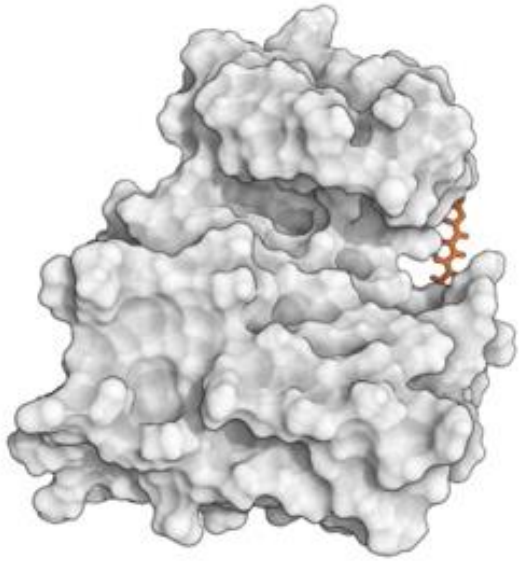
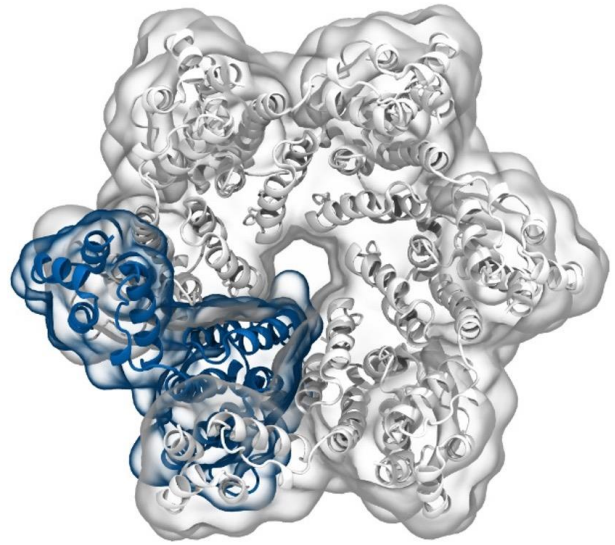
Simulation of Biomolecules



Setting up a protein simulation



Dr Matteo Degiacomi

Durham University

matteo.t.degiacomi@durham.ac.uk

Dr Antonia Mey

University of Edinburgh

antonia.mey@ed.ac.uk

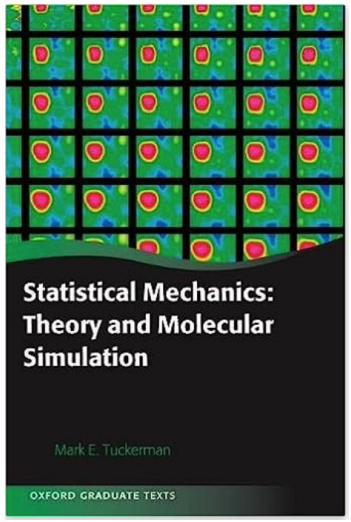
Useful resources to learn running simulations

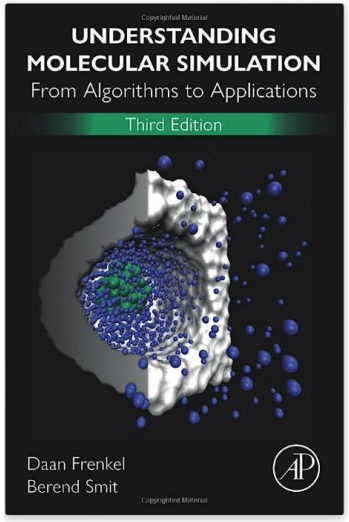
Best Practices for Foundations in Molecular Simulations [Article v1.0]

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

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

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



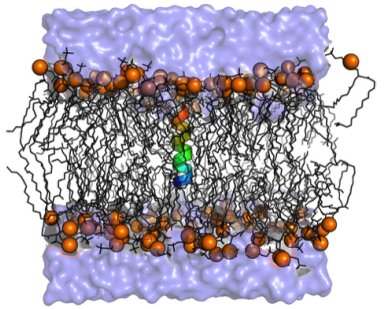


From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]

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

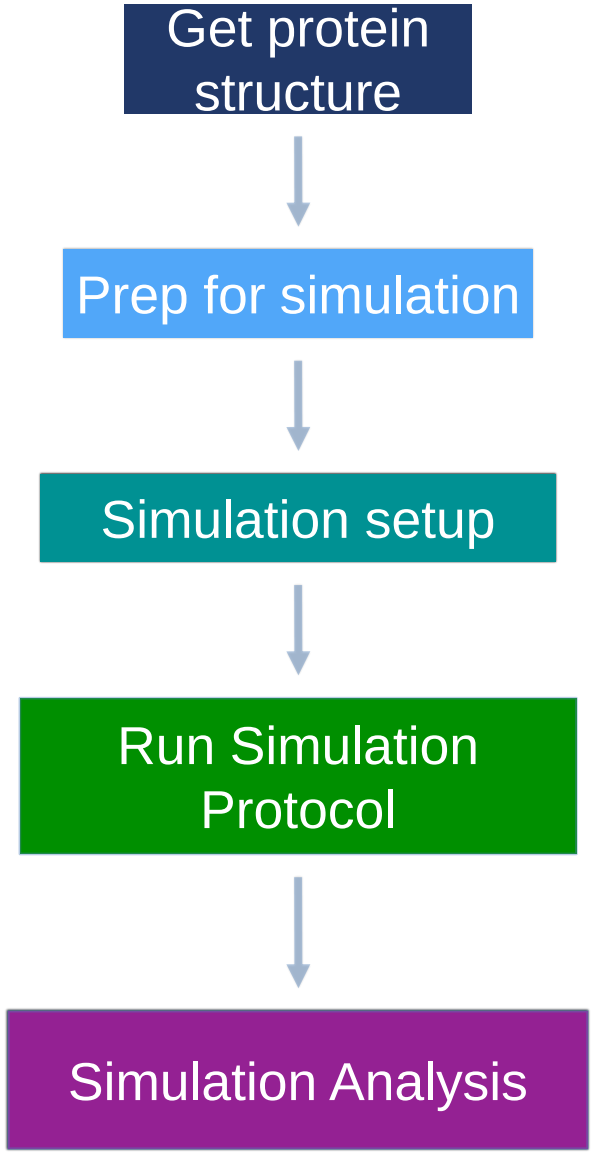
ARTICLE CODE REPOSITORY

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

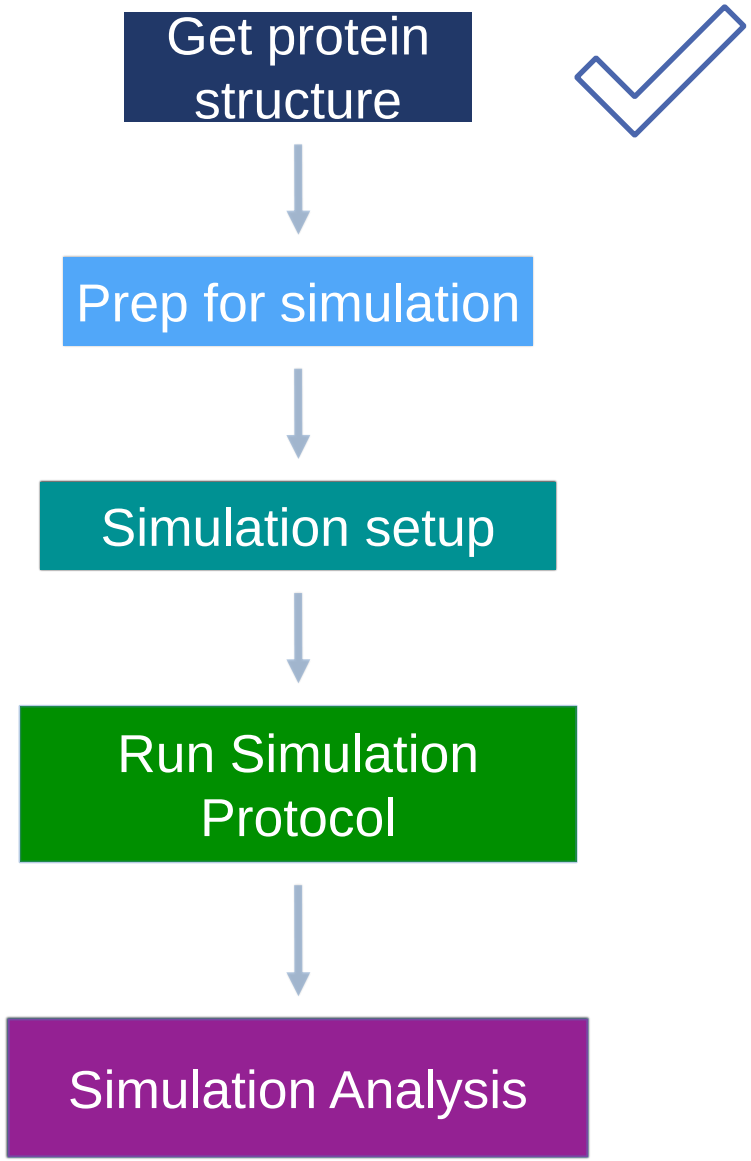
RCSB PDB PROTEIN DATA BANK

Getting ligands/co-factors

ZINC20

RCSB PDB PROTEIN DATA BANK

A typical workflow for molecular dynamics



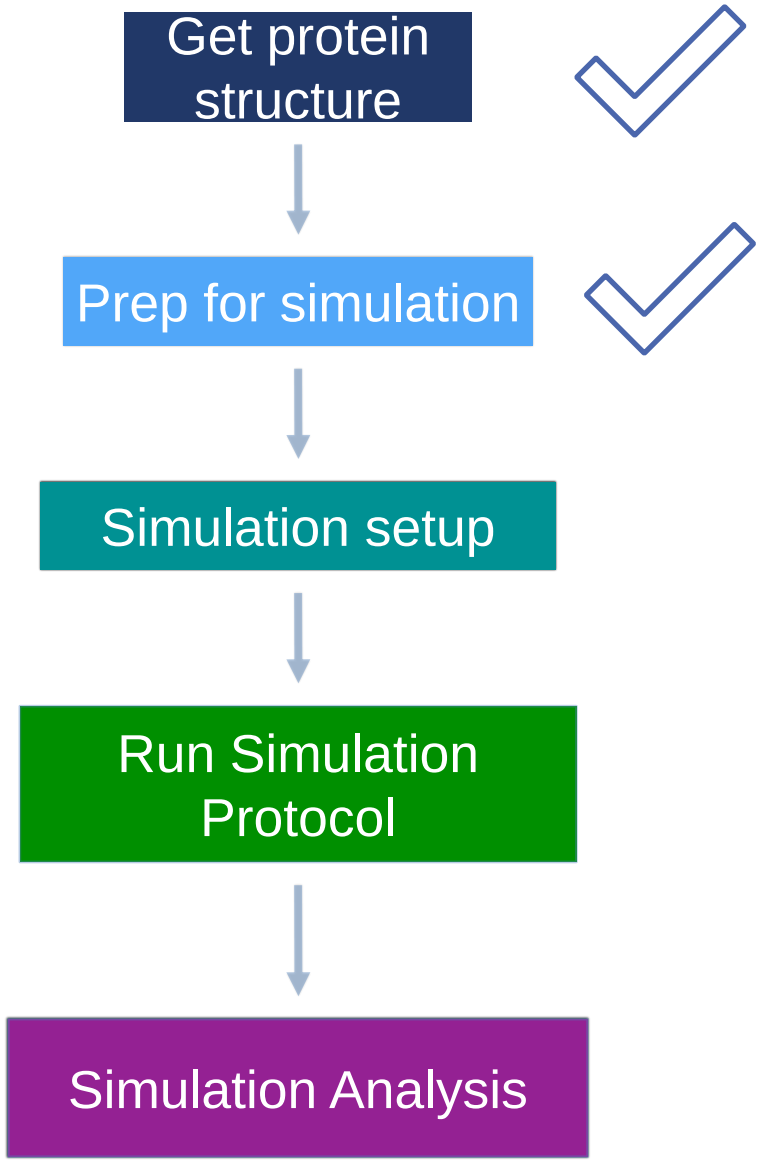
Prepping your protein and/or ligand(s)

SCHRÖDINGER
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

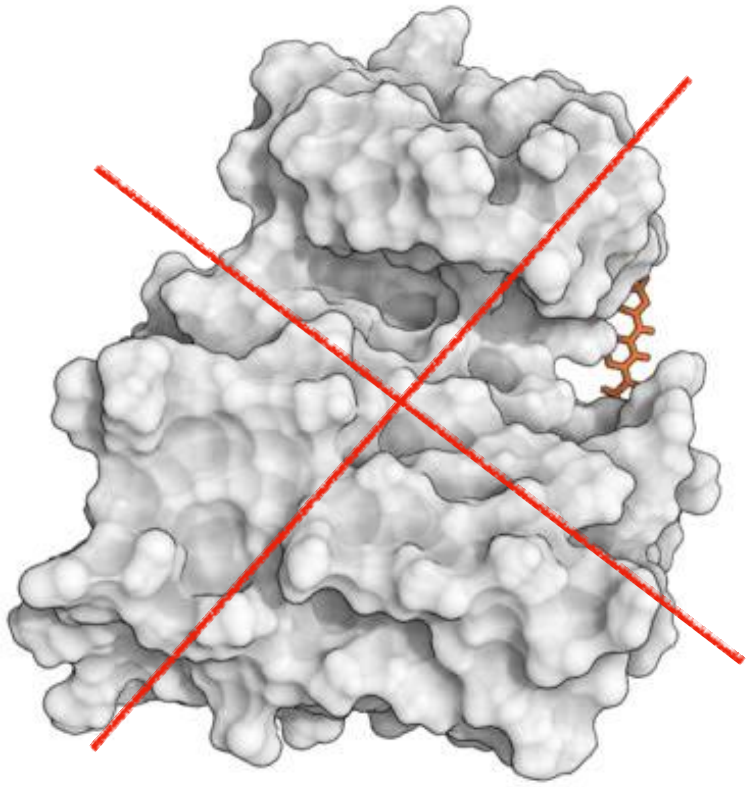
DLPoly

LAMMPS

mostly C++
command line

Python API

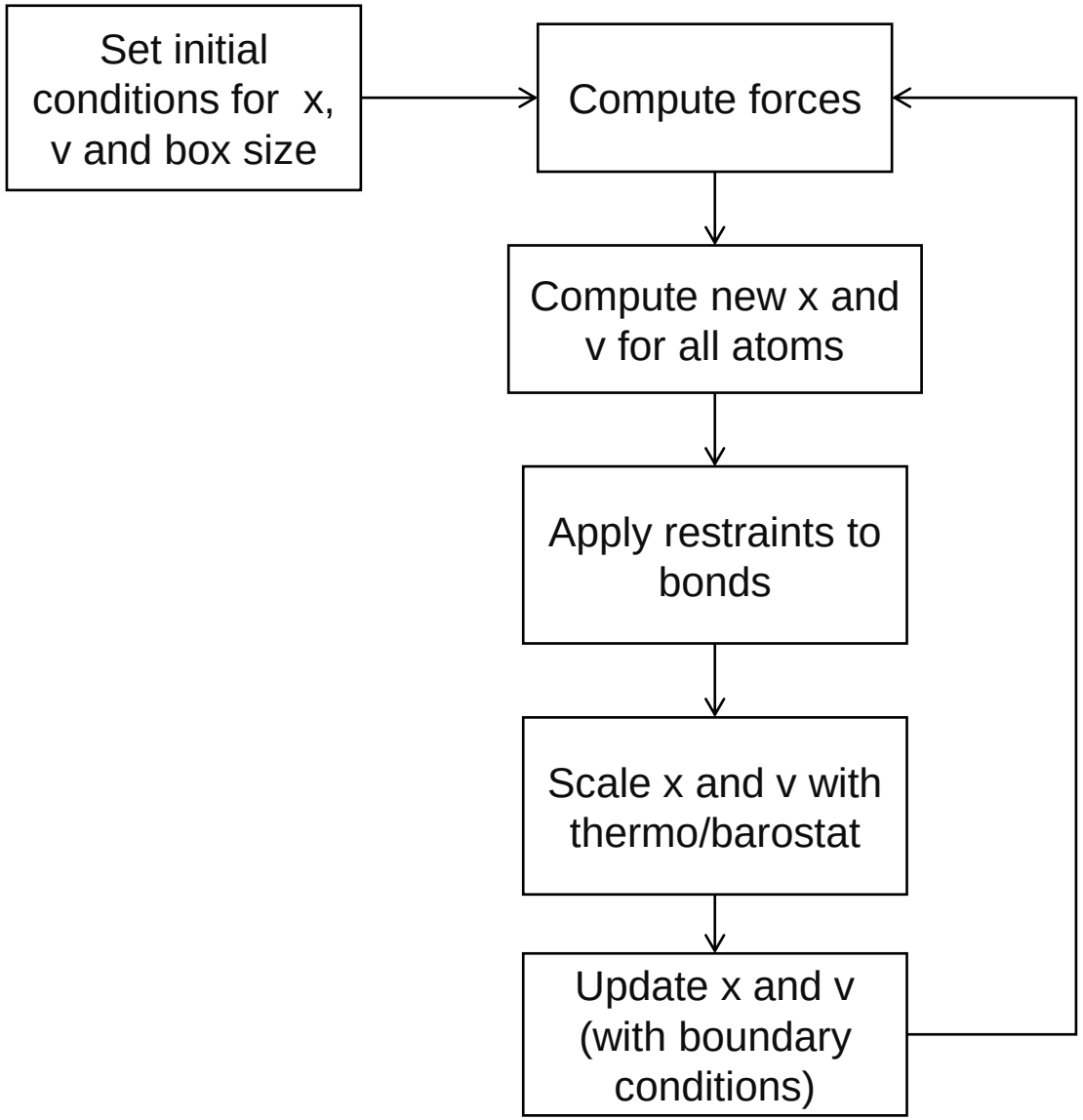
Disclaimer!



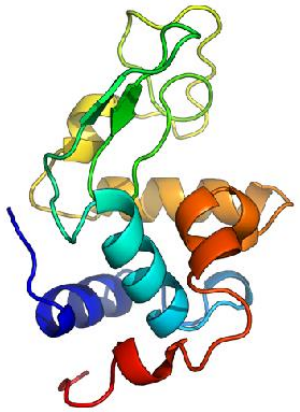
Running biomolecular MD can take days on specialised hardware.

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

A Molecular Dynamics timestep



Molecular dynamics require multiple steps for the setup of simulations



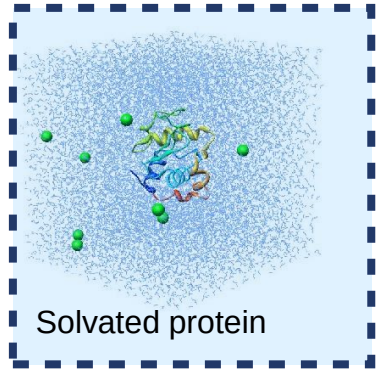
Disulfide bridges

Protonation

Crystal Waters Cofactors

Alternative side chains

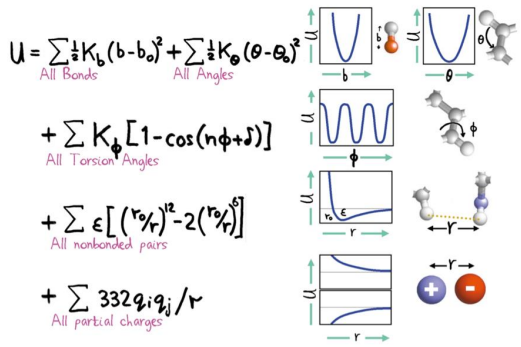
Solvate in water and add ions (*typically, neutralize the system*)

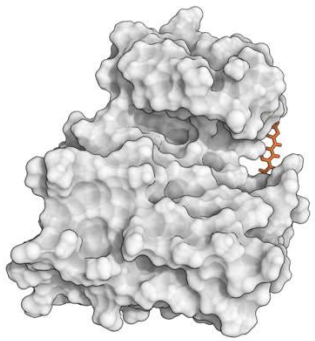


Apply forcefield description

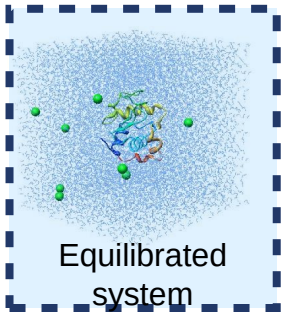
$$\vec{F}_i = -\frac{dU}{d\vec{x}_i}$$

$$U = \sum_{\text{All Bonds}} \tfrac{1}{2}K_b(b-b_0)^2 + \sum_{\text{All Angles}} \tfrac{1}{2}K_\theta(\theta-\theta_0)^2 + \sum_{\text{All Torsion Angles}} K_\phi[1-\cos(n\phi+\delta)] + \sum_{\text{All nonbonded pairs}} \epsilon\left[\left(\tfrac{r_0}{r}\right)^{12} - 2\left(\tfrac{r_0}{r}\right)^{6}\right] + \sum_{\text{All partial charges}} 332 q_i q_j / r$$

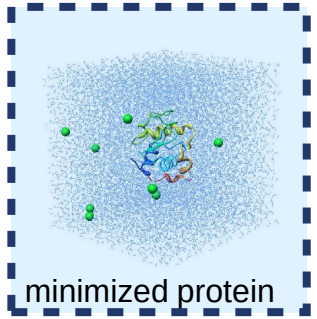




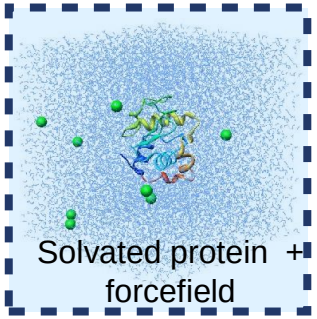
Production simulation



NVT/NPT equilibration



Minimize the positions

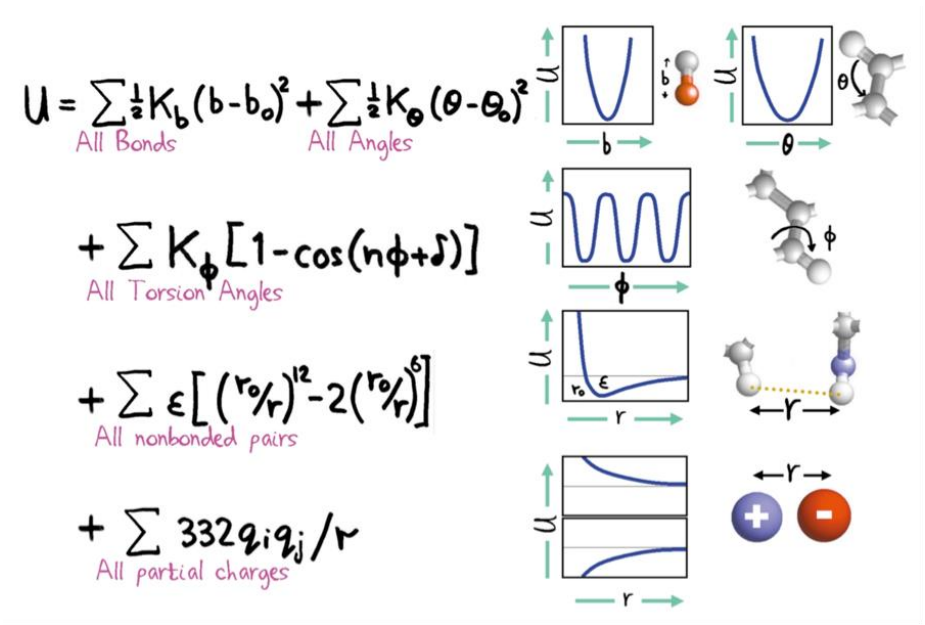


http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/lysozyme/01_pdb2gmx.html

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

There are many choices for force fields to be made

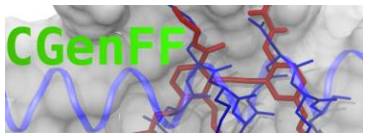
$$\vec{F}_i = -\frac{dU}{d\vec{x}_i}$$



- **Amber** (glycam params cover most sugars)
- **CHARMM** (incl. POPC, POPE, DPPC lipids)
- **OPLS**
- **GROMOS**
- **…**

There is no "best force field"!

Small molecule force fields

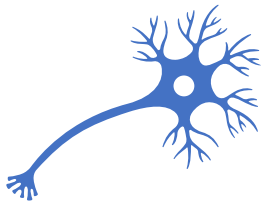


open forcefield

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

GAFF

Machine learned force fields

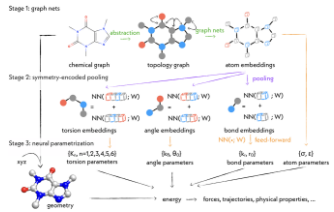


Ani-2x

SchNet

PhysNet

Espaloma



BandNN

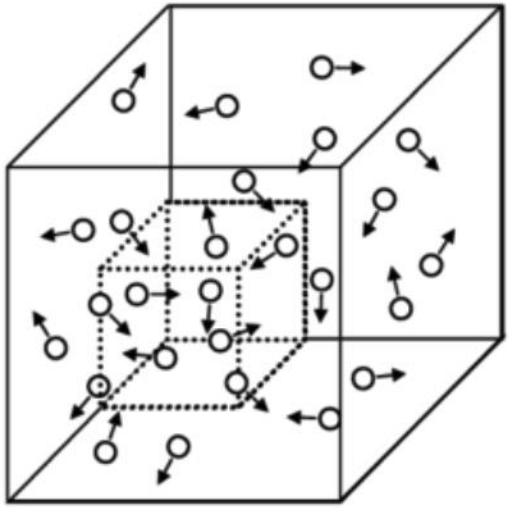
Coarse-grained force fields…

Choosing your thermodynamic ensemble

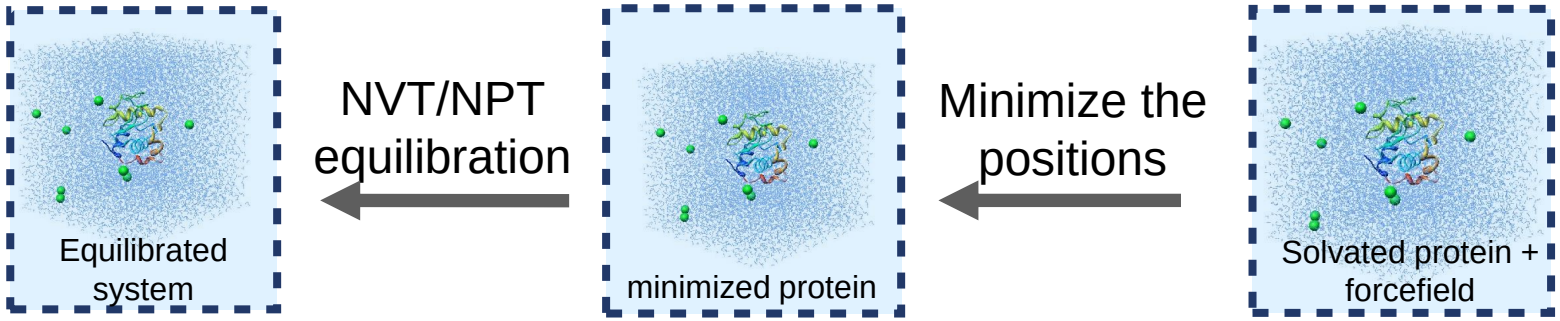
Simulations replicate a specific *thermodynamic ensemble* (typically NVT or NPT), or even grand canonical (μVT)

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
- …

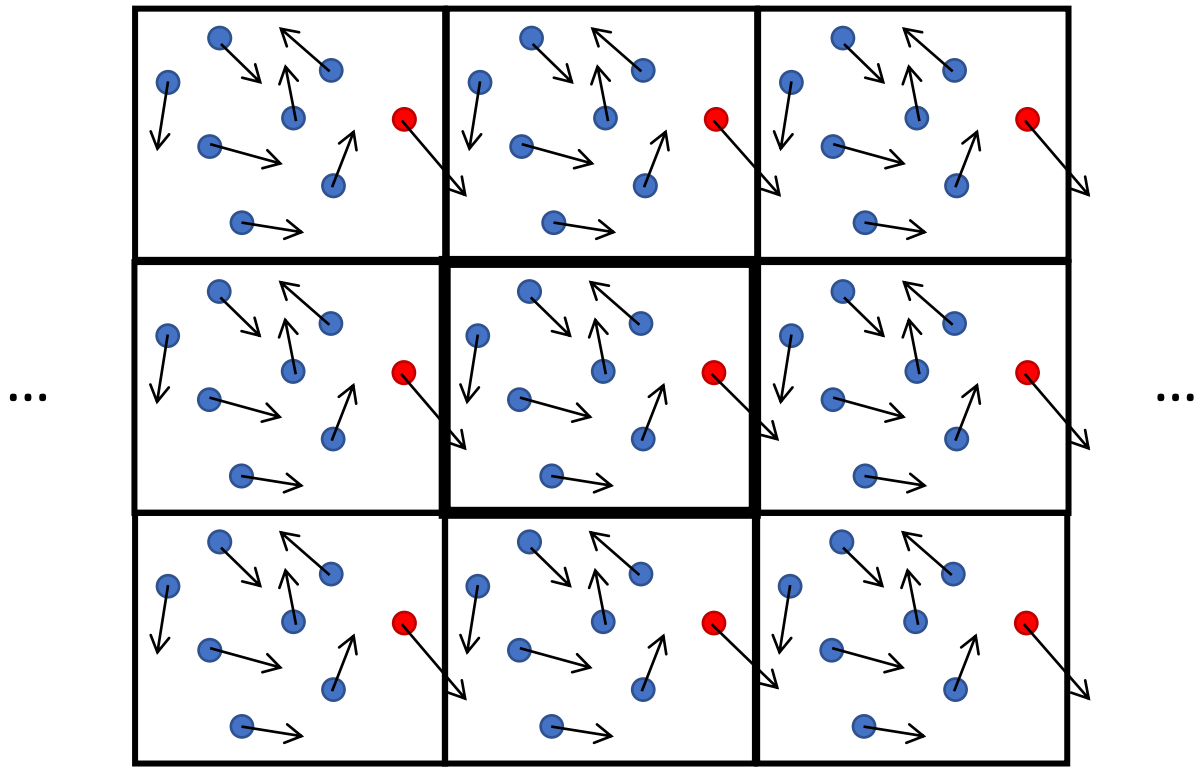


μVT

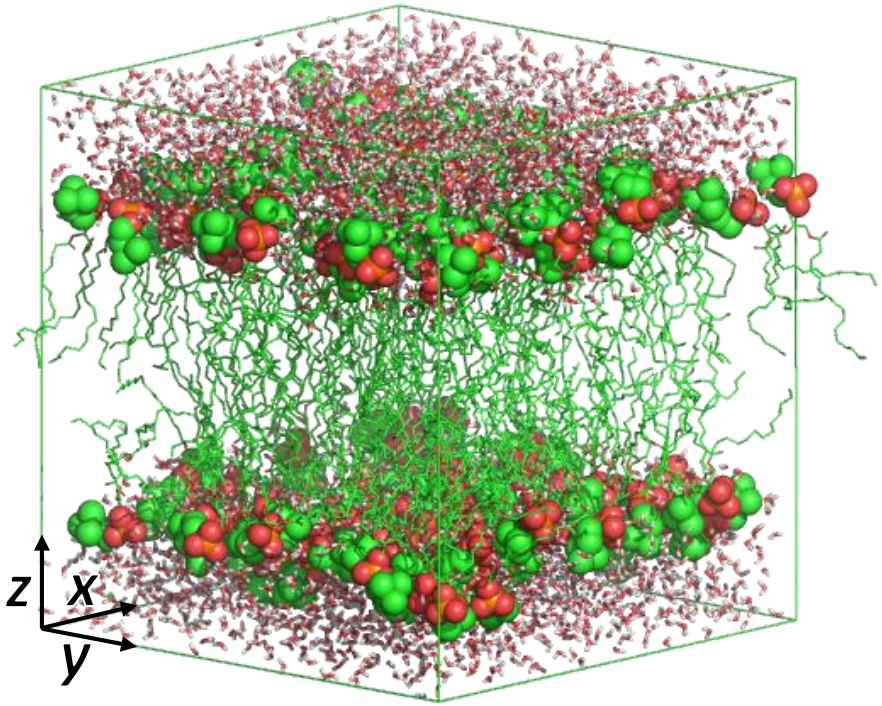


Periodic boundary conditions (PBC) and pressure coupling

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)

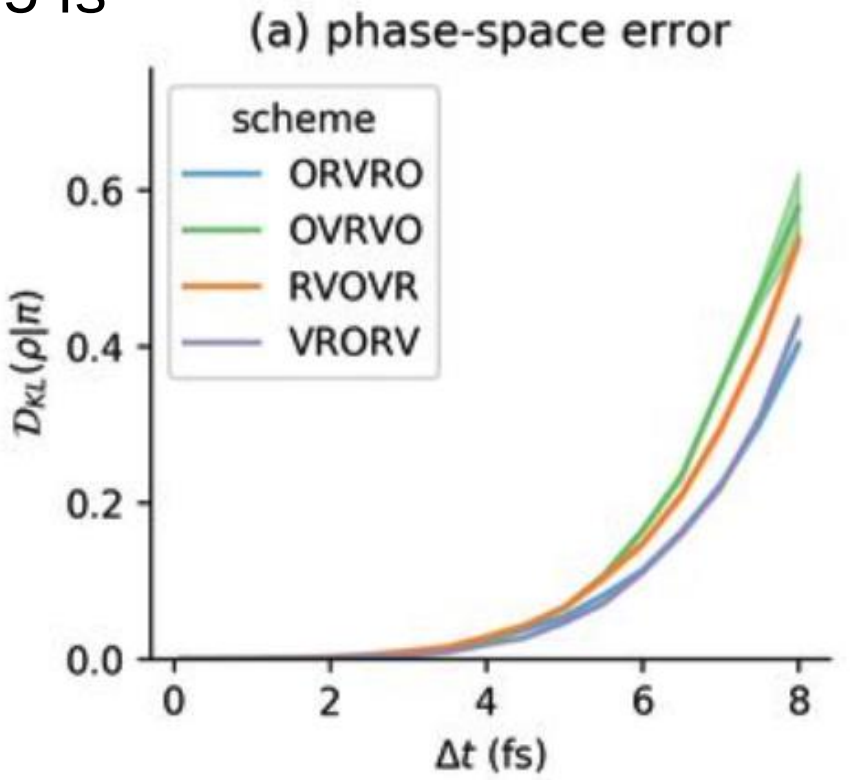
Sampling timescales for protein systems

The steepest gradient determines the smallest timestep.

Timestep size is imposed by the fastest phenomenon we want to observe.
In atomistic simulations:

- 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?

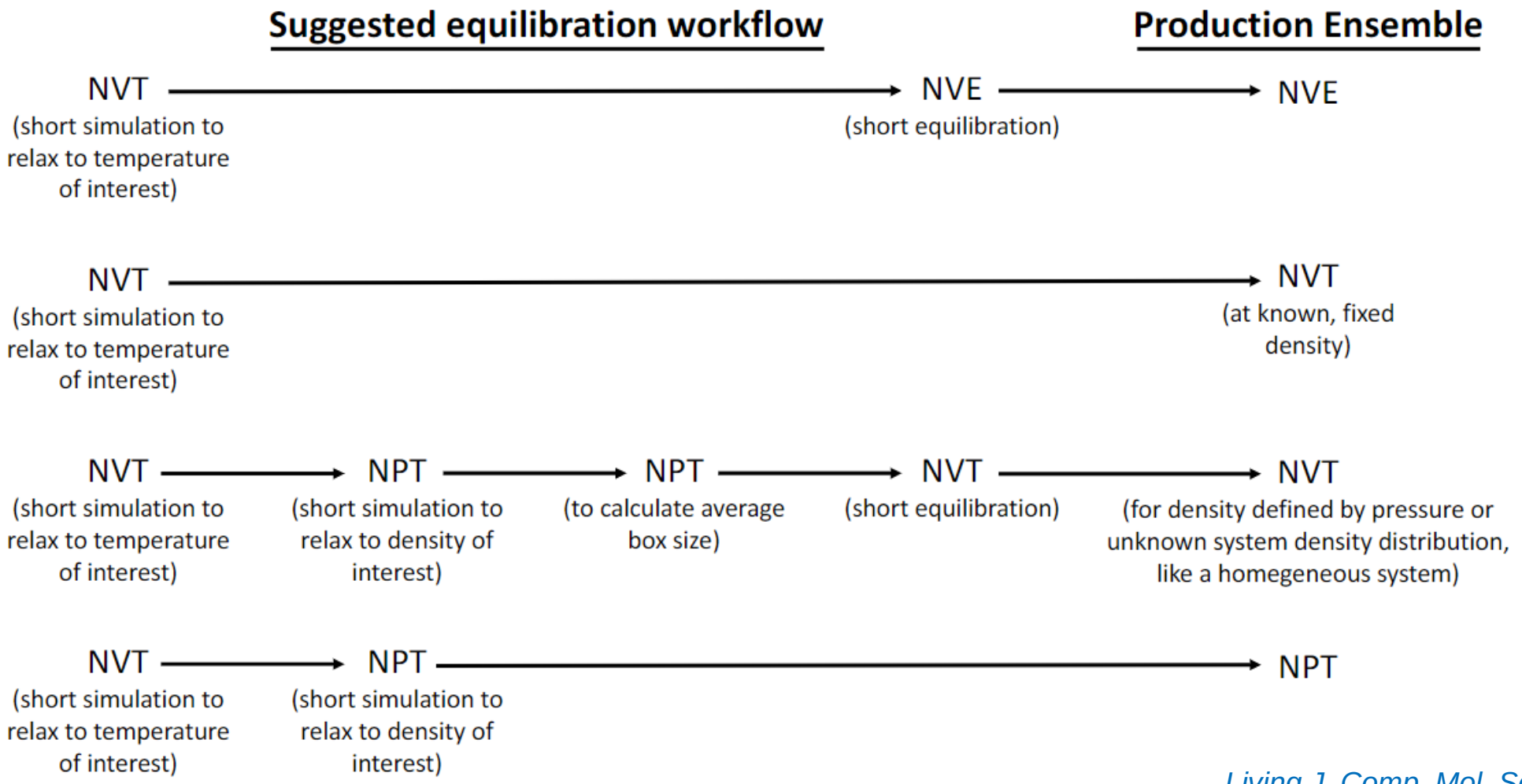
Equilibration phase: is the system in a “relaxed” state?

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)



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

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 μs 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

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

Which MD engine should I use?

Consider:

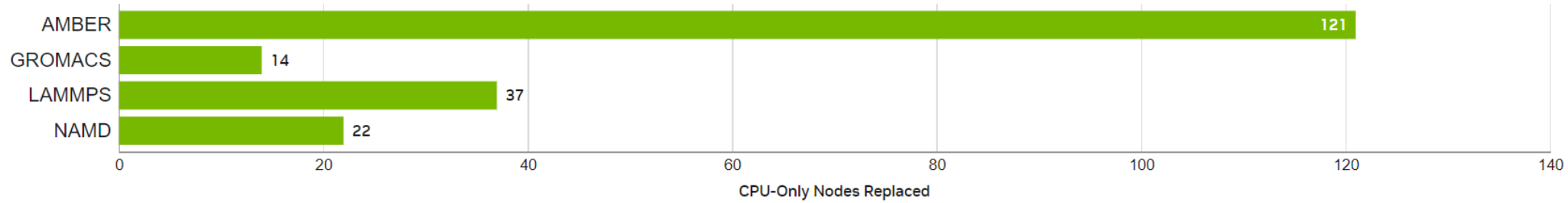
- Support for force field of choice
- Enables running desired simulation protocol
- Performance for available hardware
- Ease of use

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

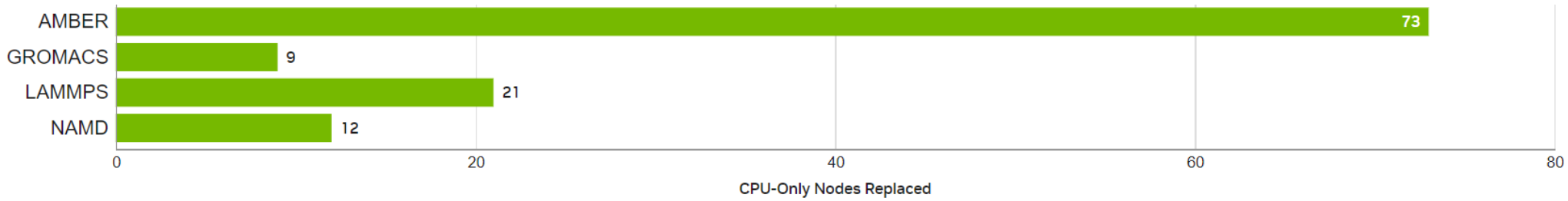
Graphical Processing Units (GPUs) are especially effective for MD

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

NVIDIA H200

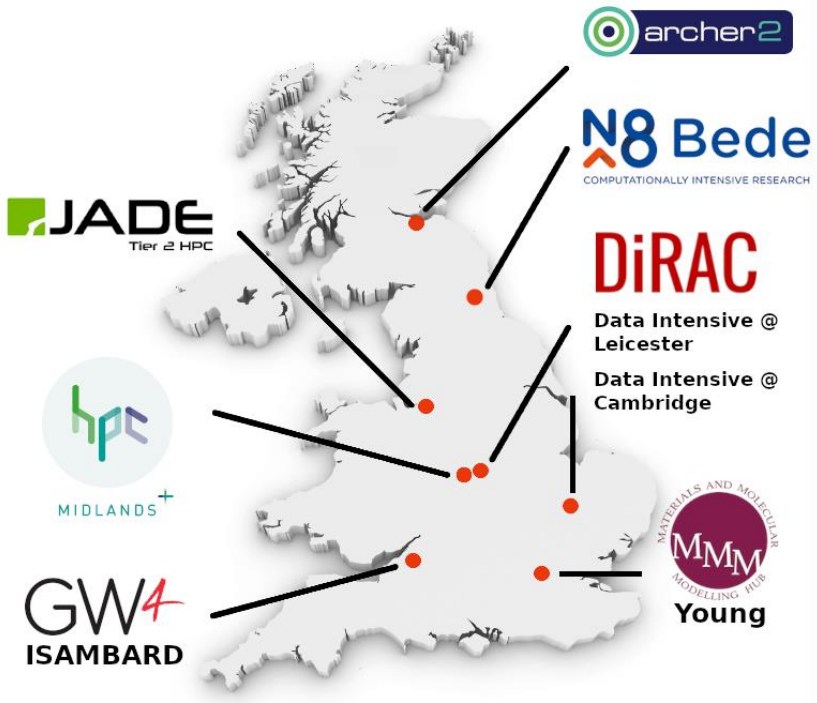


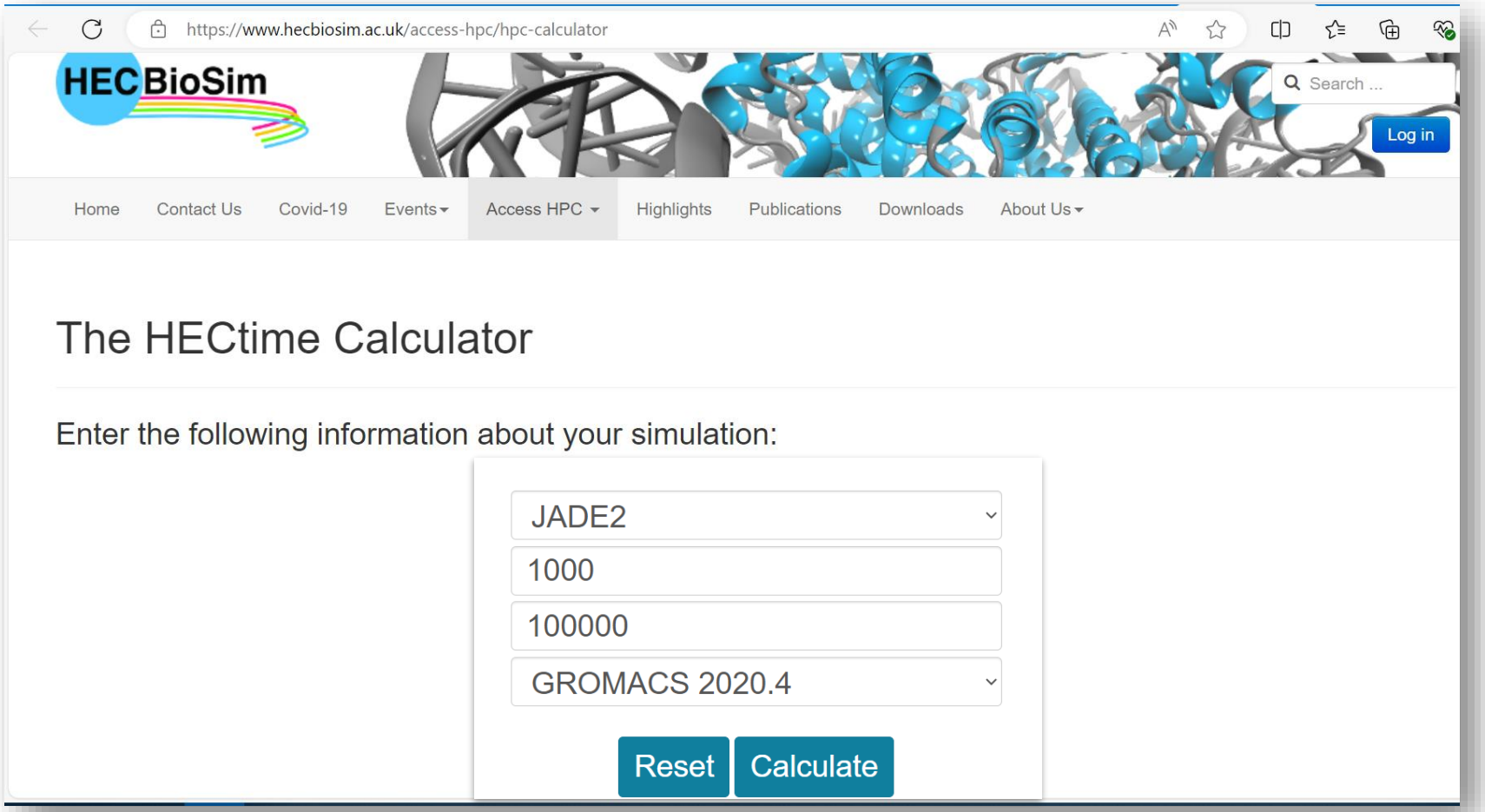
NVIDIA A100



CPU Server: Dual Xeon Platinum 8480+ @2GHz | GPU Server: Dual EPYC 7742@2.25GHz with 4x NVIDIA A100 SXM 80GB | AMBER Benchmark: DC-Cellulose_NVE, CUDA Version: 12.2 | GROMACS Benchmark: ADH Dodec, CUDA Version: 12.4 | LAMMPS Benchmark: Tersoff, CUDA Version: 12.2 | NAMD Benchmark: apoa1_nve_cuda, CUDA Version: 12.2

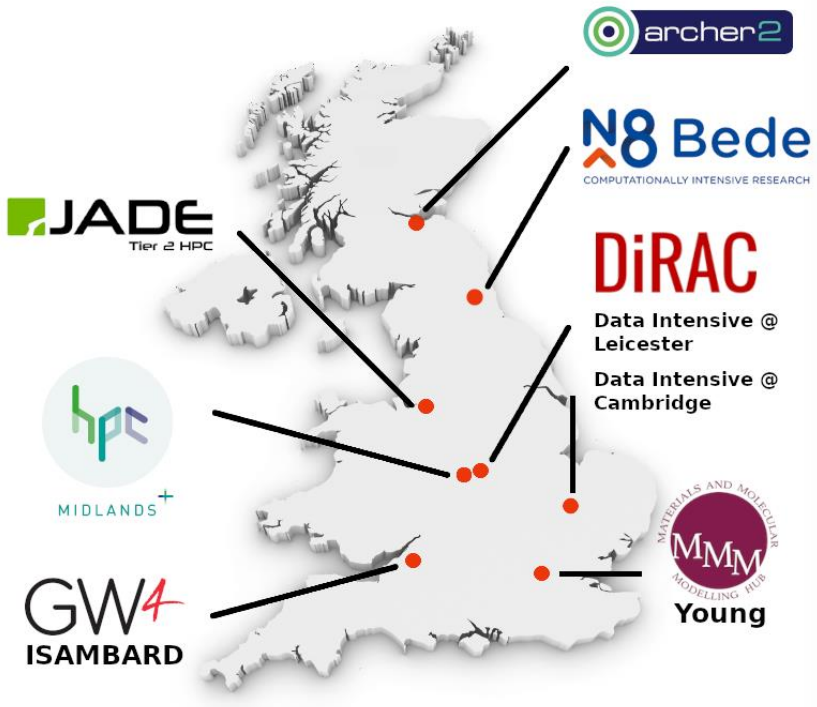
Calculating runtimes: example on UK Tier 2 systems



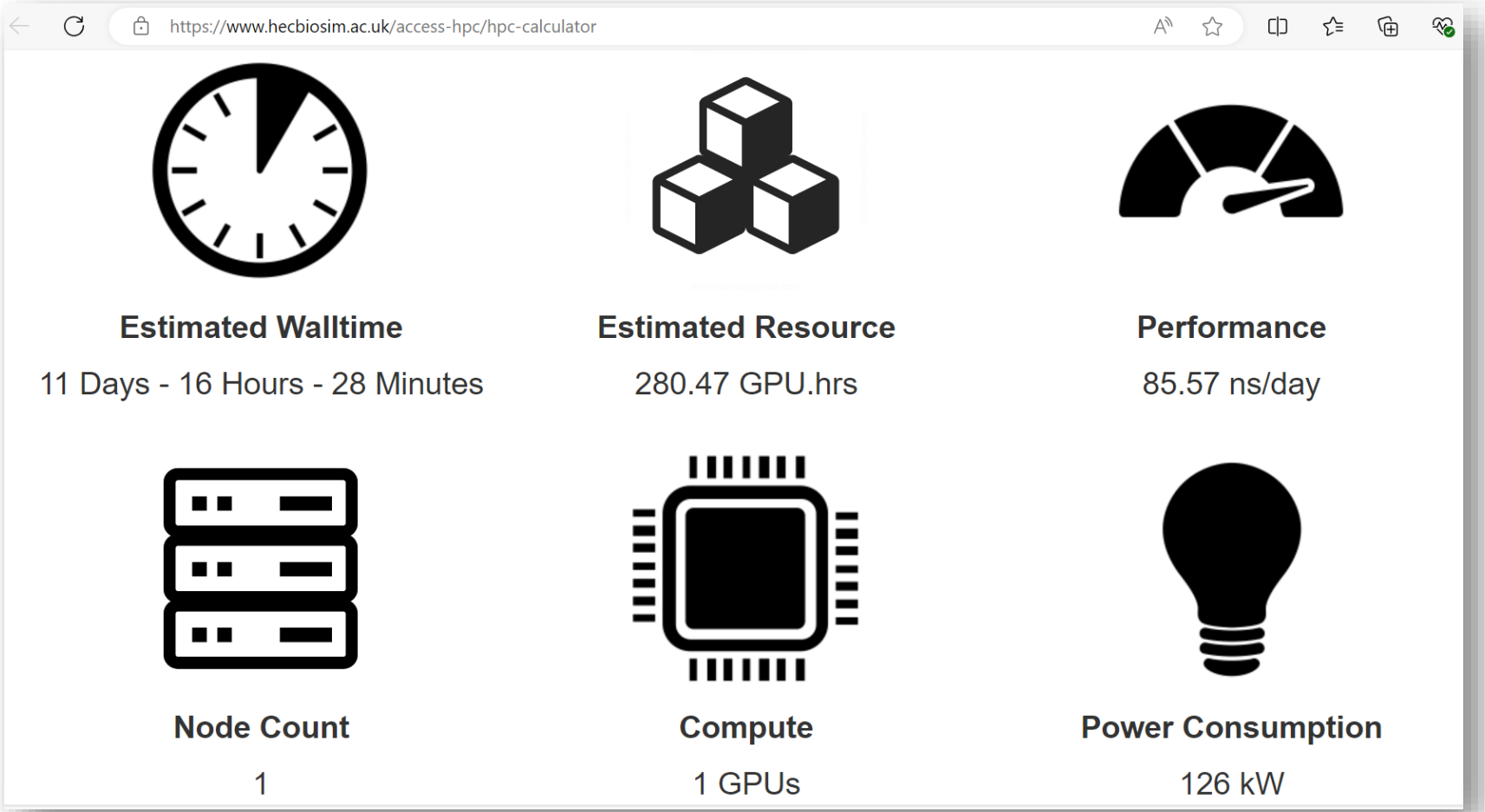


https://www.hecbiosim.ac.uk/access-hpc/hpc-calculator

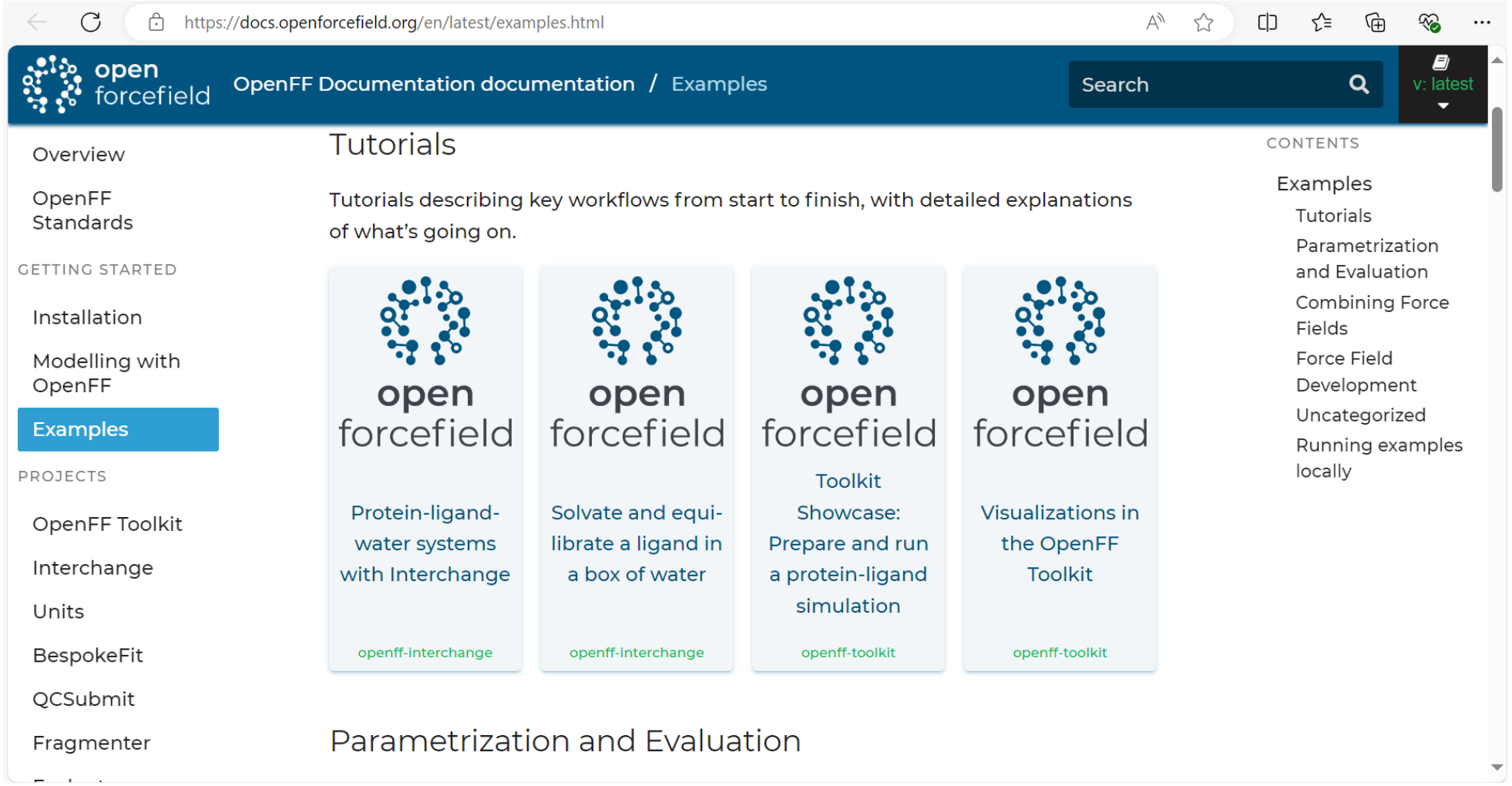
Calculating runtimes: example on UK Tier 2 systems



[as of 2024]



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



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

open forcefield