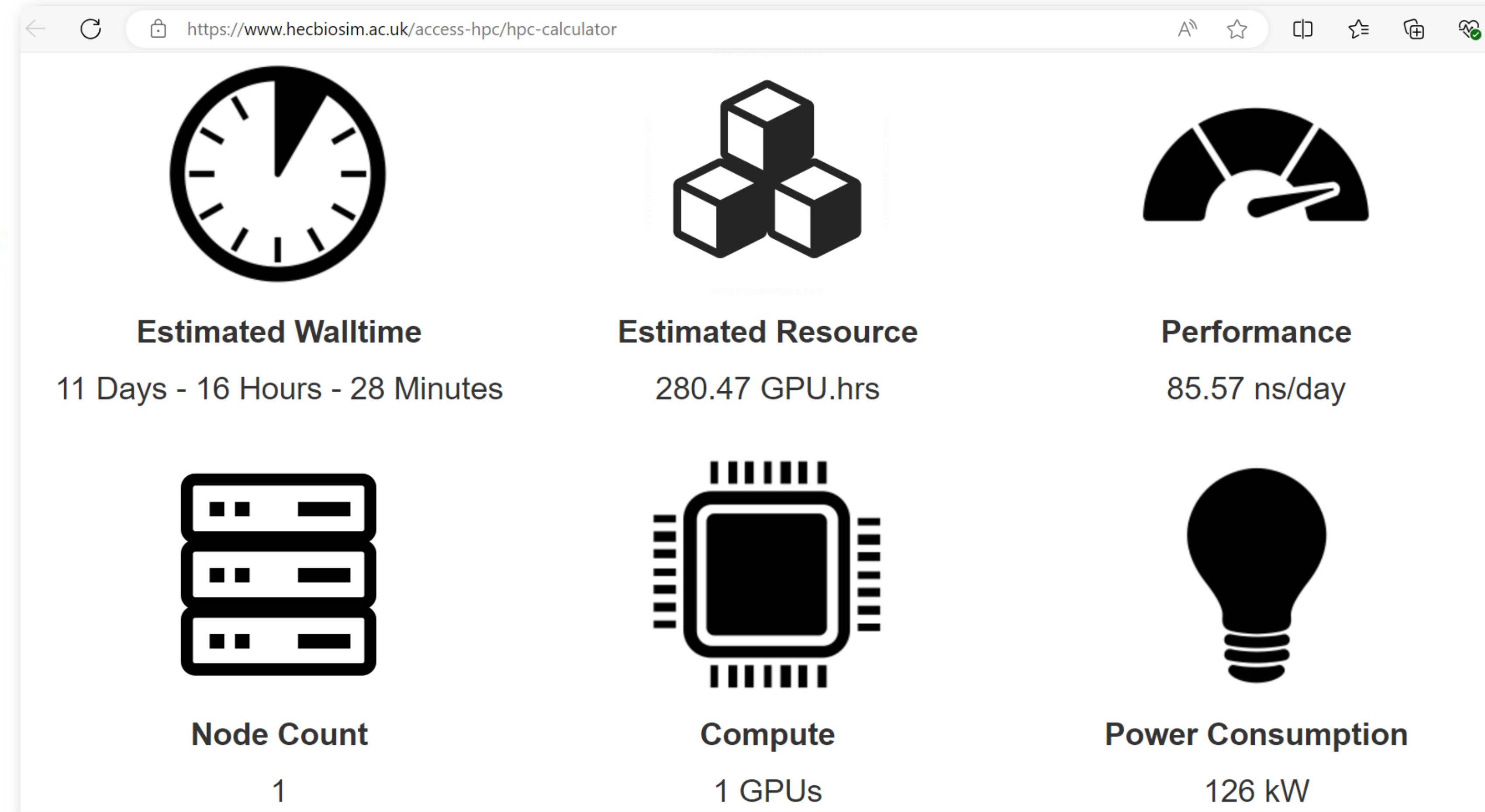
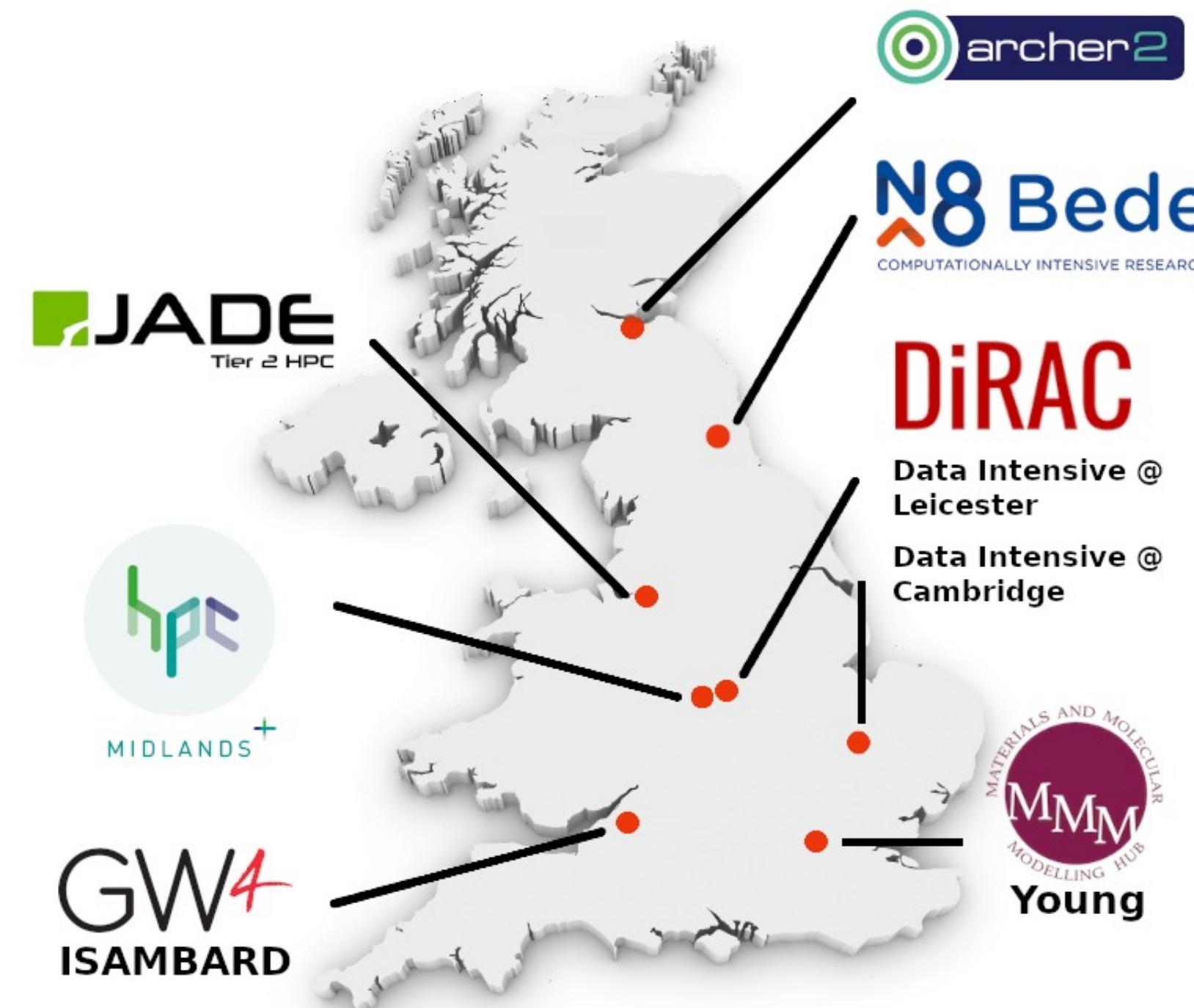
Enter the following information about your simulation:

JADE2  
1000  
100000  
GROMACS 2020.4

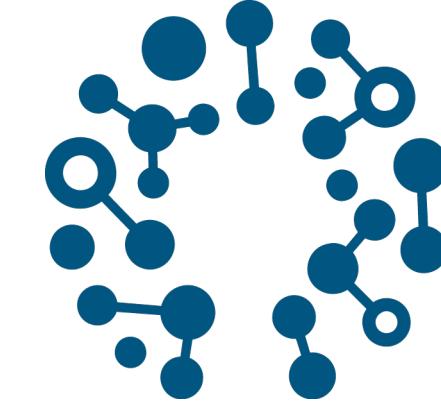
Reset Calculate

<https://www.hecbiosim.ac.uk/access-hpc/hpc-calculator>

# Calculating runtimes: example on UK Tier 2 systems

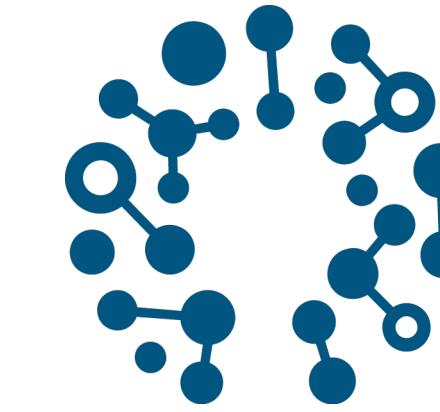


<https://www.hecbiosim.ac.uk/access-hpc/hpc-calculator>



# open forcefield

<https://docs.openforcefield.org/en/latest/examples.html>



open  
forcefield

