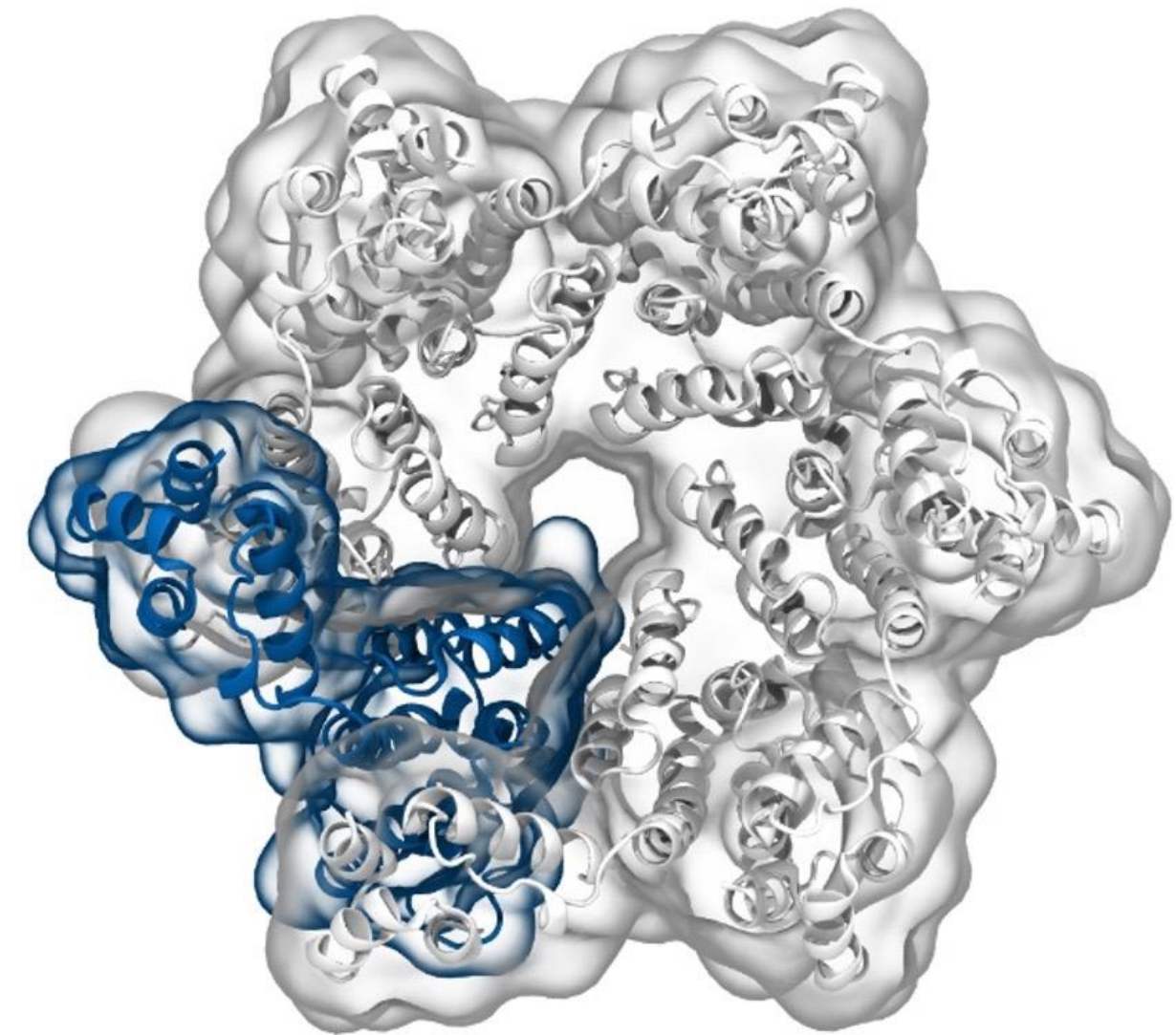
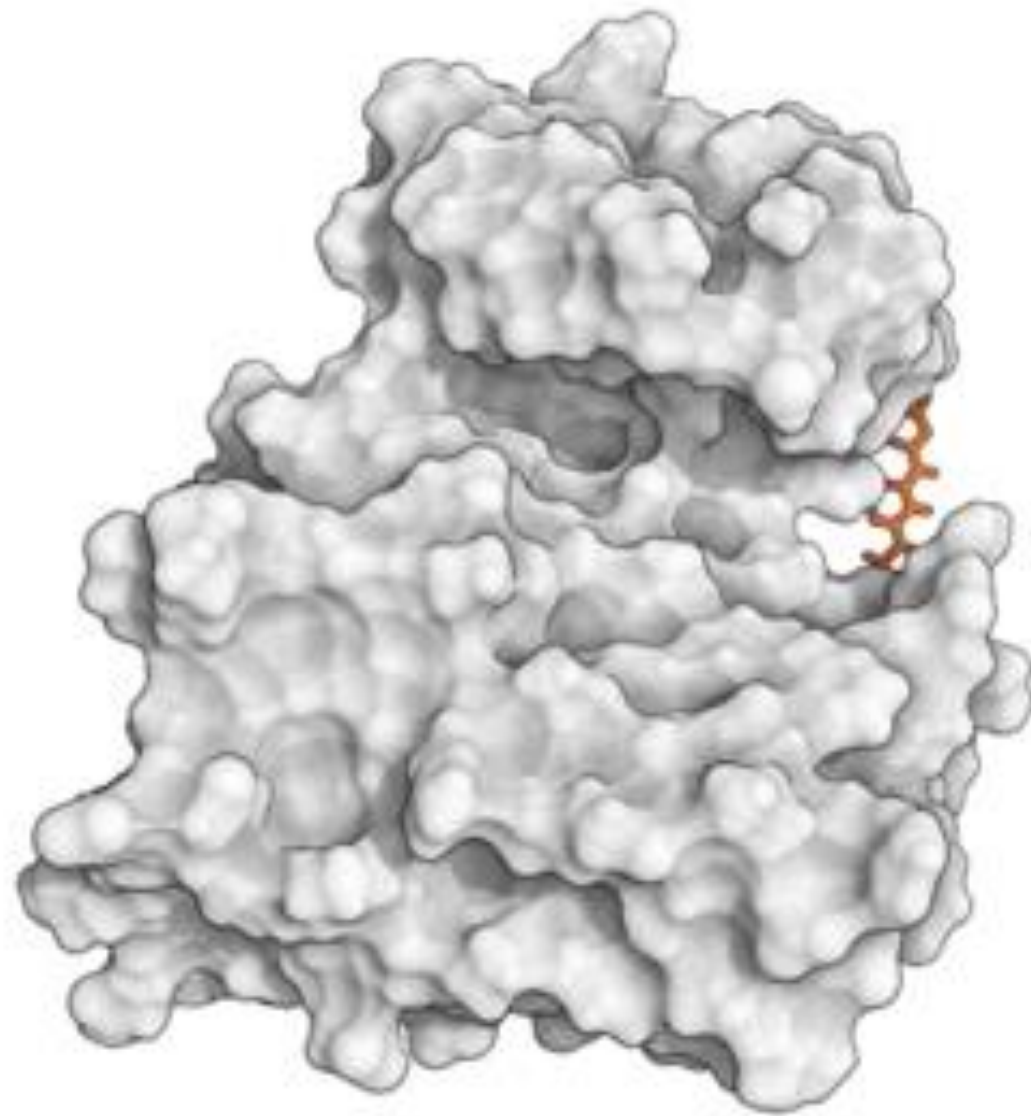


Simulation of Biomolecules

Setting up a protein simulation

2024 CCP5 Summer School



Dr Matteo Degiacomi
Durham University

matteo.t.degiacomini@durham.ac.uk

Dr Antonia Mey
University of Edinburgh

antoniamey@ed.ac.uk

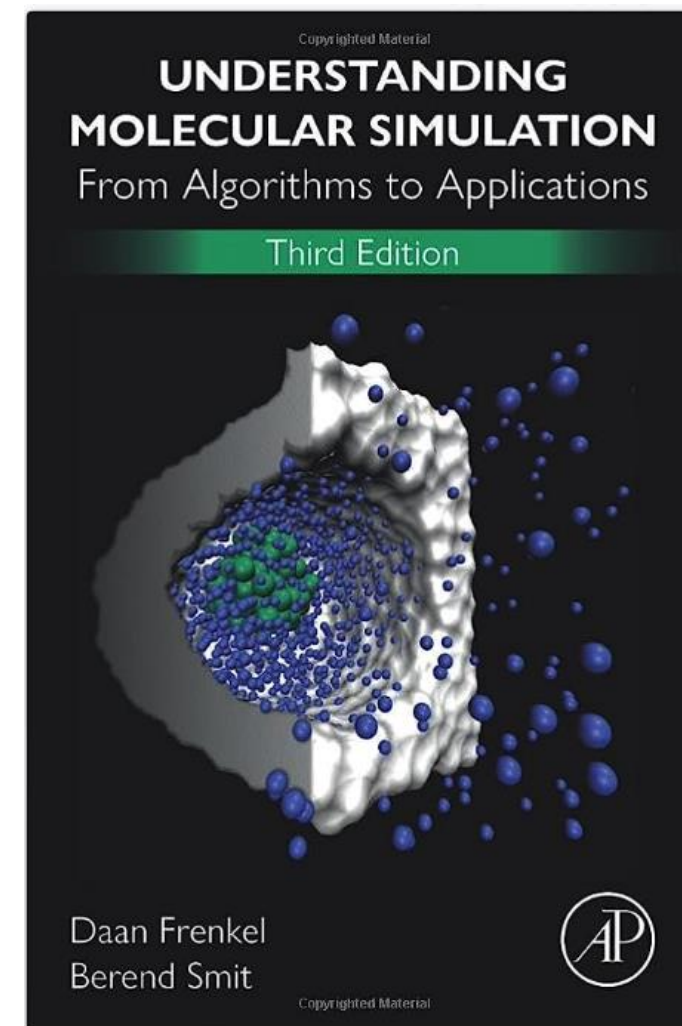
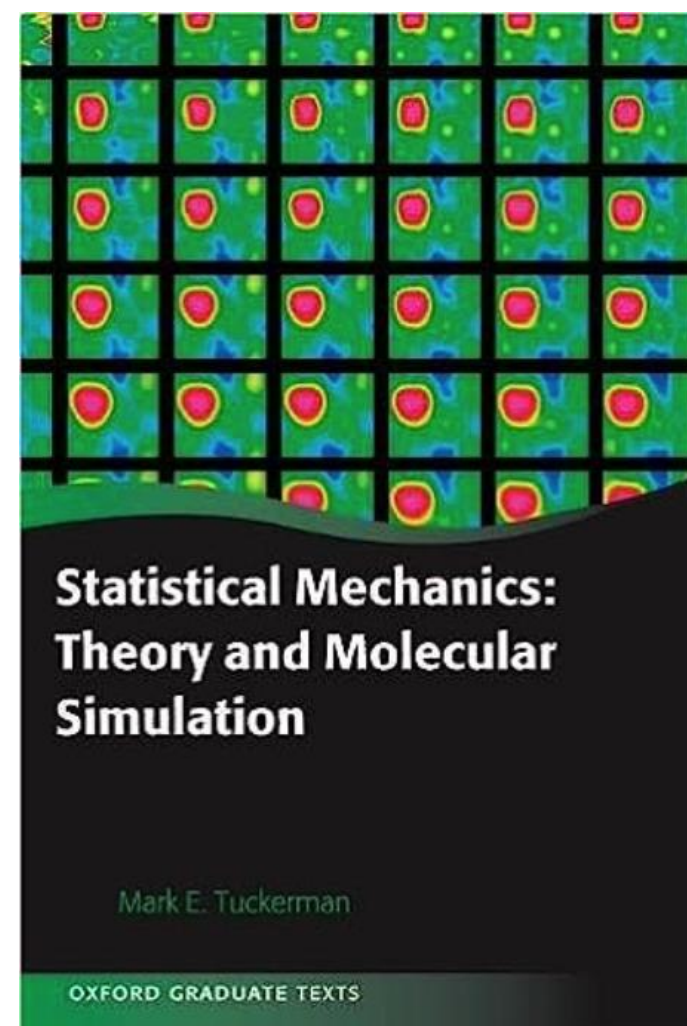
Useful resources to learn running simulations

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

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

Lester O. Hedges^{1,2*}, Sofia Bariami^{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]

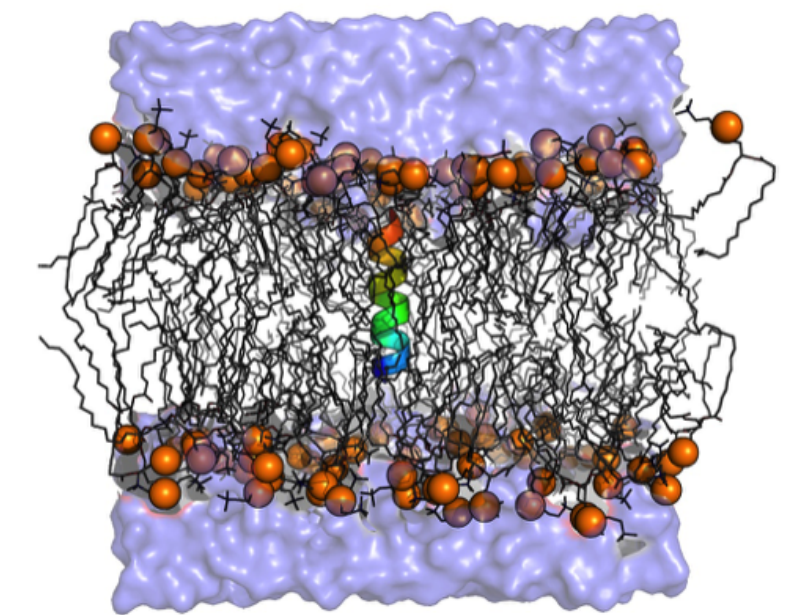
Justin A. Lemkul

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

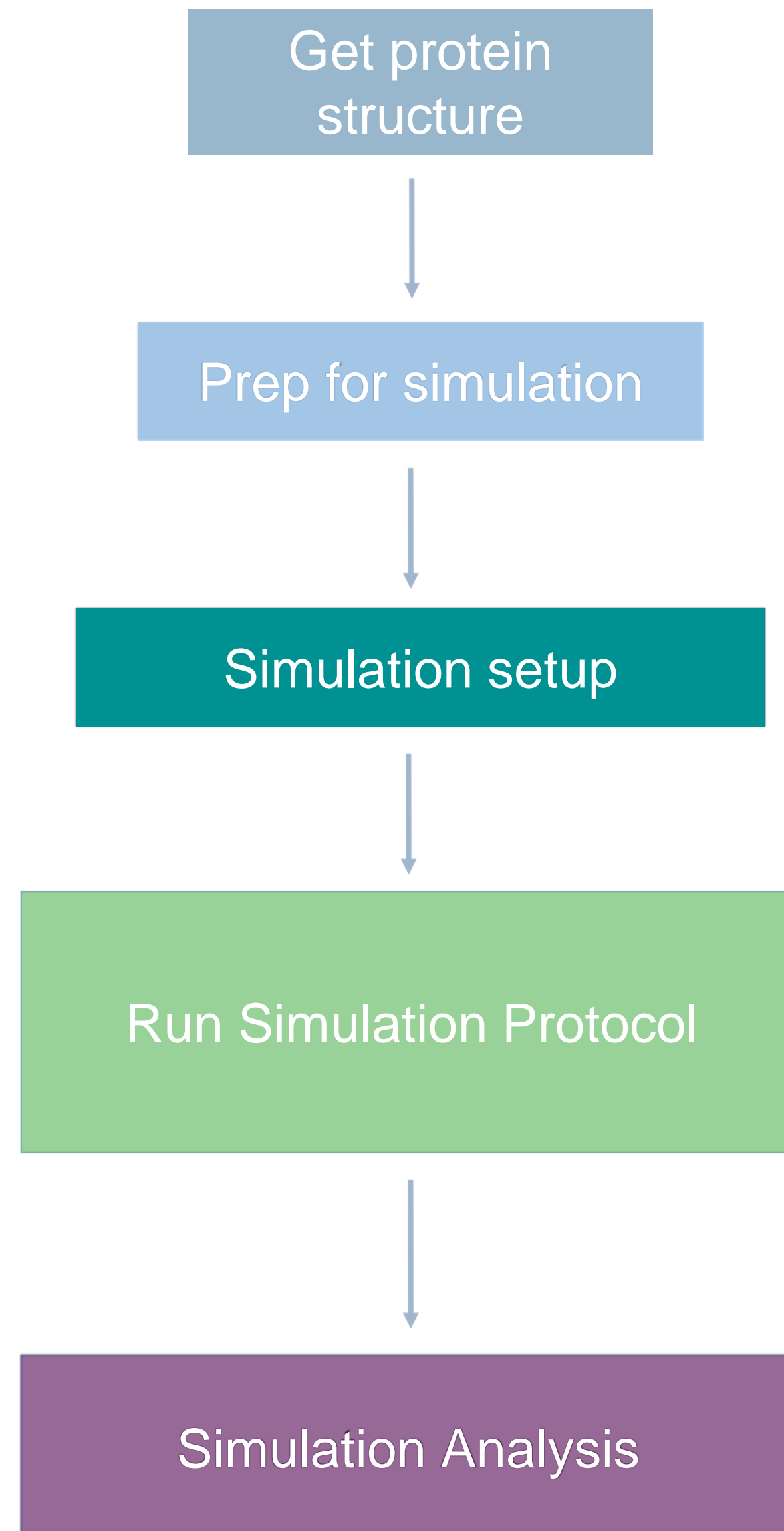
 ARTICLE CODE REPOSITORY

GROMACS: tutorials.gromacs.org

Amber: ambermd.org/tutorials

OpenMM: docs.openmm.org/latest/userguide/library/03_tutorials.html

A typical workflow for molecular dynamics



Getting your protein structure

AlphaFold Protein Structure Database

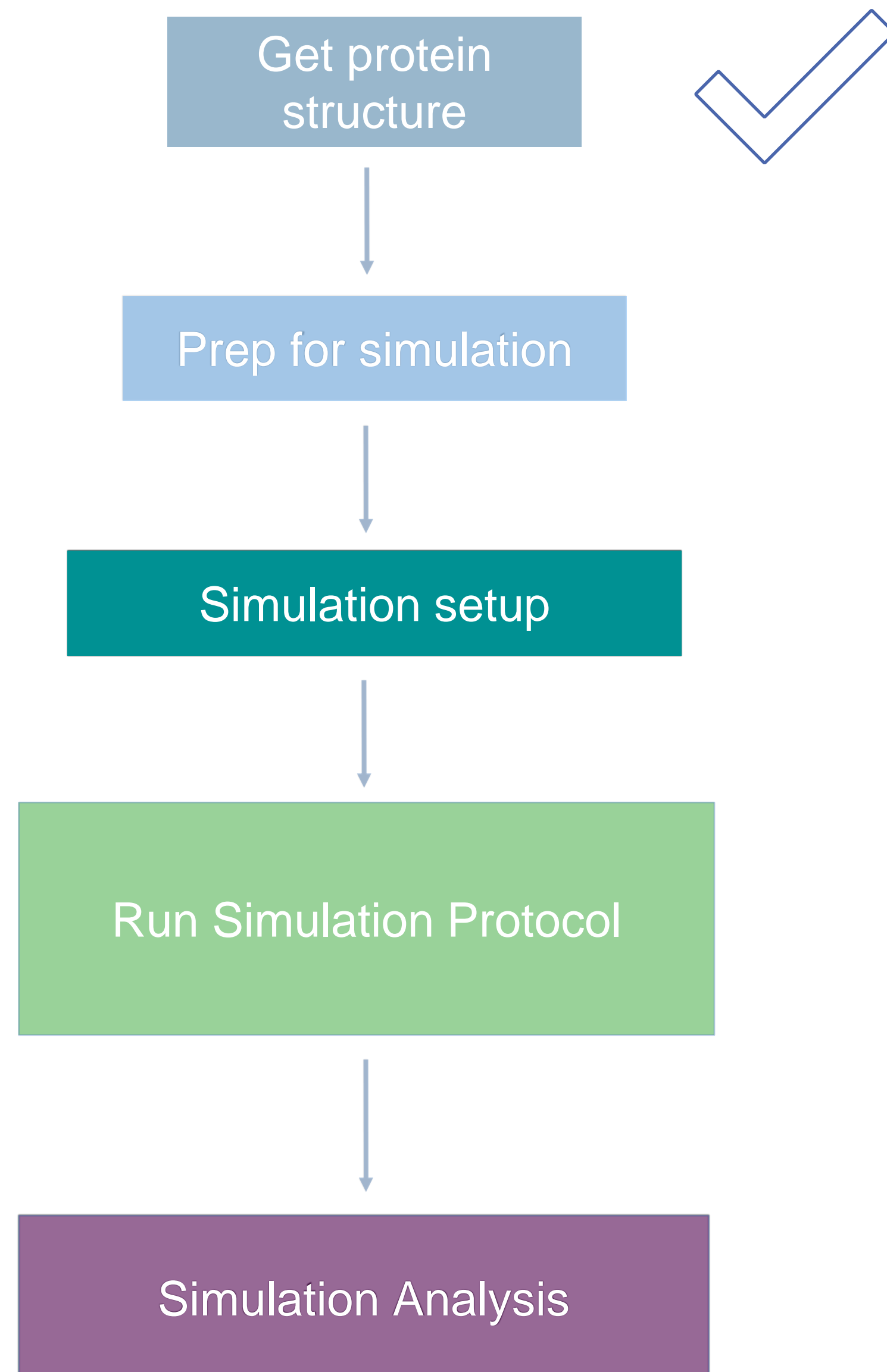


Getting ligands/co-factors

ZINC20



A typical workflow for molecular dynamics



**Prepping your protein
and/or ligand(s)**

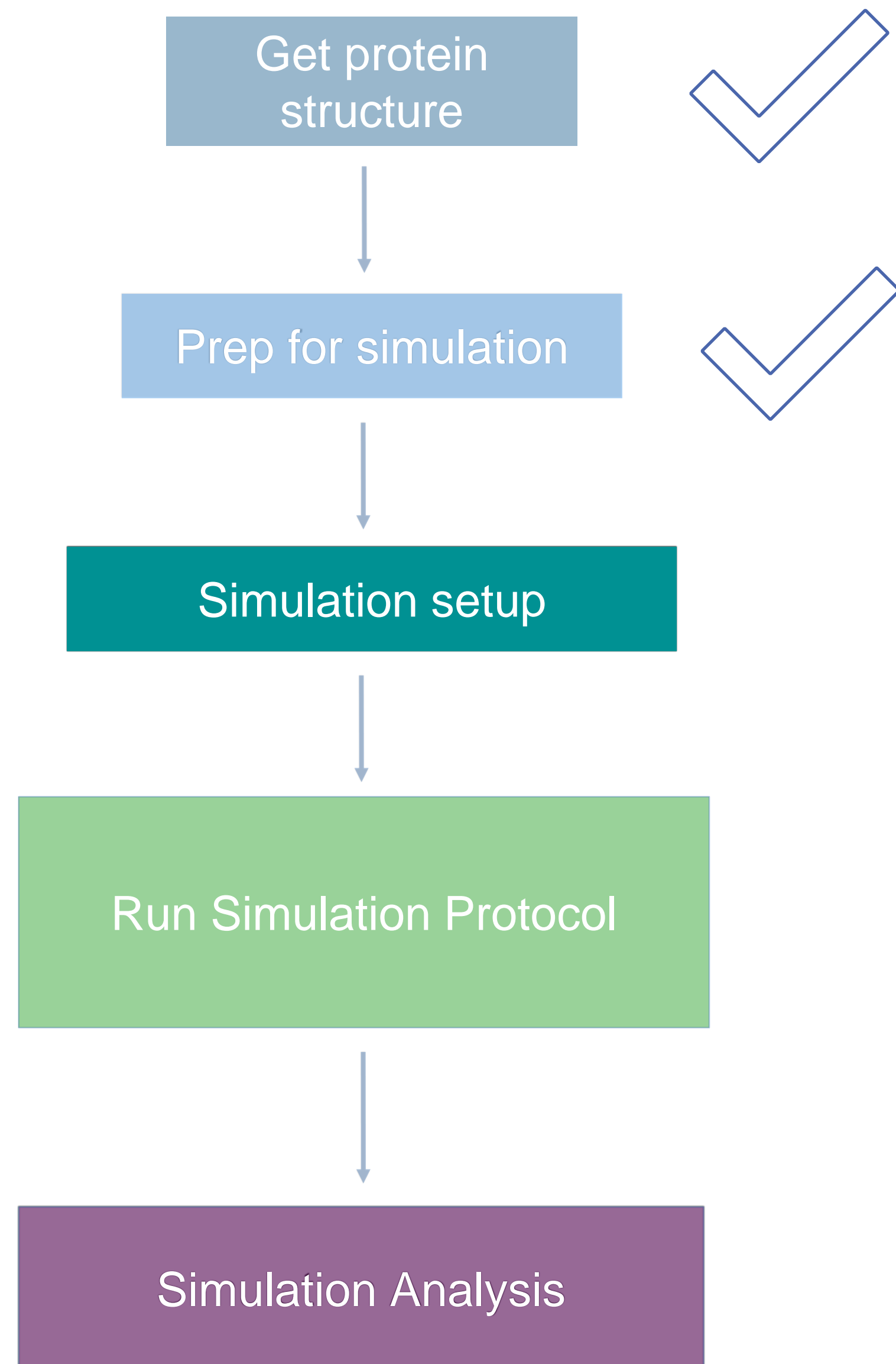
SCHRODINGER
Maestro



SeeSAR
The drug design dashboard

[...]

A typical workflow for molecular dynamics



Setup and run on your favourite MD engine

Gromacs

OpenMM

Amber

NAMD

BioSimSpace

SOMD

Charmm

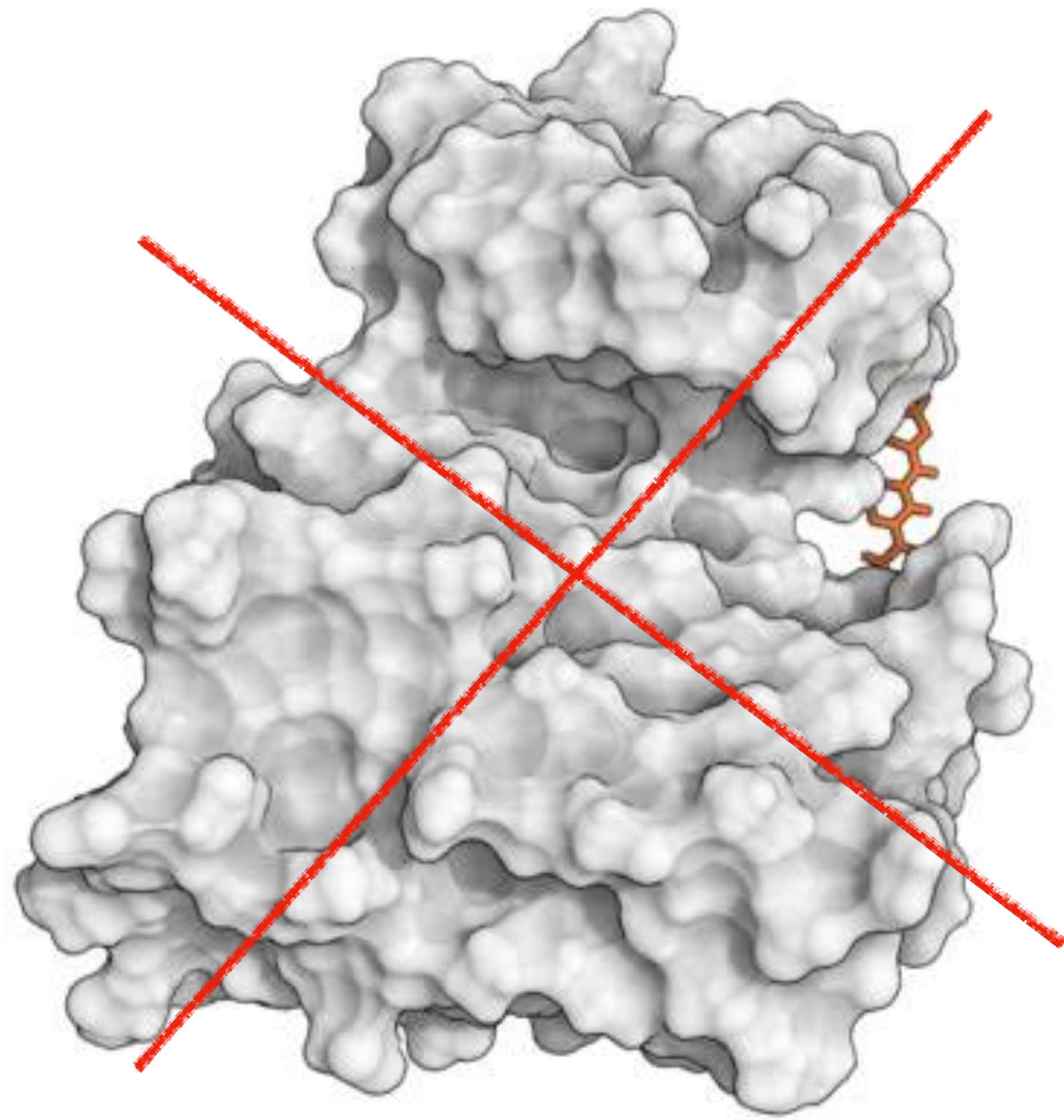
DLPoly

LAMMPS

*mostly C++
command line*

Python API

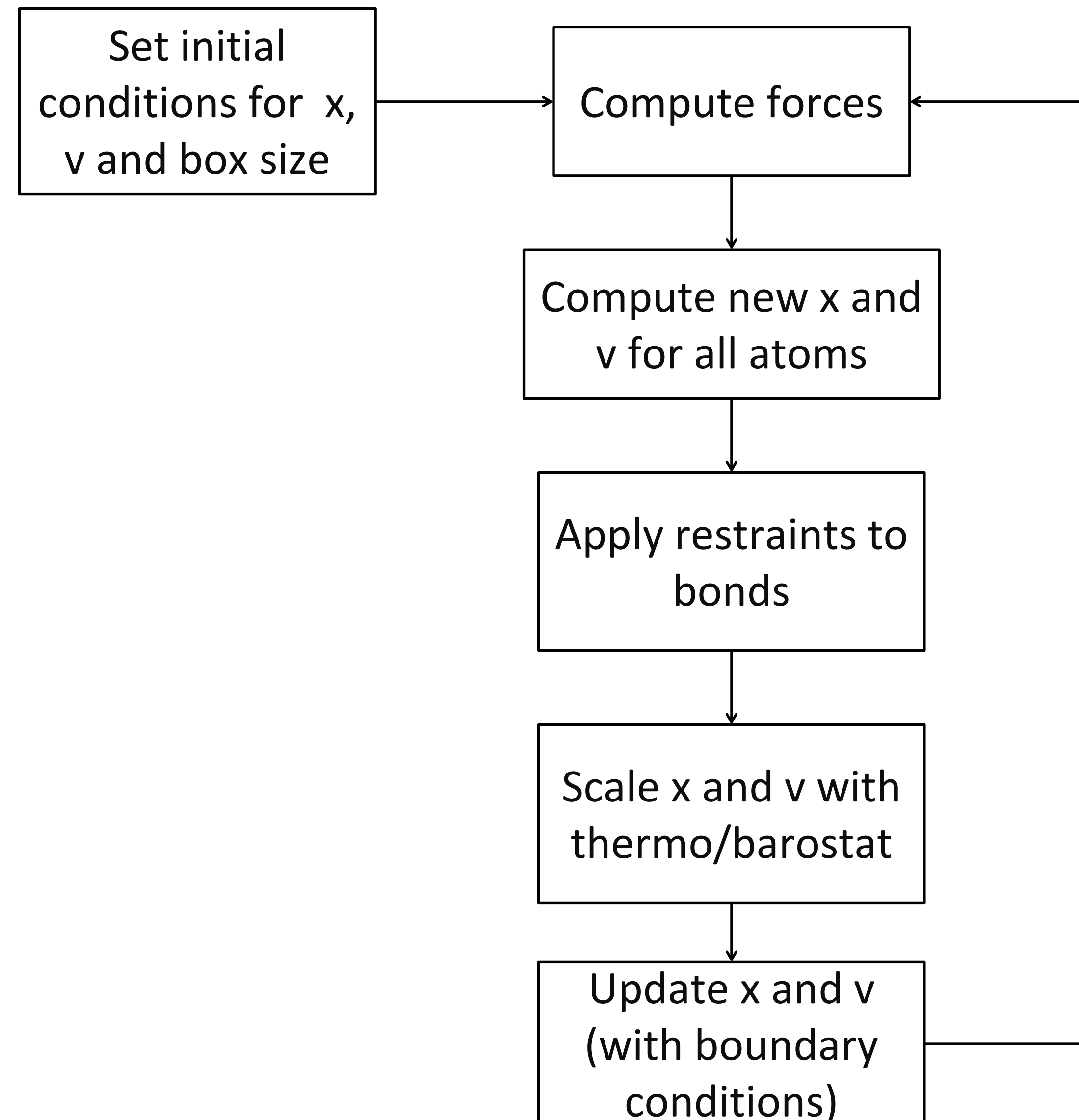
Disclaimer!



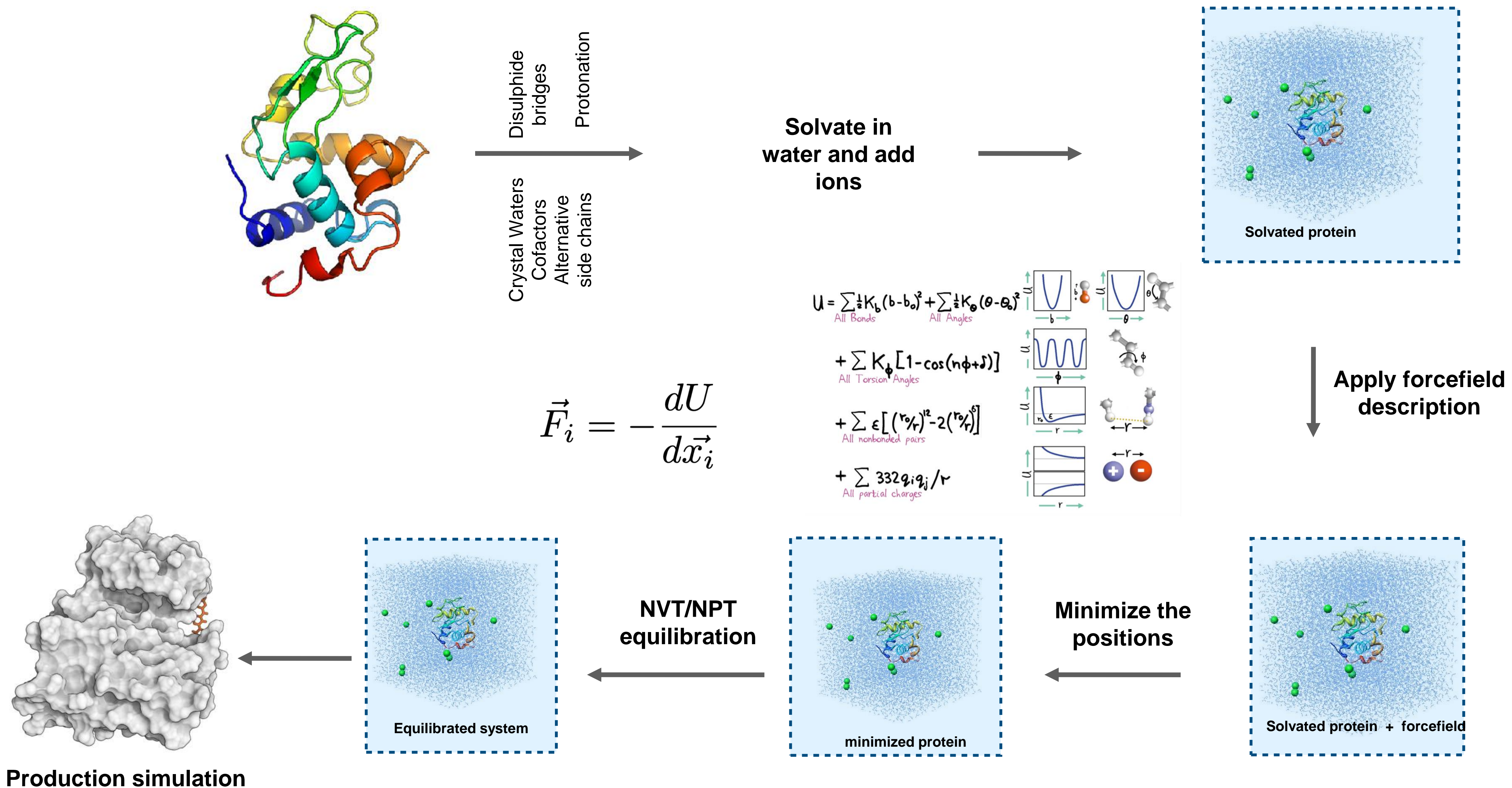
Running biomolecular MD can take days on specialised hardware.

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

A Molecular Dynamics timestep



Molecular dynamics require multiple steps for the setup of simulations



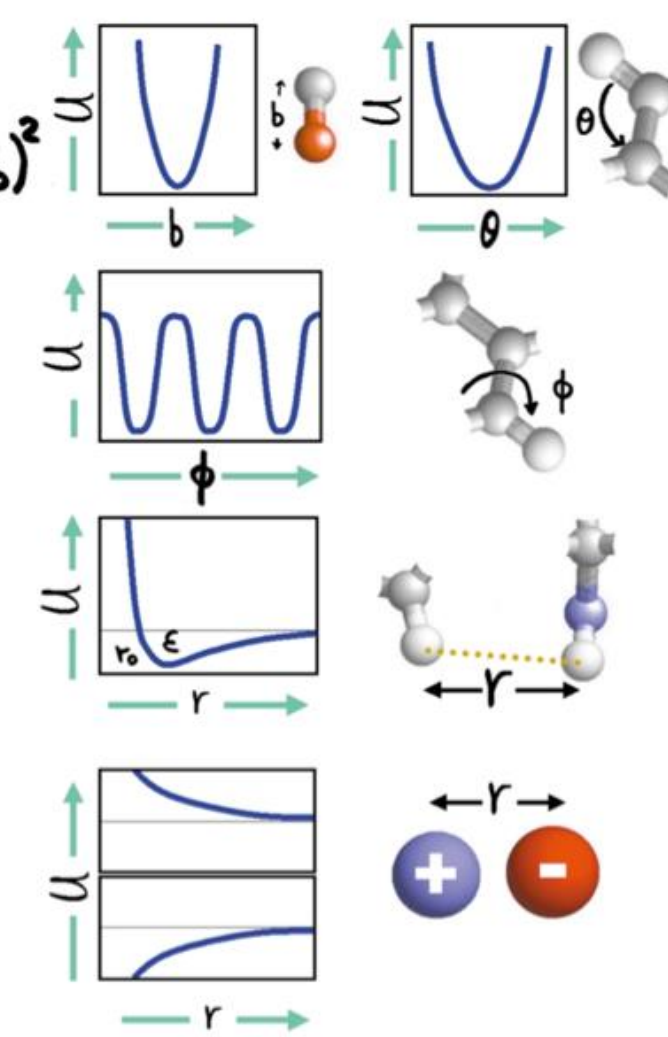
There are many different choices for force fields to be made

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

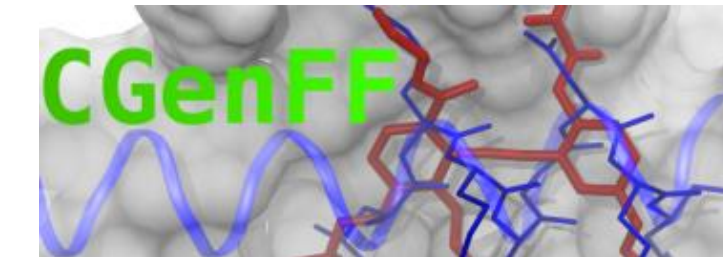
$$U = \sum_{\text{All Bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{All Angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

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

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

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


Small molecule force fields



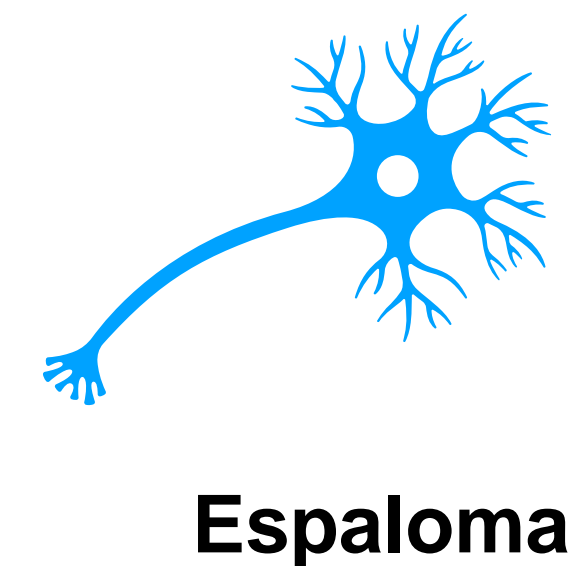
$$E_{\text{pair}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 +$$

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

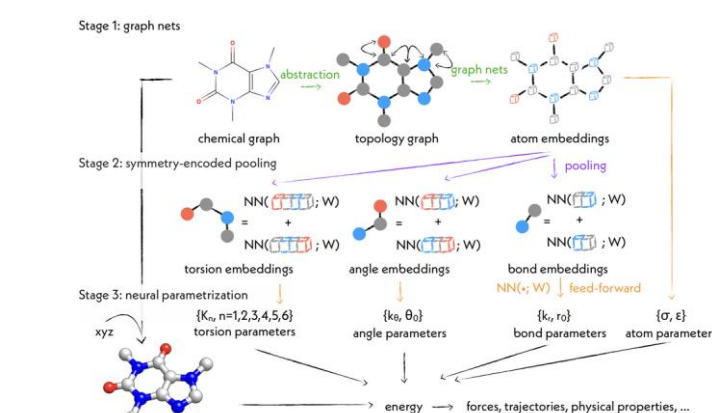
GAFF

Machine learned force fields

- **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)



Ani-2x



SchNet

PhysNet

BandNN

There is no “best force field”!

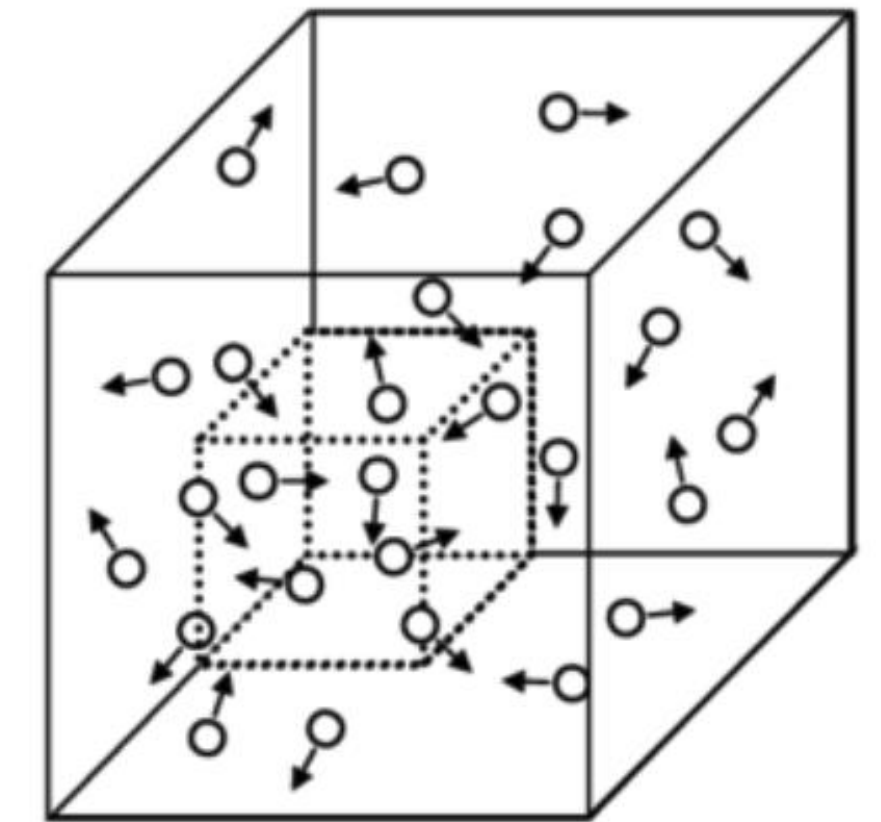
Coarse-grained force fields...

Choosing your thermodynamic ensemble

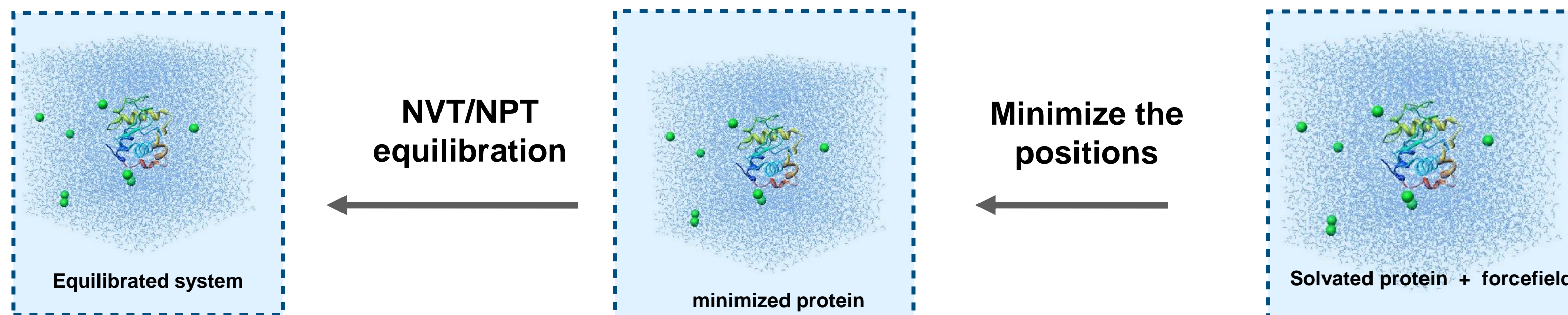
Simulations replicate a specific *thermodynamic ensemble* (typically NVT or NPT), or even grand canonical (μ VT)

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

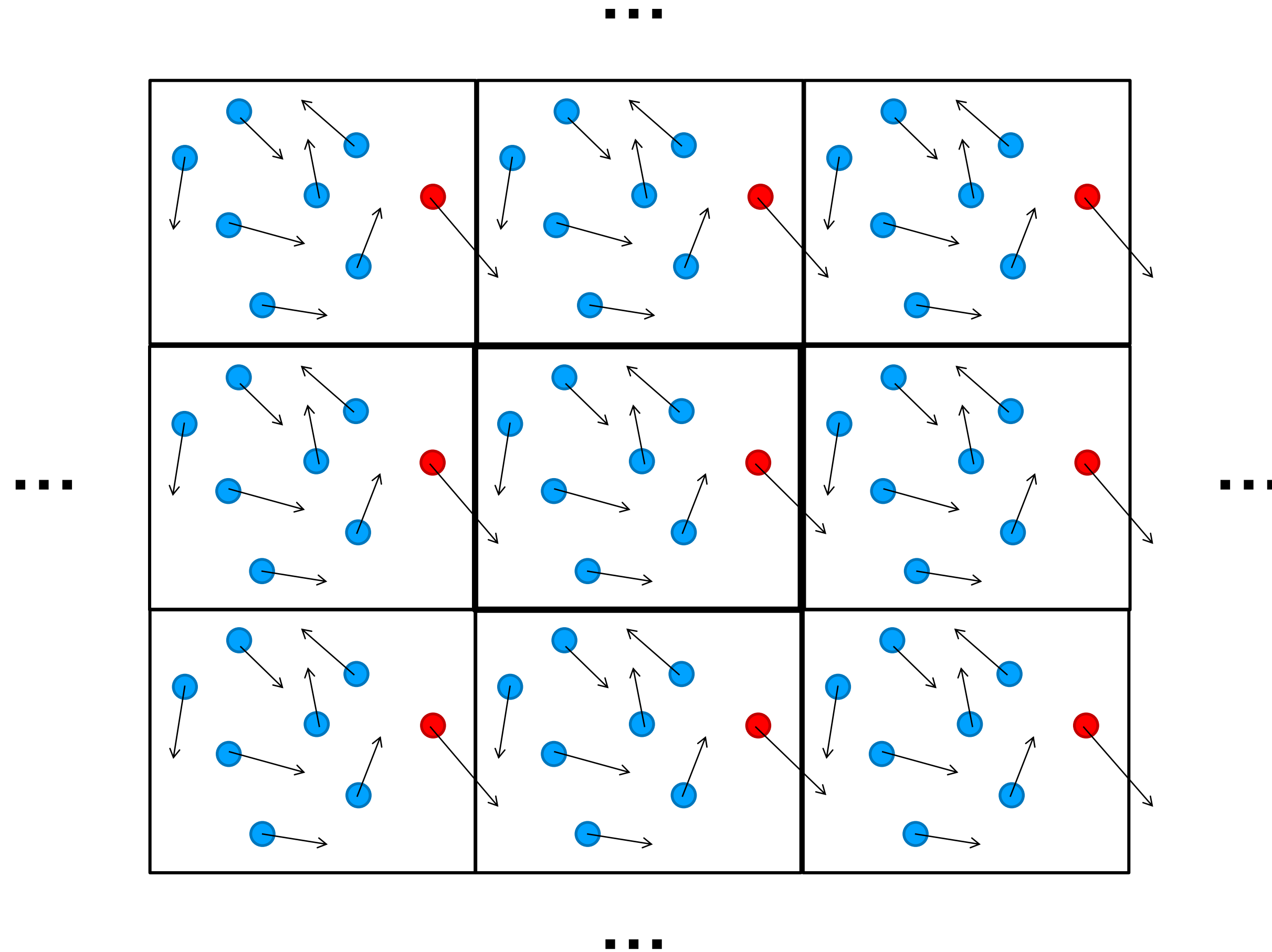
- Nose-Hoover
- Berendsen
- Parrinello-Rahman
- Langevin piston
- ...



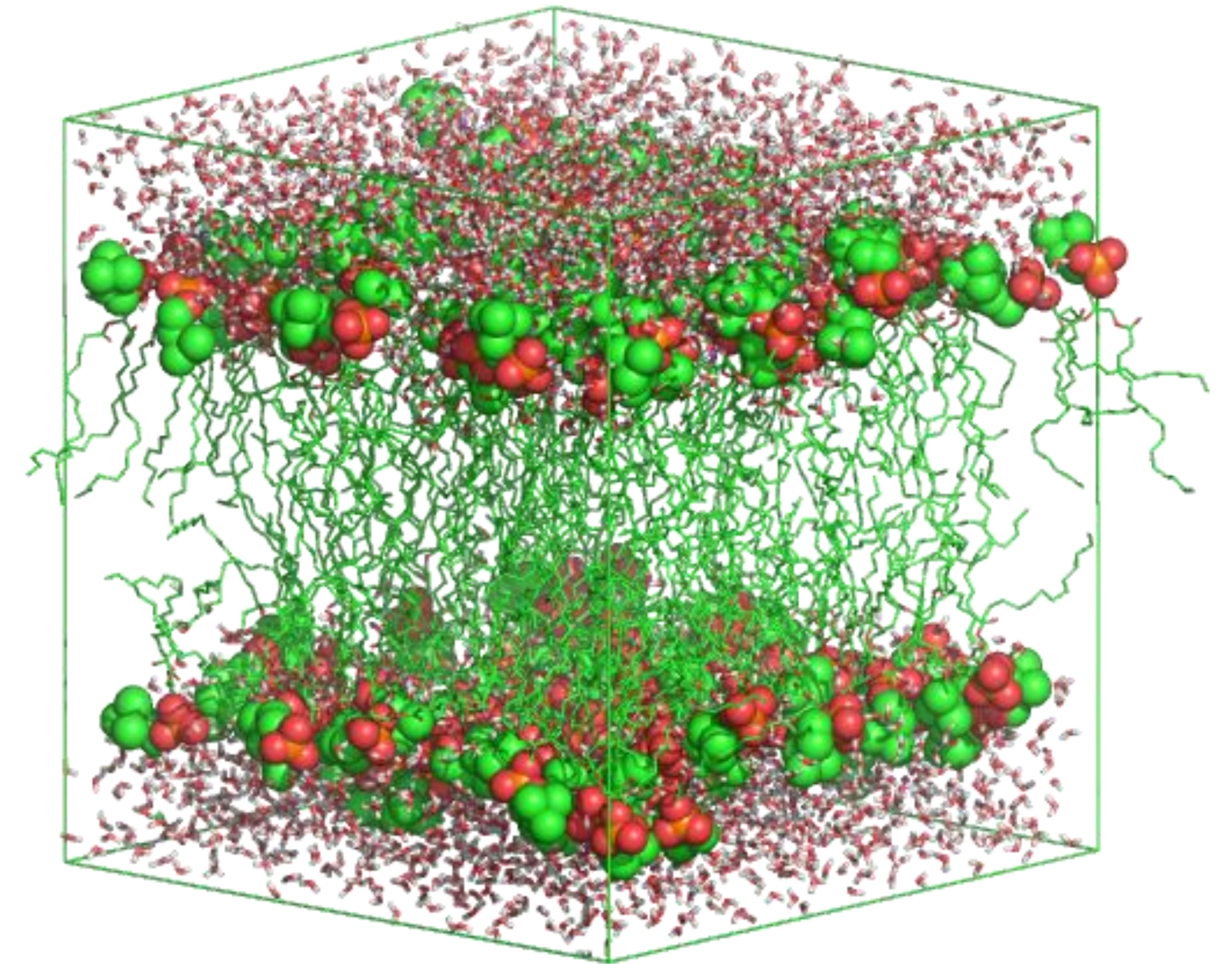
μ VT



Periodic boundary conditions (PBC) and pressure coupling



Typically, PBC in x-y-z



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

Sampling timescales for protein systems

The steepest gradient determines the smallest timestep:

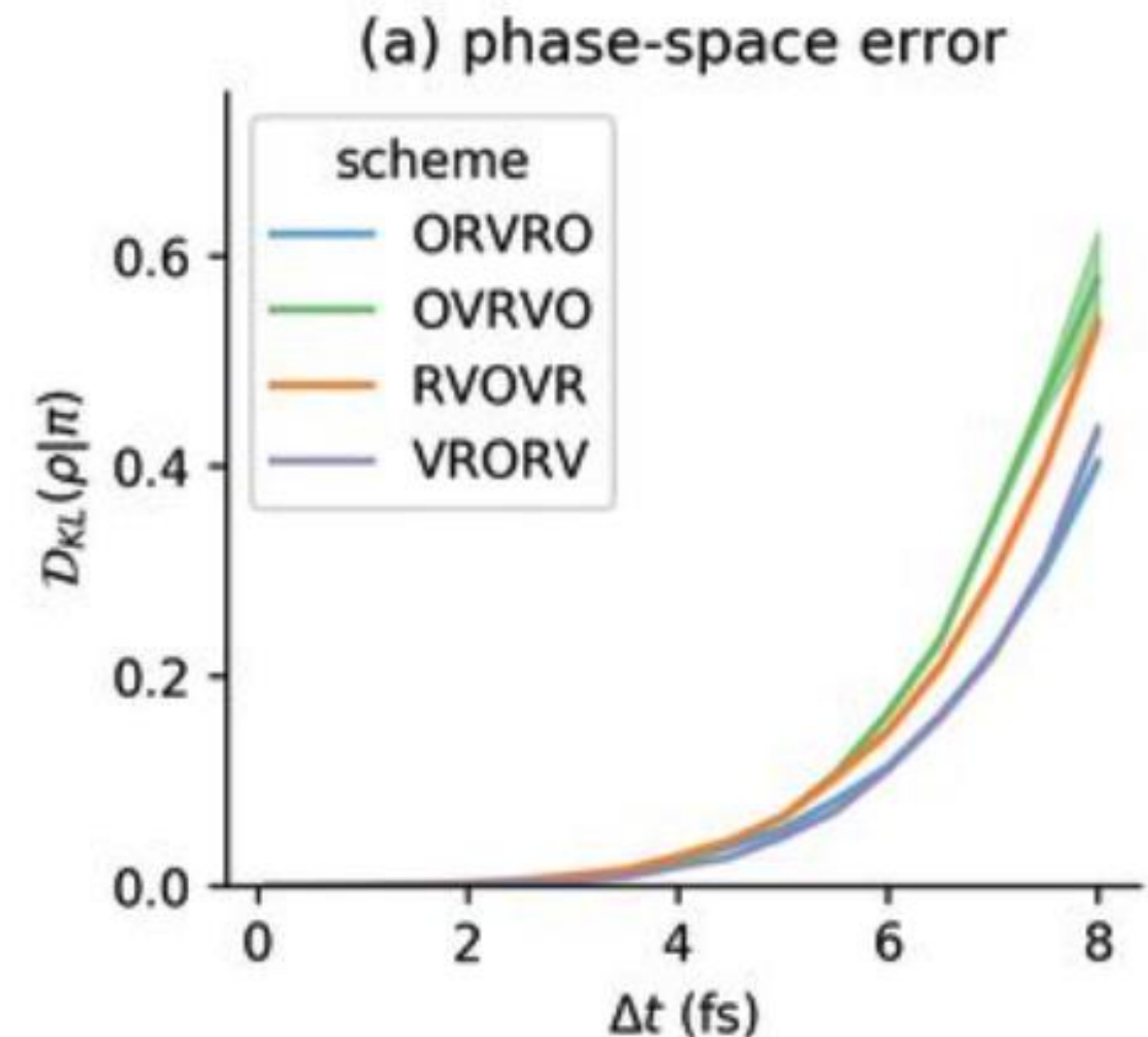
Timestep size is imposed by the fastest phenomenon we want to observe :

- 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?

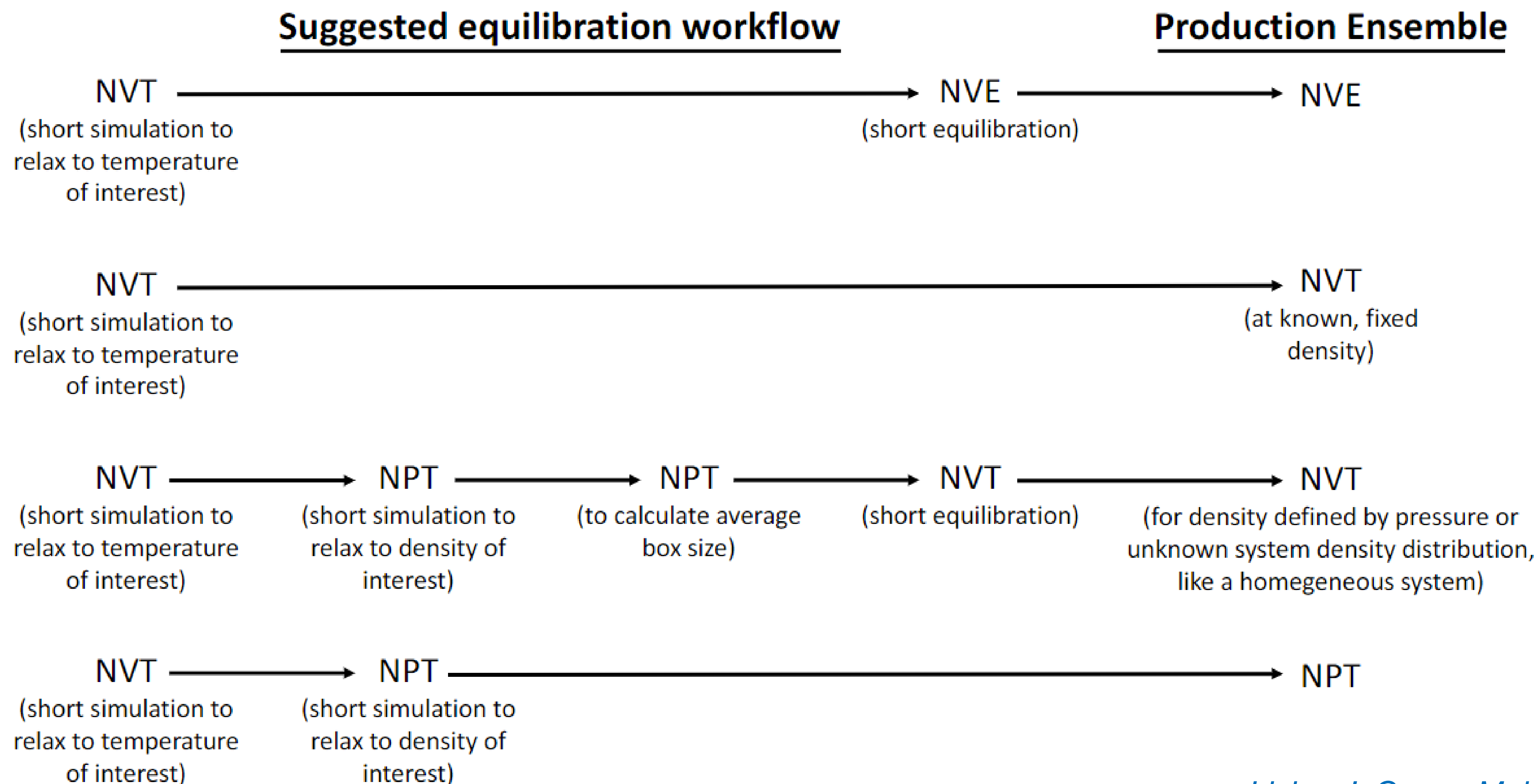
Production phase: do
we have good sampling
and convergence?



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. α -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 μ 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:

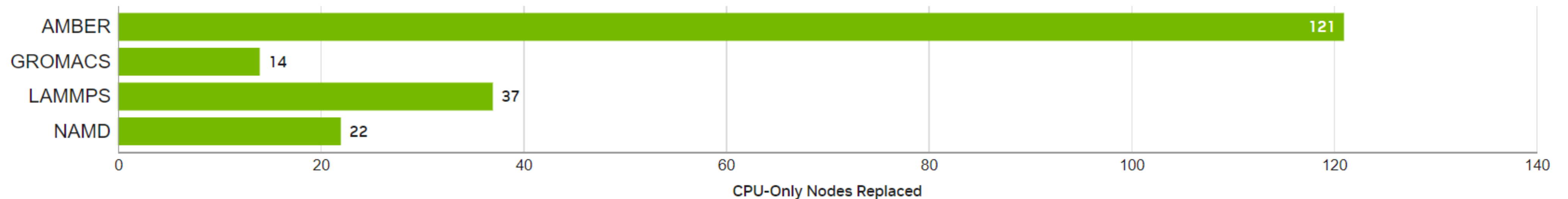
- Support for force field of choice
- Enables running desired simulation protocol
- Performance for available hardware
- Ease of use

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

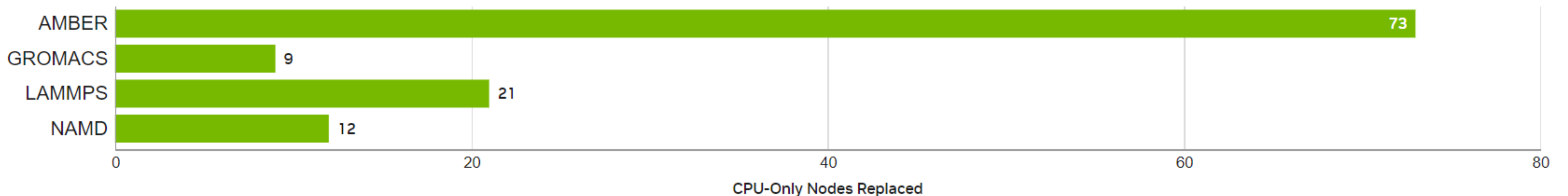
Graphical Processing Units (GPUs) are especially effective for MD

From: developer.nvidia.com/hpc-application-performance

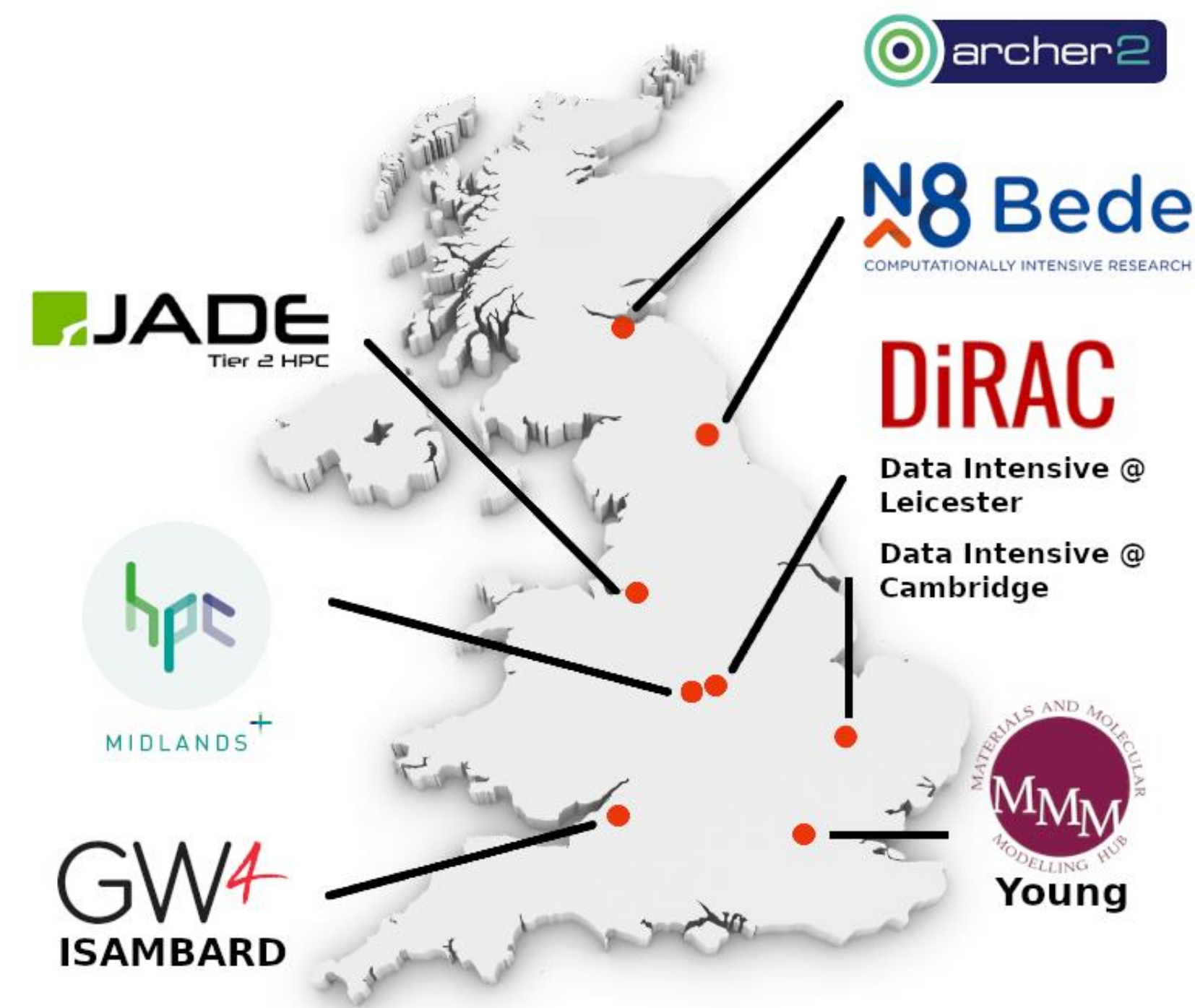
NVIDIA H200



NVIDIA A100



Calculating runtimes: example on UK Tier 2 systems



← ↻ 🔒 <https://www.hecbiosim.ac.uk/access-hpc/hpc-calculator> 🔊 ☆ 📄 ⌵ 🏠 🌱

HECBioSim 🔍 Search ... 🔑 Log in

Home Contact Us Covid-19 Events ▾ Access HPC ▾ Highlights Publications Downloads About Us ▾

The HECtime Calculator

Enter the following information about your simulation:

JADE2 ▾

1000

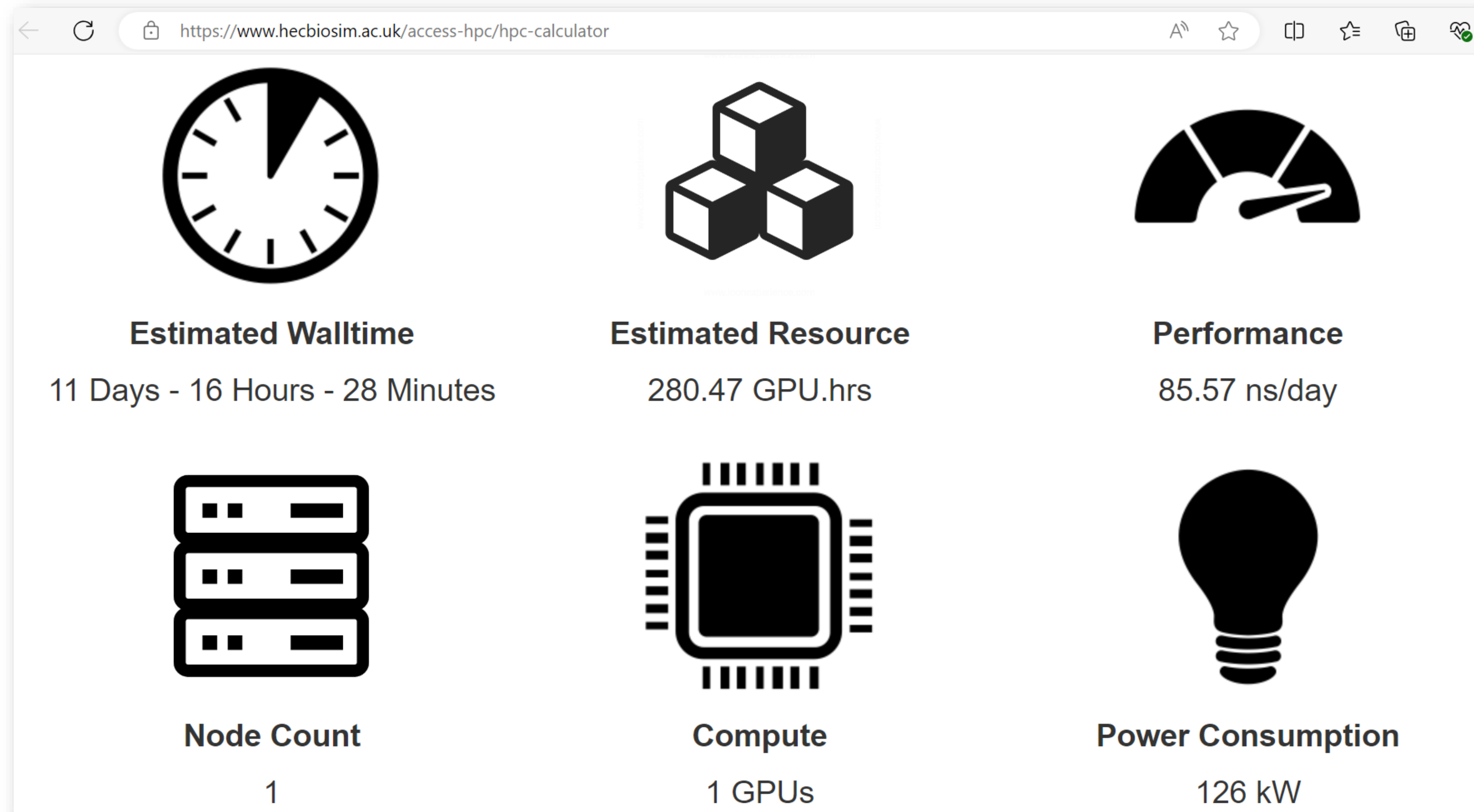
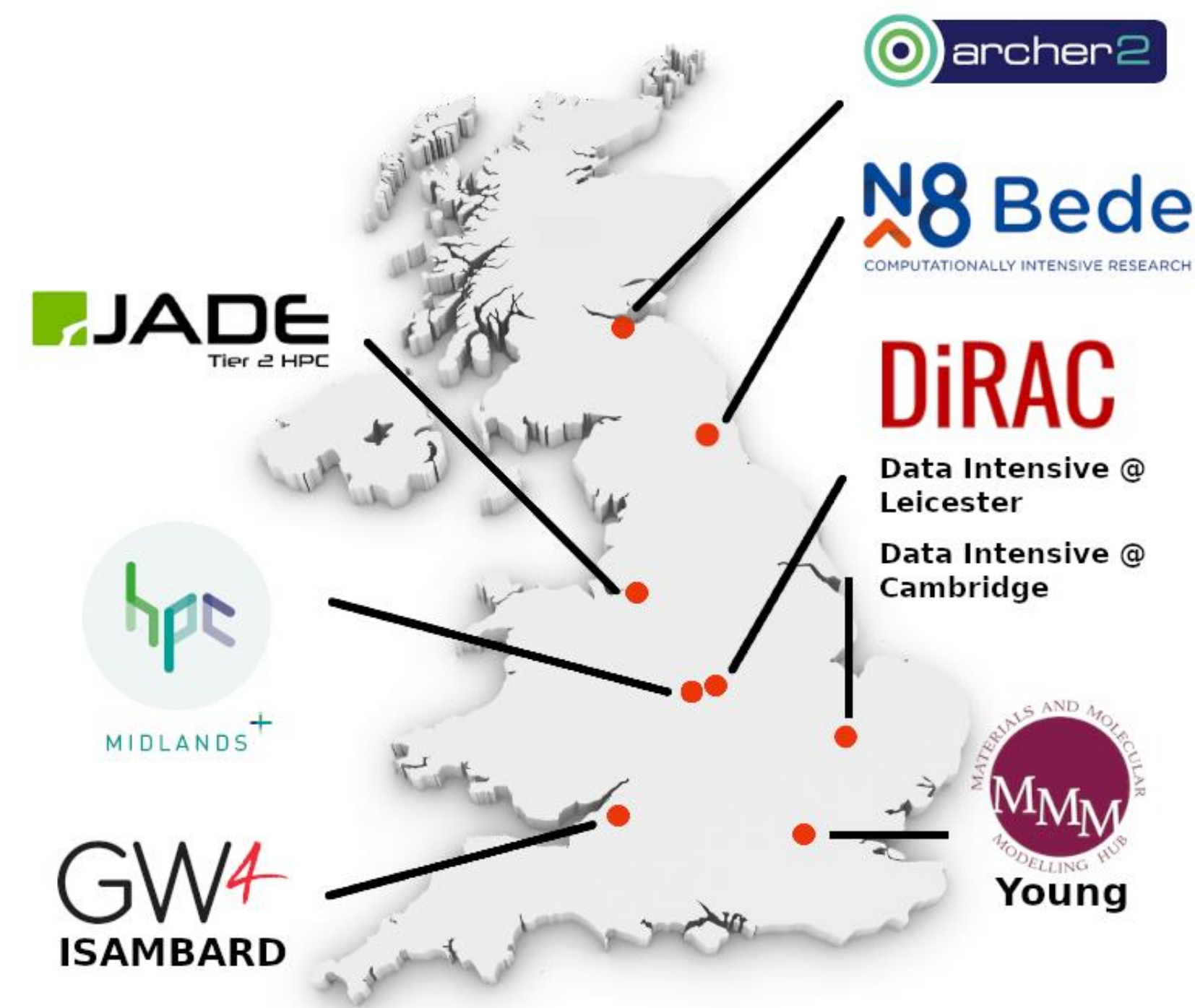
100000

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>

← ↻ 🔒 <https://docs.openforcefield.org/en/latest/examples.html> 🔊 ☆ 📄 ☆ 📌 📌 📌 ...

 **open**
forcefield OpenFF Documentation documentation / Examples 🔍 Search v: latest

Overview

OpenFF Standards

GETTING STARTED

Installation

Modelling with OpenFF

Examples

PROJECTS

OpenFF Toolkit

Interchange

Units


BespokeFit

QCSubmit

Fragmenter

Tutorials


Tutorials describing key workflows from start to finish, with detailed explanations of what's going on.



open
forcefield

Protein-ligand-water systems with Interchange


openff-interchange



open
forcefield

Solvate and equilibrate a ligand in a box of water

openff-interchange




open
forcefield

Toolkit

Showcase: Prepare and run a protein-ligand simulation

openff-toolkit



open
forcefield

Visualizations in the OpenFF Toolkit

openff-toolkit

Parametrization and Evaluation

CONTENTS

Examples

Tutorials

Parametrization and Evaluation

Combining Force Fields

Force Field Development

Uncategorized

Running examples locally

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

