bioexcel
Centre of Excellence for Computational Biomolecular Research

Horizon 2020
European Union Funding
for Research & Innovation

Brief introduction to GROMACS

Alessandra Villa
PDC-Center for High Performance Computing,
KTH-Royal Institute of Technology,
Stockholm, Sweden

avilla@kth.se

Molecular dynamics

GROMACS

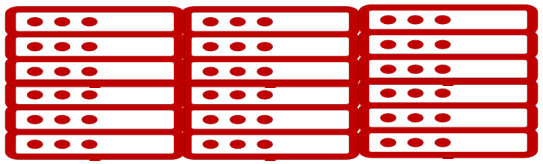
One goal of a molecular dynamics simulation is to generate enough representative conformations of the molecular system in such a way that accurate values of a property can be obtained.

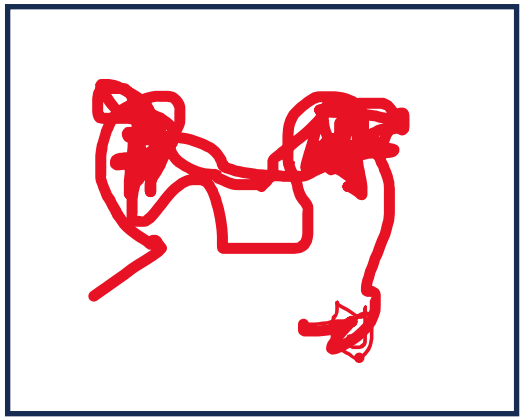
The ergodic hypothesis states $\langle A \rangle_{ensemble} = \langle A \rangle_{time}$

Two different ways

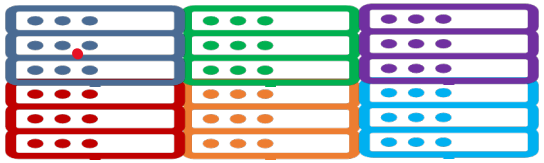
to sample the conformational space.

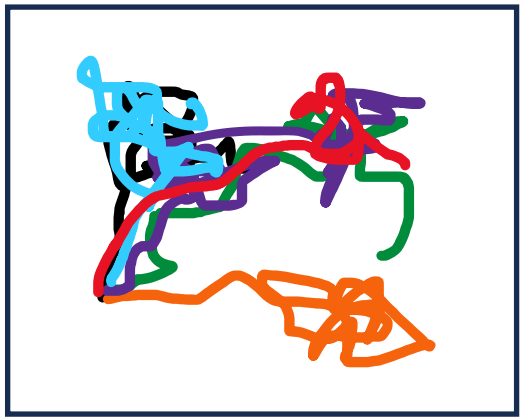
- One long simulation





- Many shorter simulations
- (*ensemble simulation)*

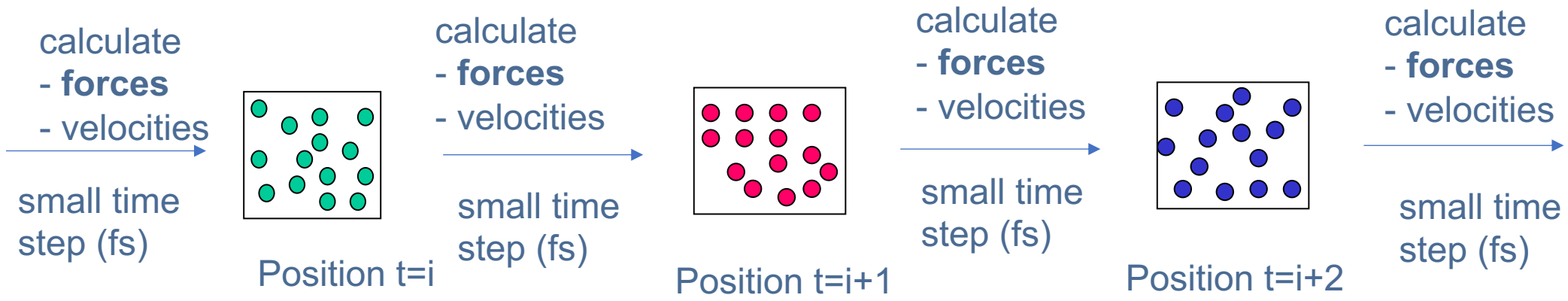




How does MD generate conformations

GROMACS

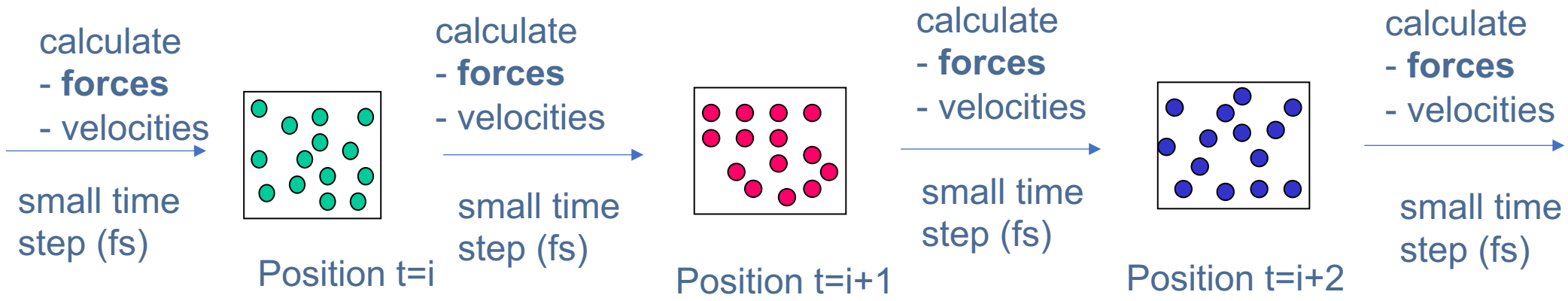
How? by iteratively solving equations of motion



How does MD generate conformations

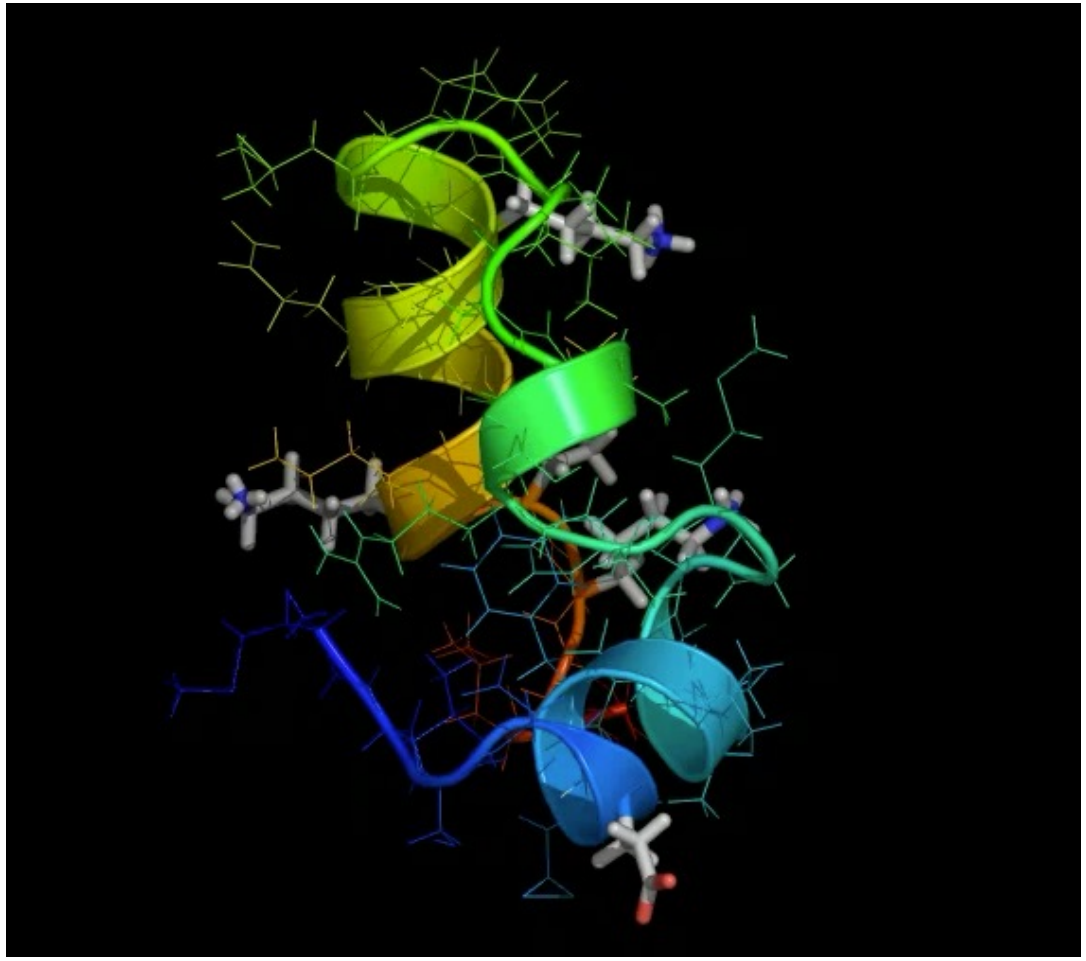
GROMACS

How? by iteratively solving equations of motion



Calculating force = main computational cost

Phase space sampling



Newton's equation of motion

acceleration = force /mass

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$

force = -slope of the potential

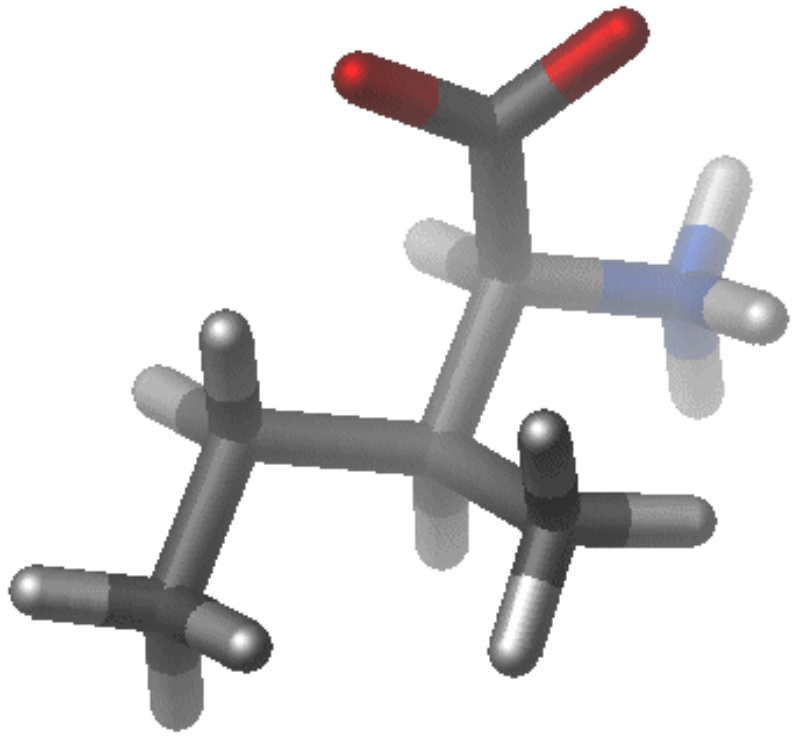
$$F_i = -\frac{\partial}{\partial r_i} V(r_1, r_2 \ldots\ldots\ldots r_N)$$

$$i = 1 \ldots\ldots \mathrm{N}_{\mathrm{particles}}$$

One small step for a human,
many small steps for a computer

Molecular Mechanics force field

A simplify representation of a molecular system should be as simple as possible



See Nobel lectures in chemistry 2013. (Michael Levitt)

Force field and analytical functions

$$V(r_1, r_2, ..., r_N) = \sum_{bonds} \frac{1}{2} K_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{torsions} \frac{1}{2} K_\xi (\xi - \xi_0)^2$$

$$+ \sum_{torsions} \frac{1}{2} K_\phi [1 + \cos(n\phi + \delta)]$$

bonded-interactions

non-bonded interactions

$$+ \sum_{pairs} [C_{12}(i,j) / r_{i,j}^{12} - C_6(i,j) / r_{i,j}^{6}] + q_i q_j / 4\pi\varepsilon_0 \varepsilon r_{i,j}$$

Every force field has his own analytical functions and set of parameters

Force field

Every force field has his own analytical functions and set of parameters, and also

- time step
- constraints
- non-bonded short-range interaction (e.i cut-off)
- non-bonded long-range interaction (PME, reaction field)

depends on force field choice (but set in mdp file)

High computational cost

GROMACS

Some interactions are more costly than others

- Non-bonded interactions
  - Calculated over every pair of atoms in the system
  - ~ to N^2 where N is the number of atoms in the system
  - More than **90% of the computing time**

=> cut-off, PME (mdp parameters)

=> pair neighbor searching is expensive operation (not every step, buffer, cluster particles => tolerance on the energy drift)

- nstlist = frequency to update the neighbor list (e.i on GPU a values of 20-40)

Other forces in MD simulations

Restraints

used for imposing restraints on the motion of the system: **Position** restraints, **Flat-bottomed position** restraints: sphere, cylinder, layer, **Angle and dihedral** restraints, **Distance restraints**: simple harmonic, piecewise linear/harmonic, complex NMR distance, **Orientation** restraints

Other external forces:

for example used to drive the system along a specific reaction coordinates or toward specific conditions

How to get good performance

- Optimal mdp parametes
  - Currently most mdp parameters do not affect performance much (except PME order and grid).
  - Automated PME tuning optimises the Coulomb cut-off and PME grid size (in GROMACS)

- Choose good options for mapping tasks in mdrun to available hardware
  => effect on performance but not easy

Two aspects :
time spent to calculate forces vs time spent to wait to transfer data

Thermostats & barostats

GROMACS

For thermostats and barostats we need to compute the temperature or pressure. This requires global communication and is currently not done on the GPU => expensive

•To reduce these costs, we now set nsttcouple and nstpcouple to 100 by default, unless needed more frequently:

- For v-rescale and c-rescale: ≥ 5 steps per tau_t/p
- For Nose-Hoover and Parrinello-Rahman: ≥ 20 steps per tau_t/p

recommendation, use:

•v-rescale for temperature coupling, with tau_t = 1 ps

•c-rescale for pressure coupling, with tau_p = 5 ps

DOI 10.5281/zenodo.7863174

How/why to get good performance

Important for you and for everybody

=> but there is not recipe

The range of possible simulations and available hardware is huge

=>

the automation often doesn't get the best result.

=>

the users need to get involved and guide the choice.

FAST. FLEXIBLE. FREE.
GROMACS

Authors/Contributors

These people have contributed to the 2023 release of GROMACS:

Mark Abraham; Andrey Alekseenko; Cathrine Bergh; Christian Blau; Eliane Briand; Mahesh Doijade; Stefan Fleischmann; Vytautas Gapsys; Gaurav Garg; Sergey Gorelov; Gilles Gouaillardet; Alan Gray; M. Eric Irrgang; Farzaneh Jalalypour; Joe Jordan; Christoph Junghans; Prashanth Kanduri; Sebastian Keller; Carsten Kutzner; Justin A. Lemkul; Magnus Lundborg; Pascal Merz; Vedran Miletić; Dmitry Morozov; Szilárd Páll; Roland Schulz; Michael Shirts; Alexey Shvetsov; Bálint Soproni; David van der Spoel; Philip Turner; Carsten Uphoff; Alessandra Villa; Sebastian Wingbermühle; Artem Zhmurov; Paul Bauer; Berk Hess; Erik Lindahl

User information and communication

GROMACS documentation -> https://manual.gromacs.org/
GROMACS forum -> https://gromacs.bioexcel.eu/
GROMACS tutorials -> https://tutorials.gromacs.org/
GROMACS webpage -> http://www.gromacs.org/

GROMACS GitLab -> https://gitlab.com/gromacs

Thank you

Co-funded by
the European Union

EuroHPC
Joint Undertaking

GROMACS

Molecular simulation: Input files

Information on the positions (and velocities) of the atoms

structure file (*pdb *gro)

Information about the molecular model

topology file (*top *itp)

Information on how we want to run the simulation

molecular dynamics parameter file (*mdp)

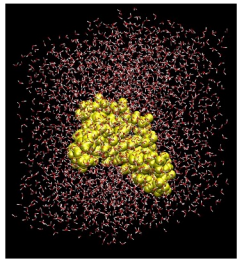
Molecular simulation in GROMACS

Input files

Structure
(*.gro, *.pdb)

Simulation parameter
(*.mdp)

Topology (*.top)



gmx grompp

*.tpr file

gmx mdrun

Output files

Ensemble of structures,
called configurations:
.xtc/.trr -> trajectory,
*.edr -> energy file,
*.log -> log file