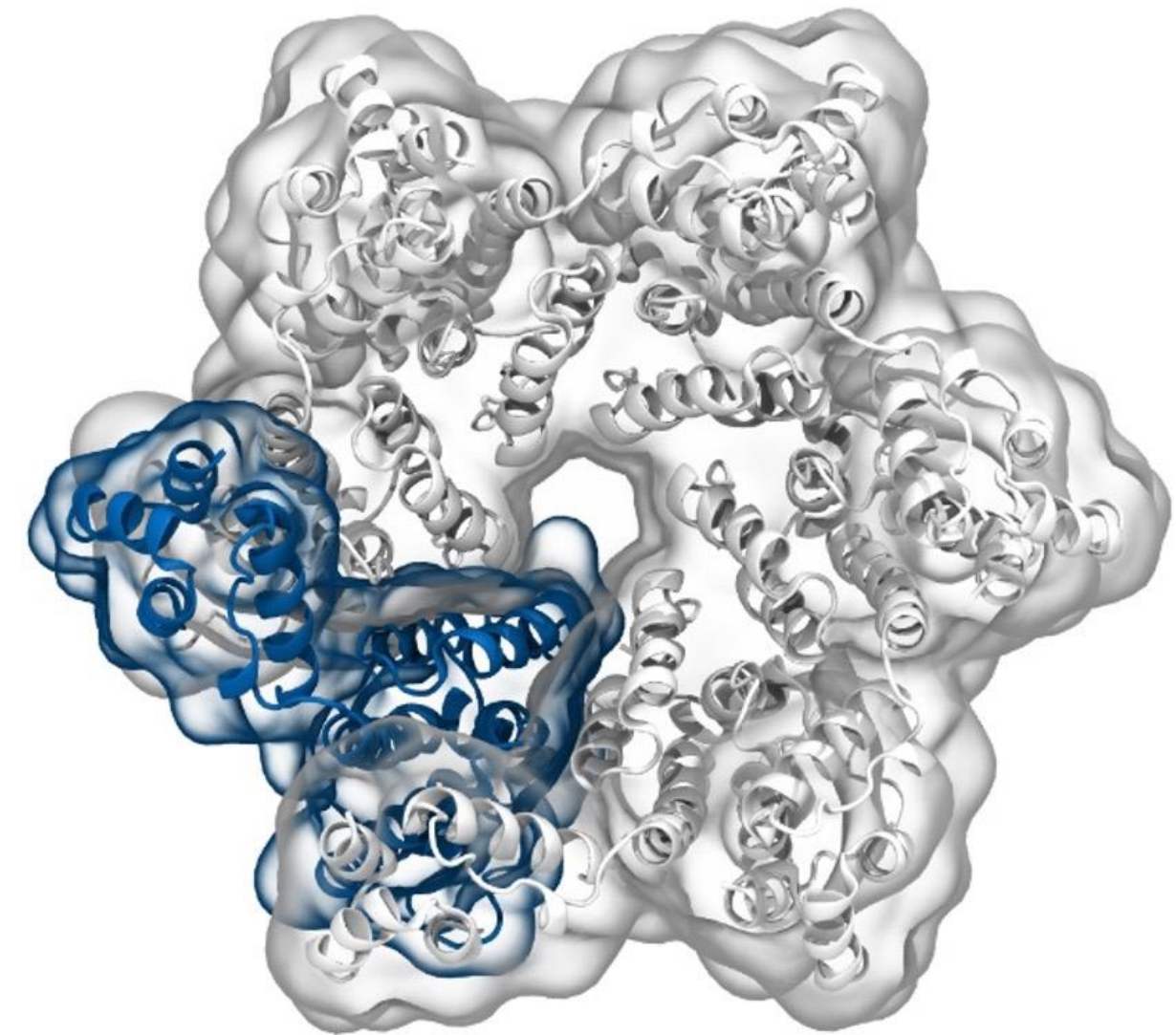
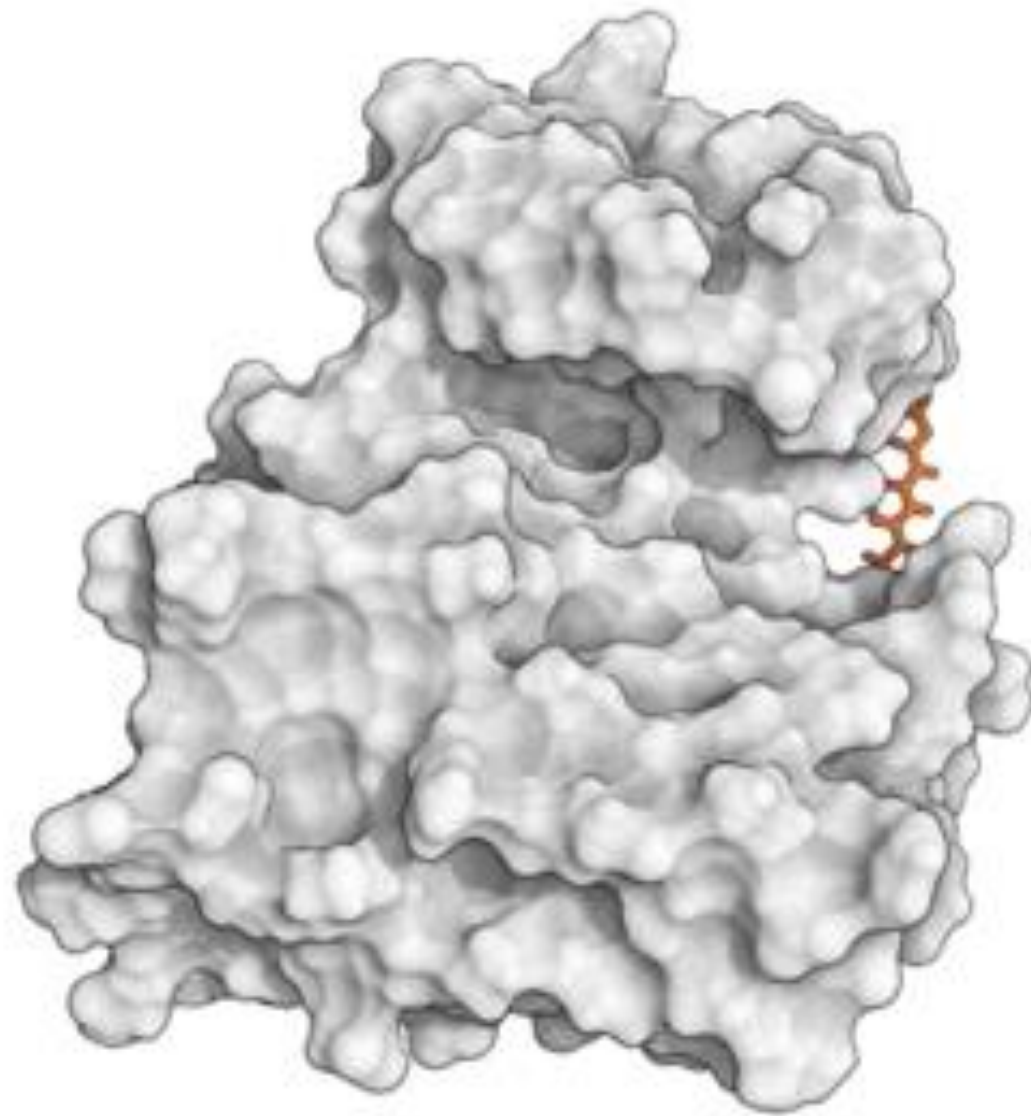


Simulation of Biomolecules

Setting up a protein simulation

2024 CCP5 Summer School



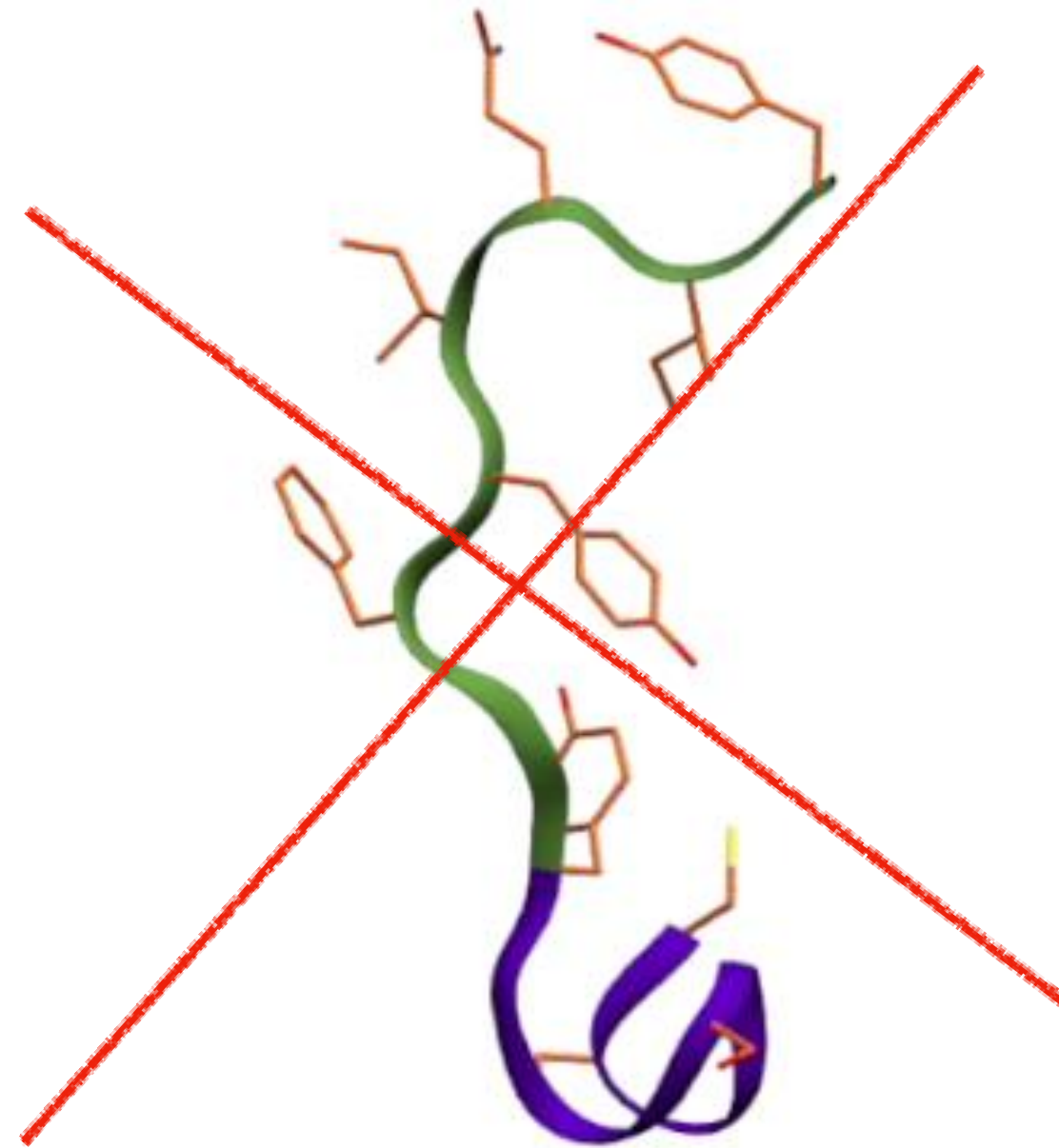
Dr Matteo Degiacomi
Durham University

matteo.t.degiacomini@durham.ac.uk

Dr Antonia Mey
University of Edinburgh

antoniamey@ed.ac.uk

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



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 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

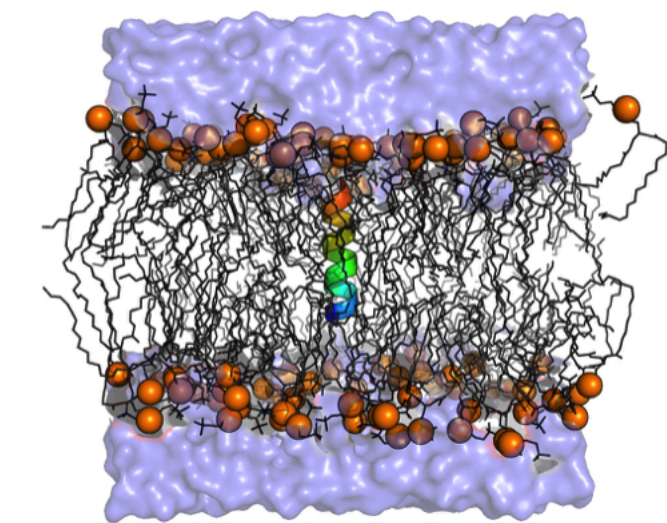
From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]

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>

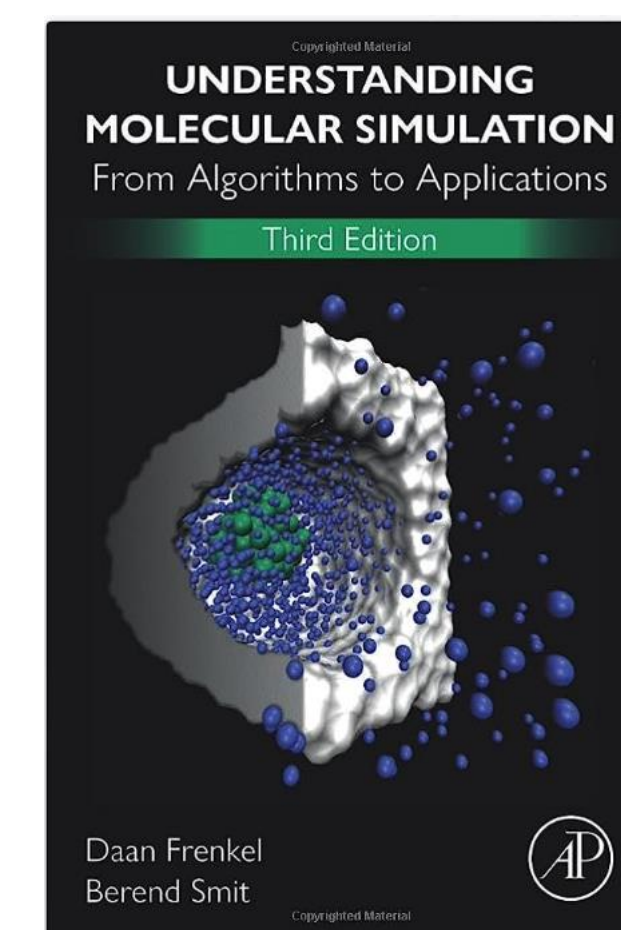
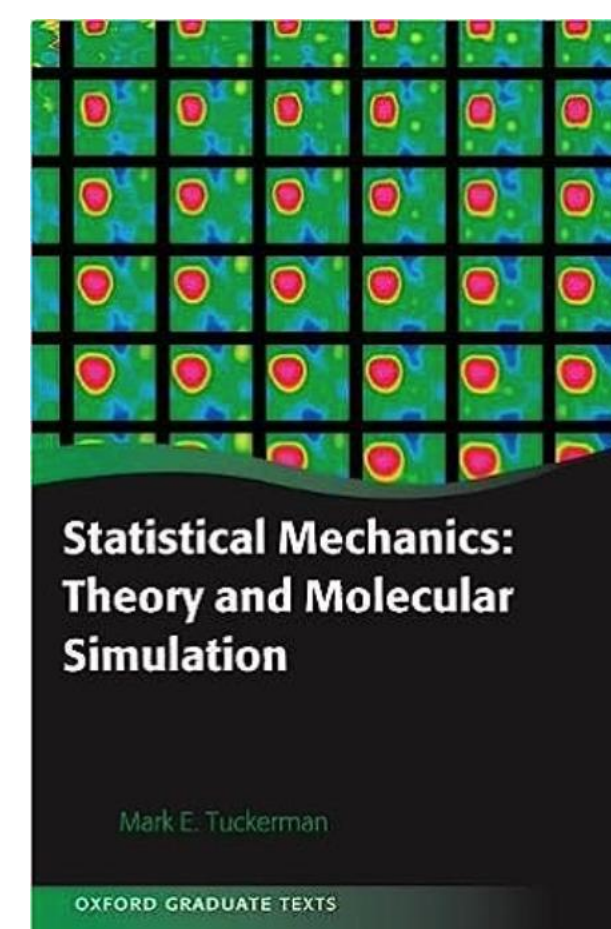
Keywords: tutorials, gromacs, molecular dynamics simulation, computational chemistry



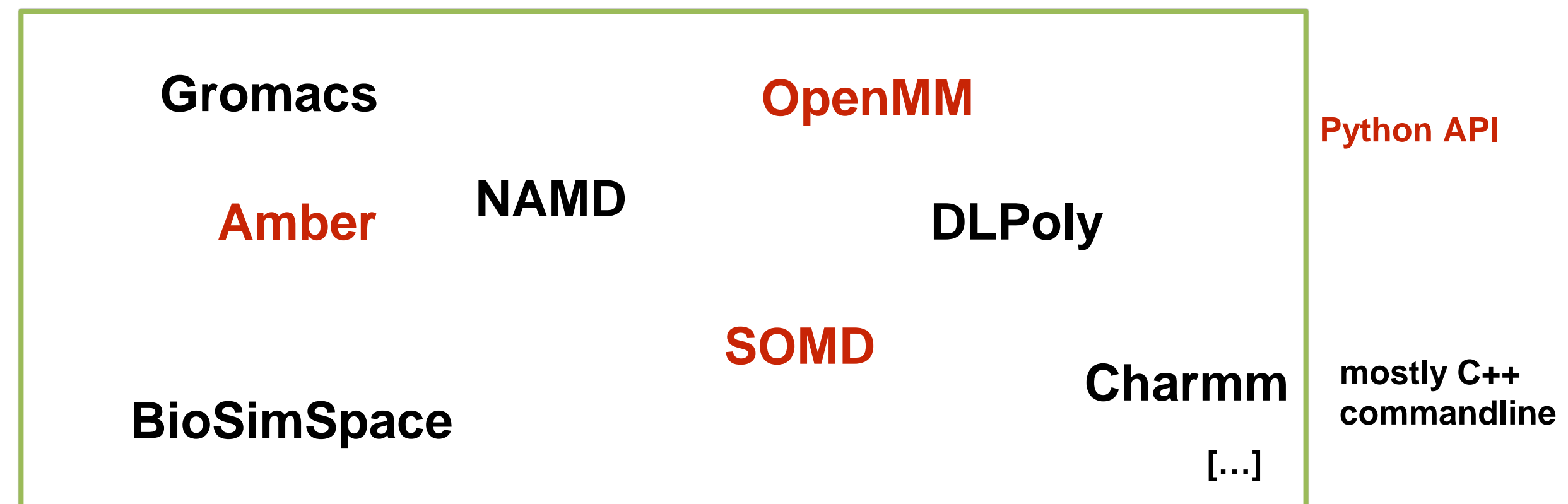
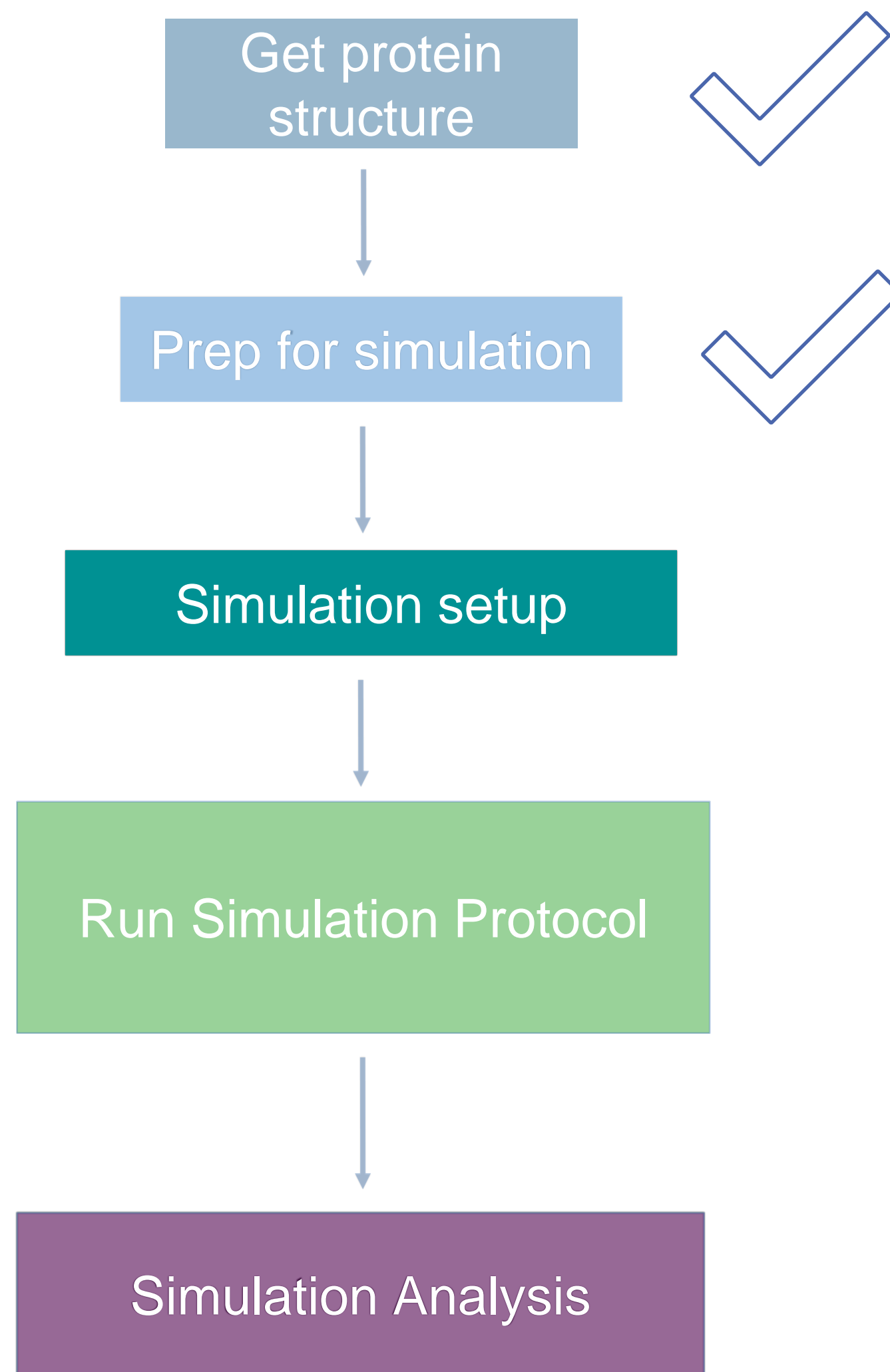
PDF

ARTICLE CODE REPOSITORY

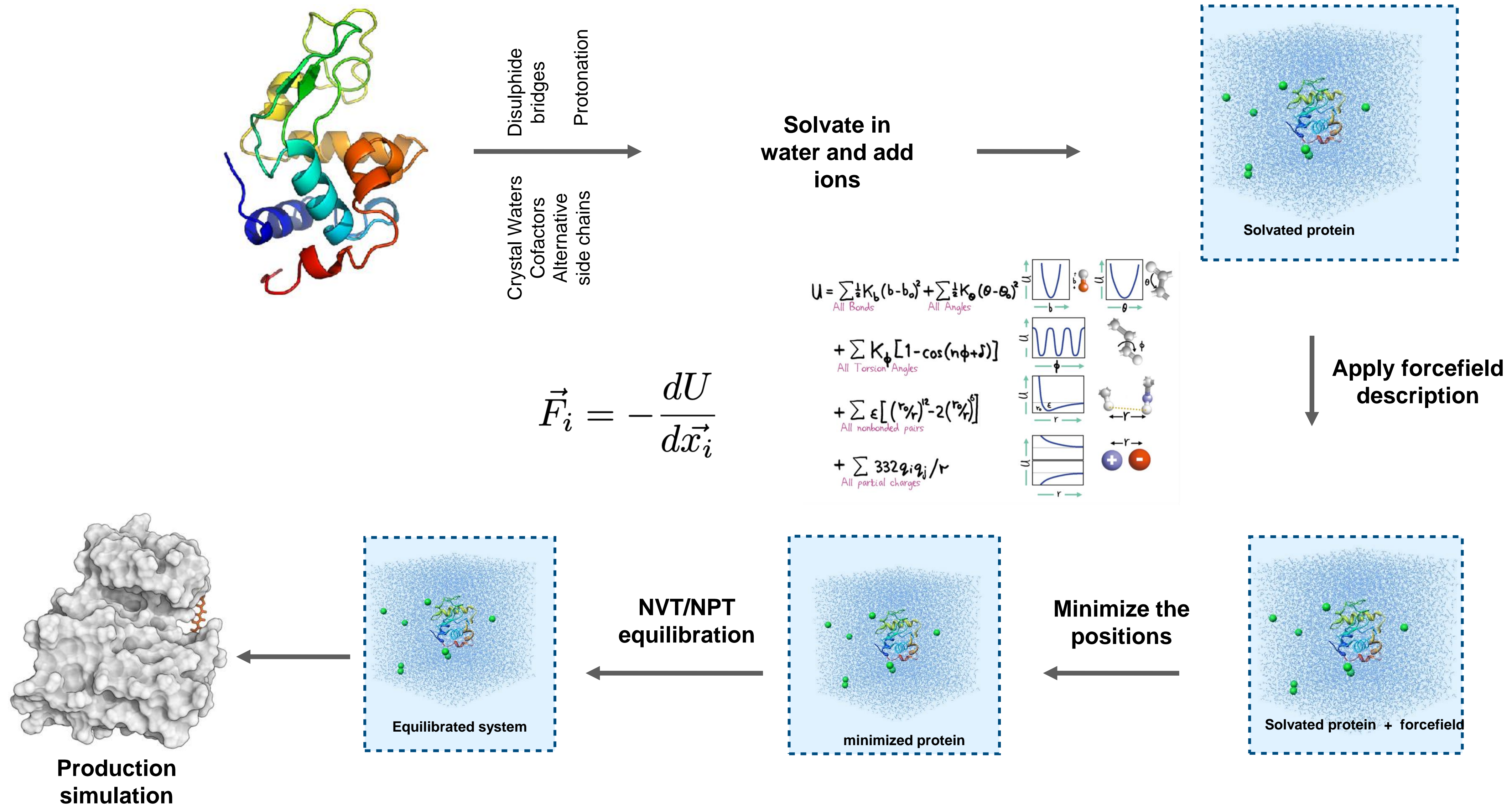
ABSTRACT



A typical workflow for molecular dynamics

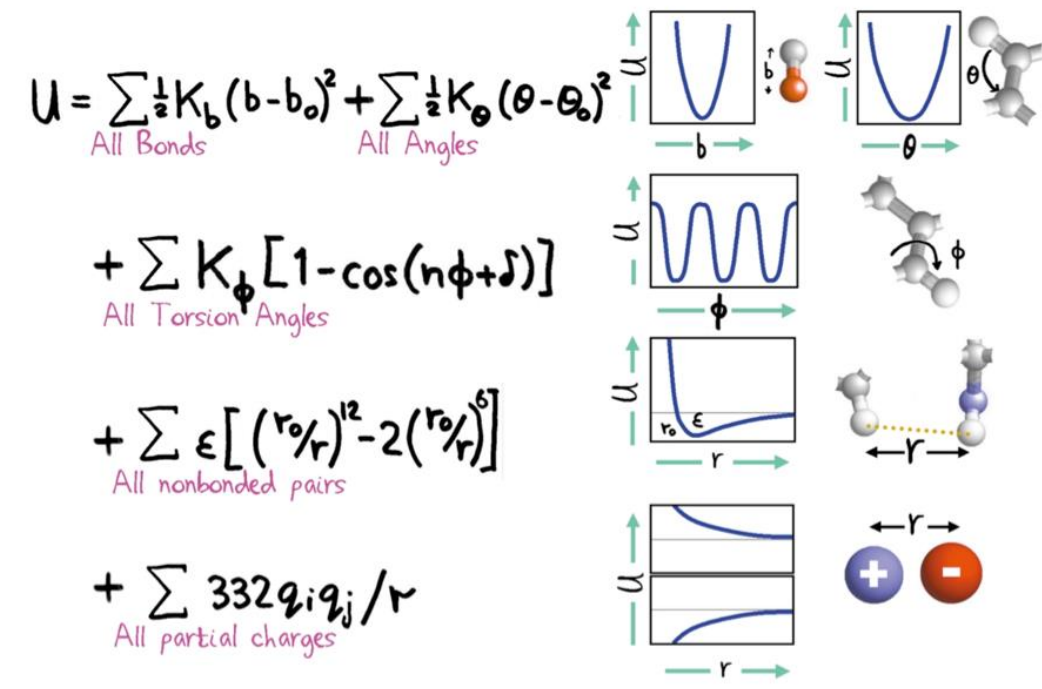


Molecular dynamics require multiple steps for the setup of simulations



There are many different choices for force fields to be made

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

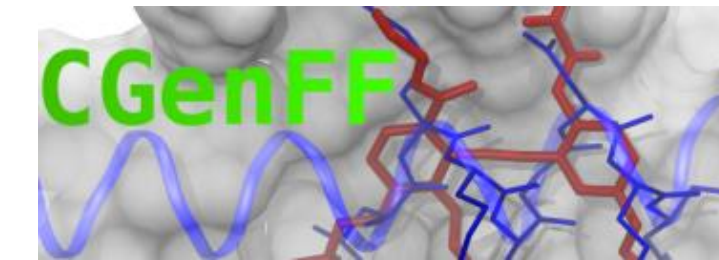


Proteins and other constituents need different force fields

- **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

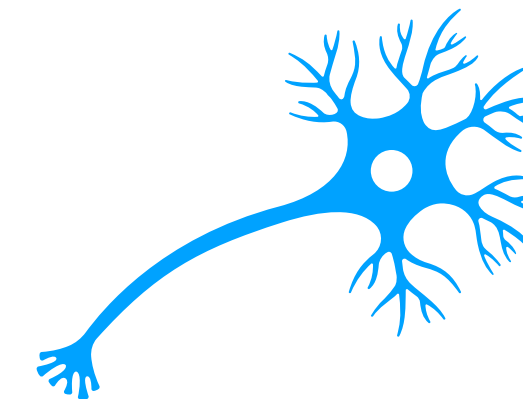


$$E_{\text{pair}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 +$$

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

GAFF

Machine learned force fields



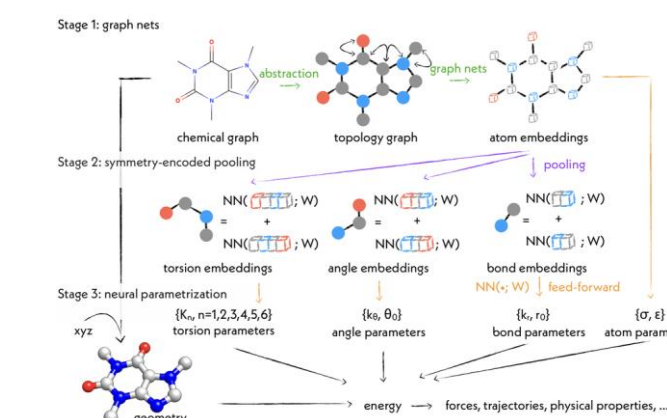
Ani-2x

SchNet

PhysNet

BandNN

Espaloma



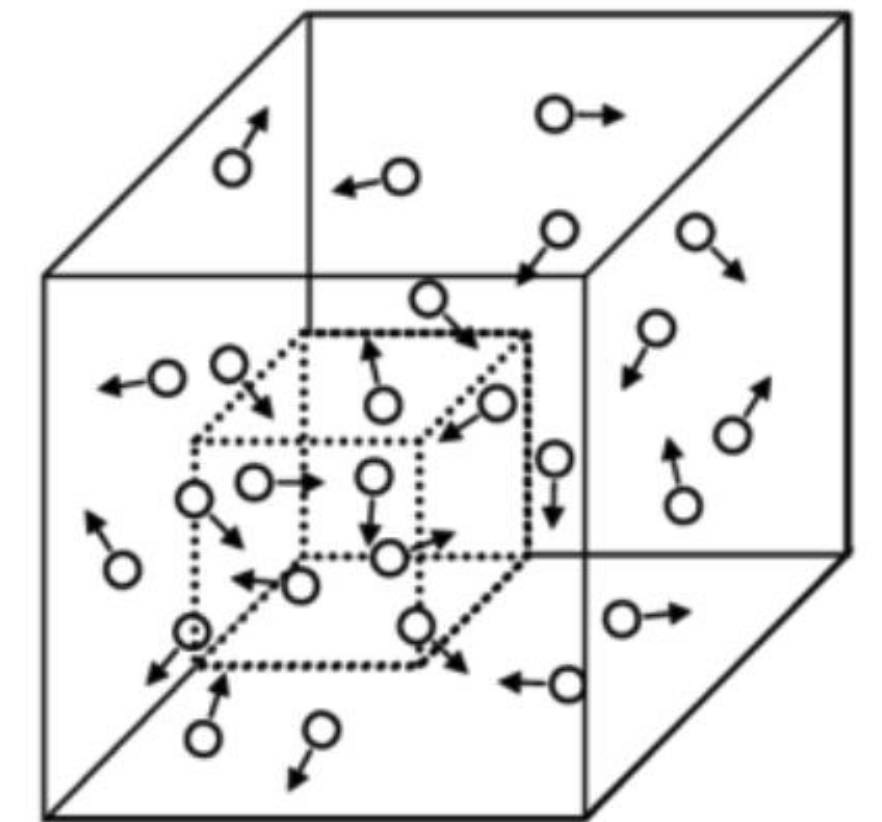
Coarse-grained force fields....

Choosing your thermodynamic ensemble

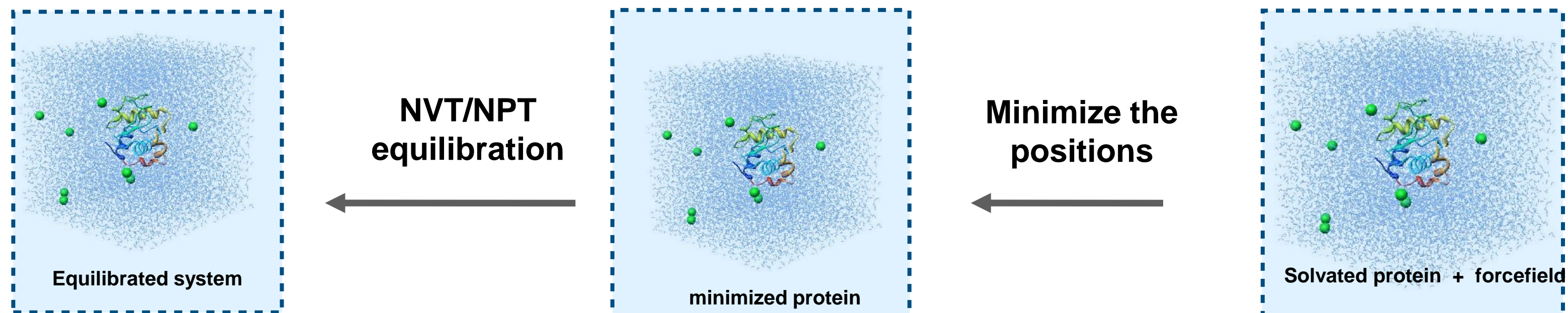
Simulations replicate a specific *thermodynamic ensemble* (typically NVT or NPT), or even grand canonical (μ VT)

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

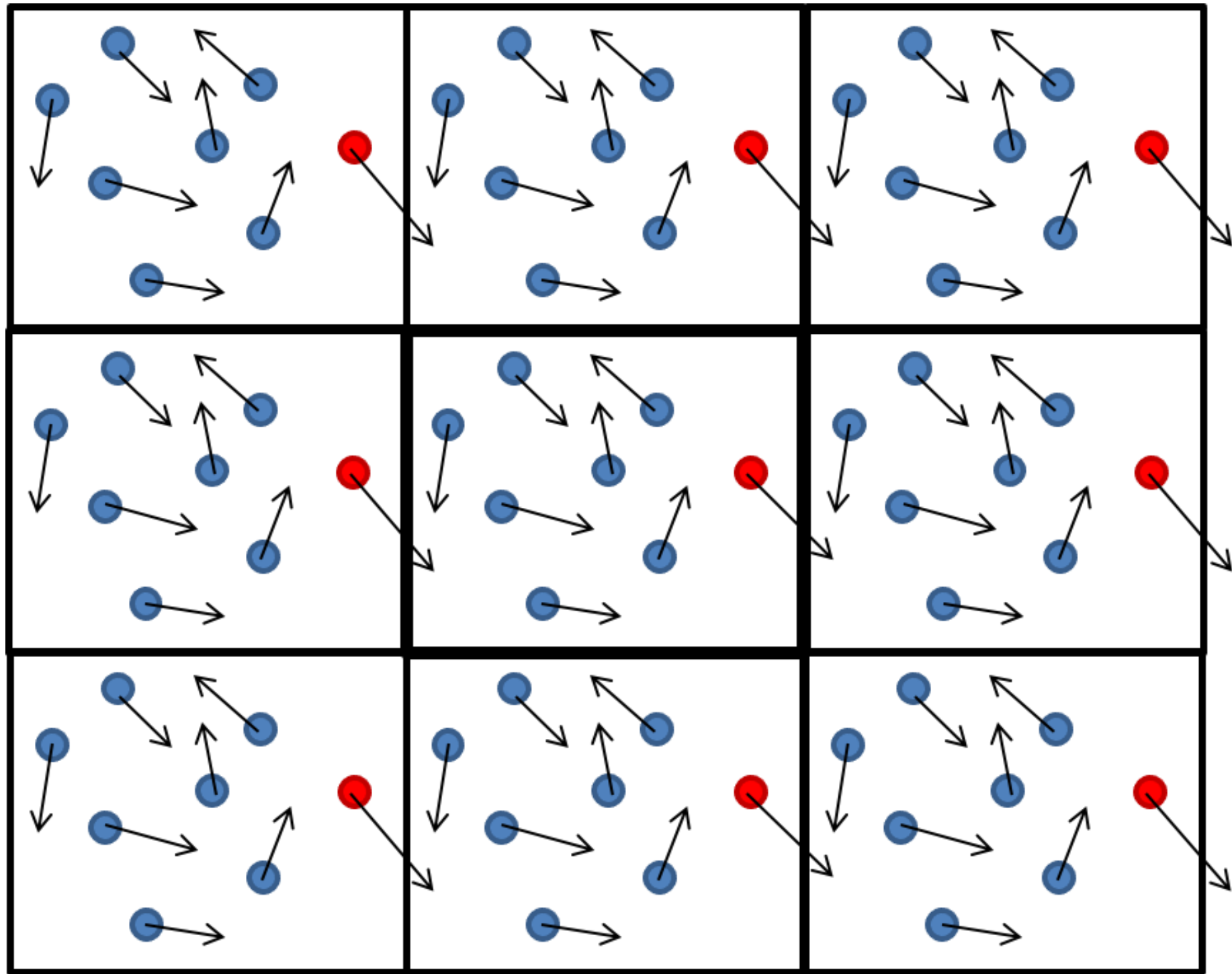
- Nose-Hoover
- Berendsen
- Parrinello-Rahman
- Langevin piston
- ...



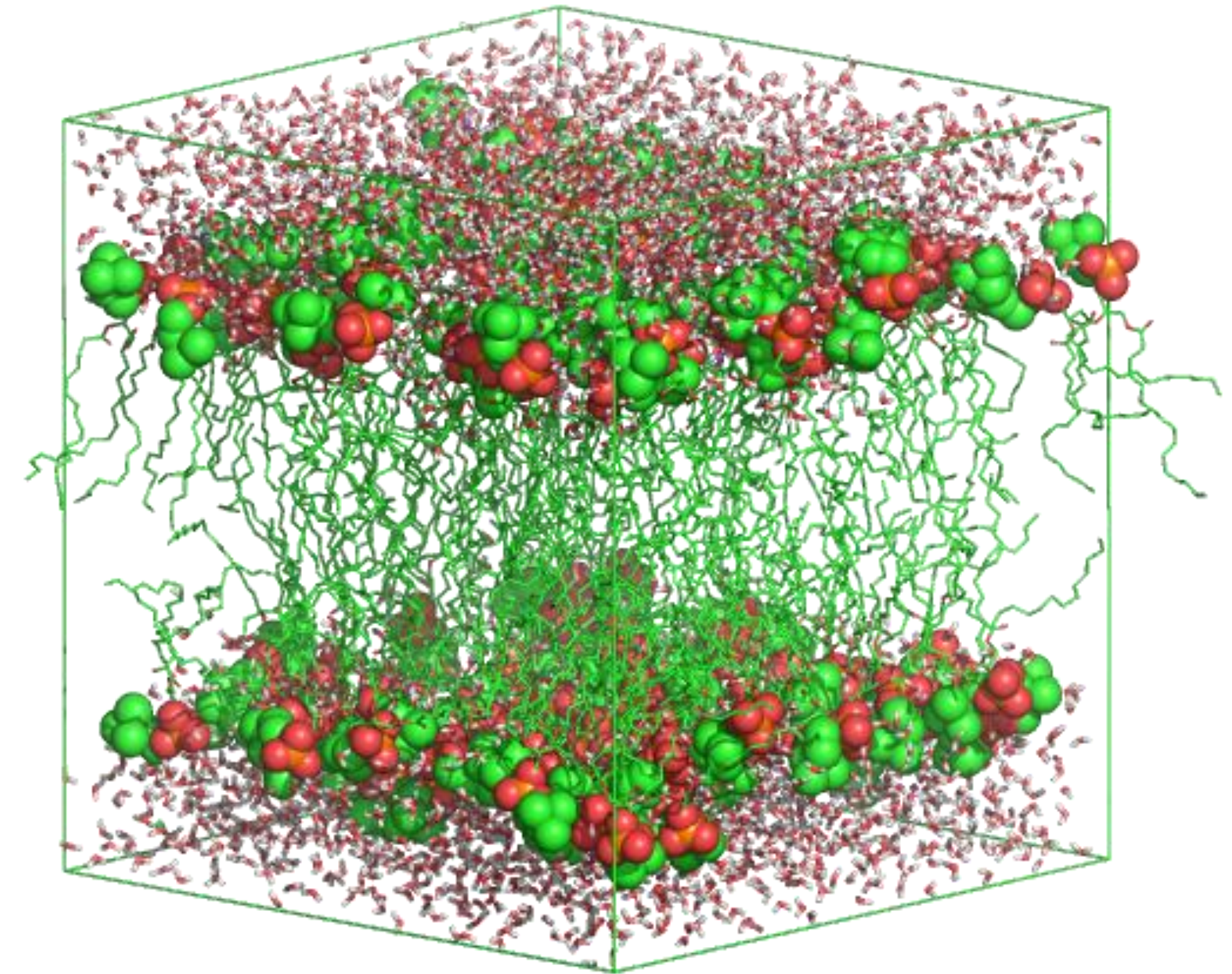
μ VT



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:

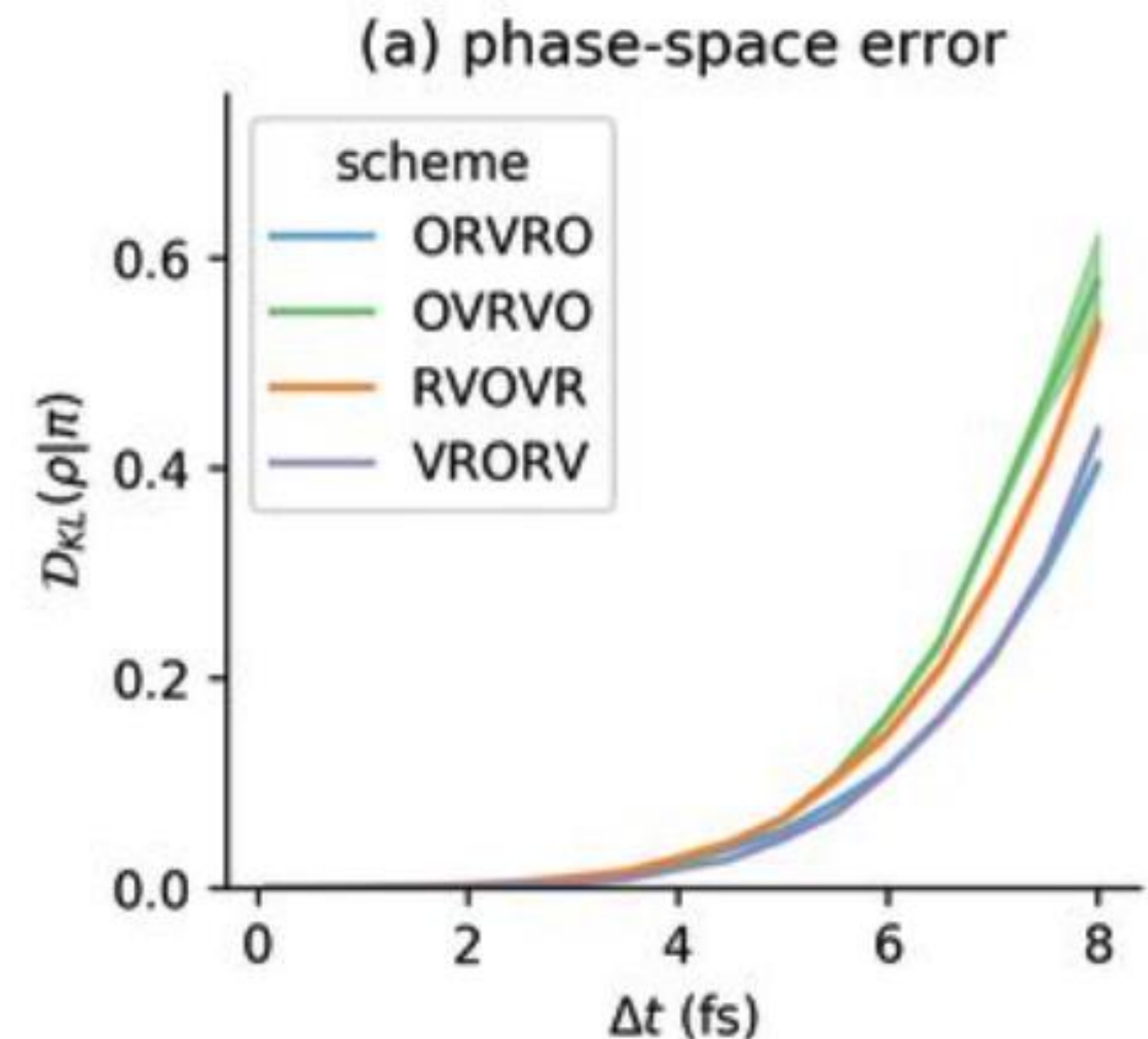
Timestep size is imposed by the fastest phenomenon we want to observe :

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

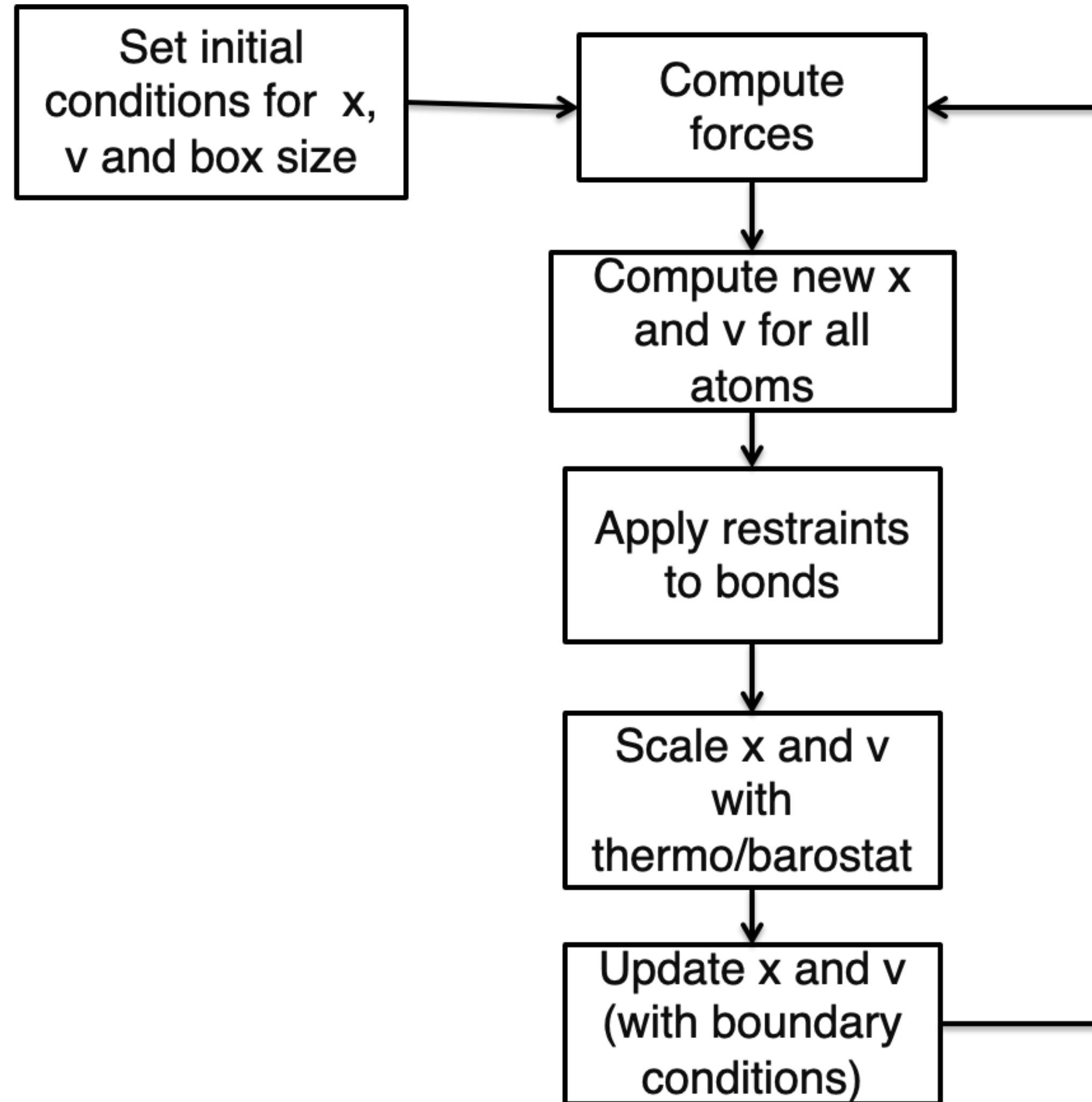
Restraining covalent bond distances allows to use 1-2 fs timesteps (restraining methods: SHAKE, RATTLE, LINCS,...)

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