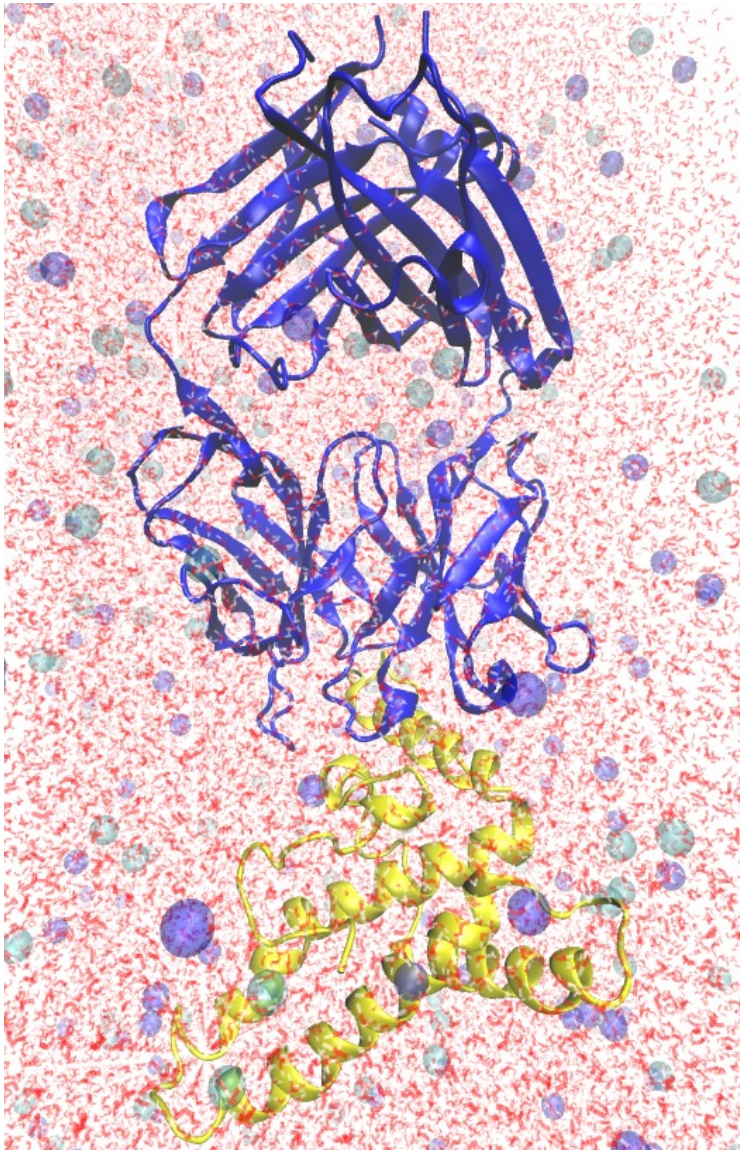
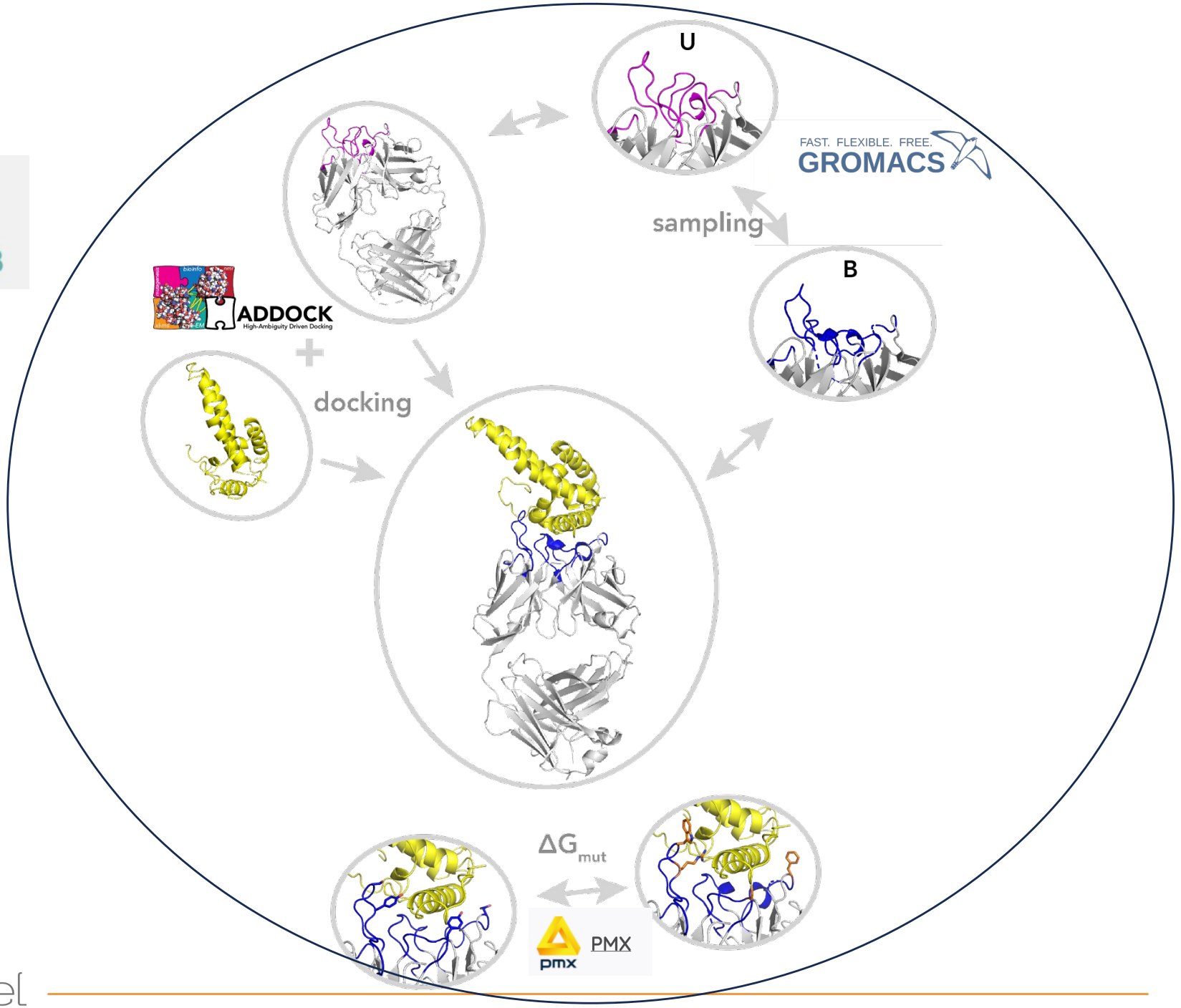
A bioexcel use-case

Centre of Excellence for Computational Biomolecular Research

How an antigen
interacts with antibody




BIOBB
U
FAST. FLEXIBLE. FREE.
GROMACS
sampling
B
ADDOCK
High-Ambiguity Driven Docking
+
docking
ΔG mut
PMX
pmx

Basis of molecular dynamics simulations

Alessandra Villa

PDC-Center for High Performance Computing,

KTH-Royal Institute of Technology,

Stockholm, Sweden

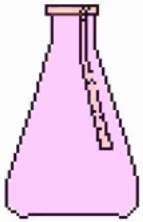
avilla@kth.se

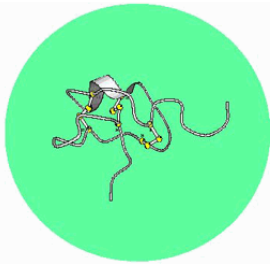
Goal of a Molecular Simulation

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

One method is Molecular Dynamics

Properties





Experimental measurements are made on macroscopic samples. An experiment measures an average property
averaged over a large numbers of molecules (macroscopic sample), usually averaged over the time of measurement.

Molecular Simulations enable us to predict system properties through the generation of several conformations of the systems in such a way that accurate values of a property can be obtained.

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

Sampling and properties

One goal of a molecular simulation is to generate enough representative conformations such the sampling of the conformational ensemble is ergodic. If this is the case, experimentally relevant information may then be calculated

One of the methods is Molecular Dynamics

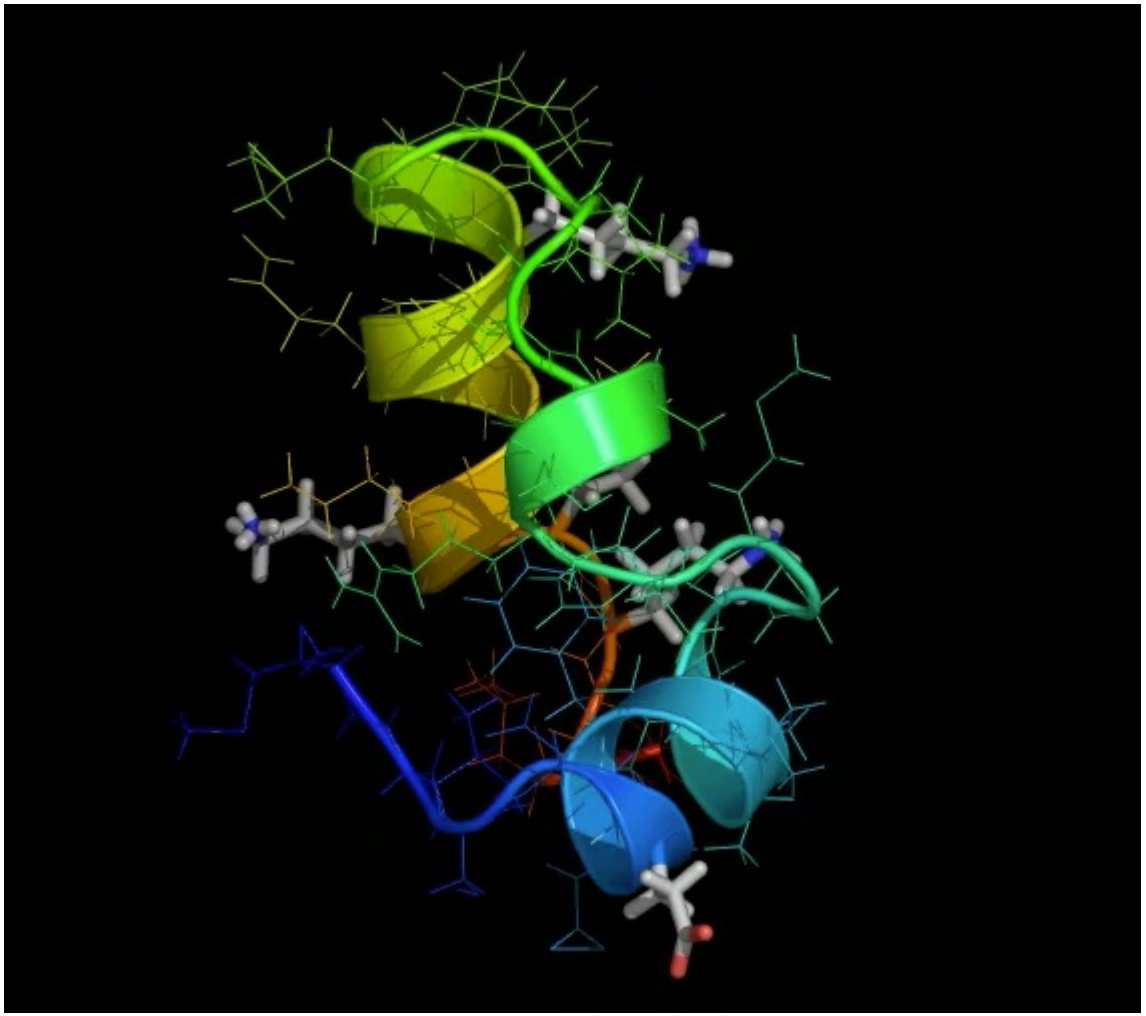
Different properties have different relaxation time

Molecular dynamics simulations enable also time-dependent behavior of the systems

Challenge for biomolecular simulations

- Biophysical processes involve hundreds of thousands of atoms often in intricate interactions that are difficult to simply
- They span a wide range of time scales: primary events (e.i. photosynthesis) occur within ps, enzymatic and regulatory processes take milliseconds, and structural reorganizations may exceed seconds.
- The small driving forces that cause molecular changes results from large, opposing energetic effect. This required careful fine-tuning of the force fields that describe interatomic interactions.

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

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

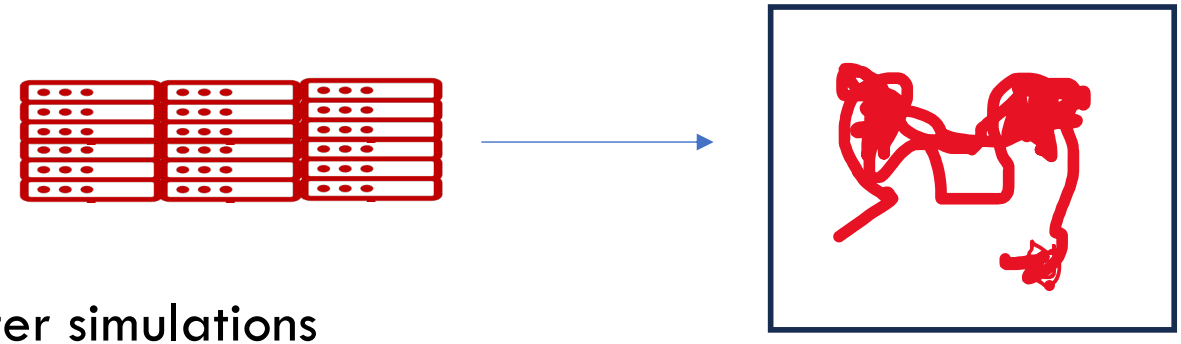
$$i = 1 \ldots\ldots\ldots \mathrm{N}_{\mathrm{particles}}$$

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

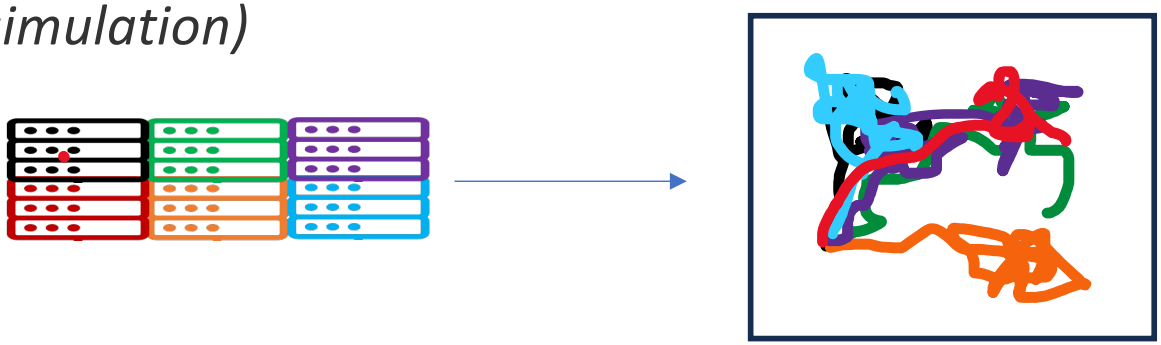
Two different ways

to sample the conformational space.

- One long simulation



- Many shorter simulations
- (*ensemble simulation*)



Molecular Model

One of the factor that governs a MD simulation

Molecular model

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

- Why?

Have a look at the Nobel lectures in chemistry 2013

(Martin Karplus, Michael Levitt, Arieh Warshel)

Choose the appropriate molecular model

- Degree of freedom (which particles)
- Energy functions or interaction potentials that describe the interactions between the particles

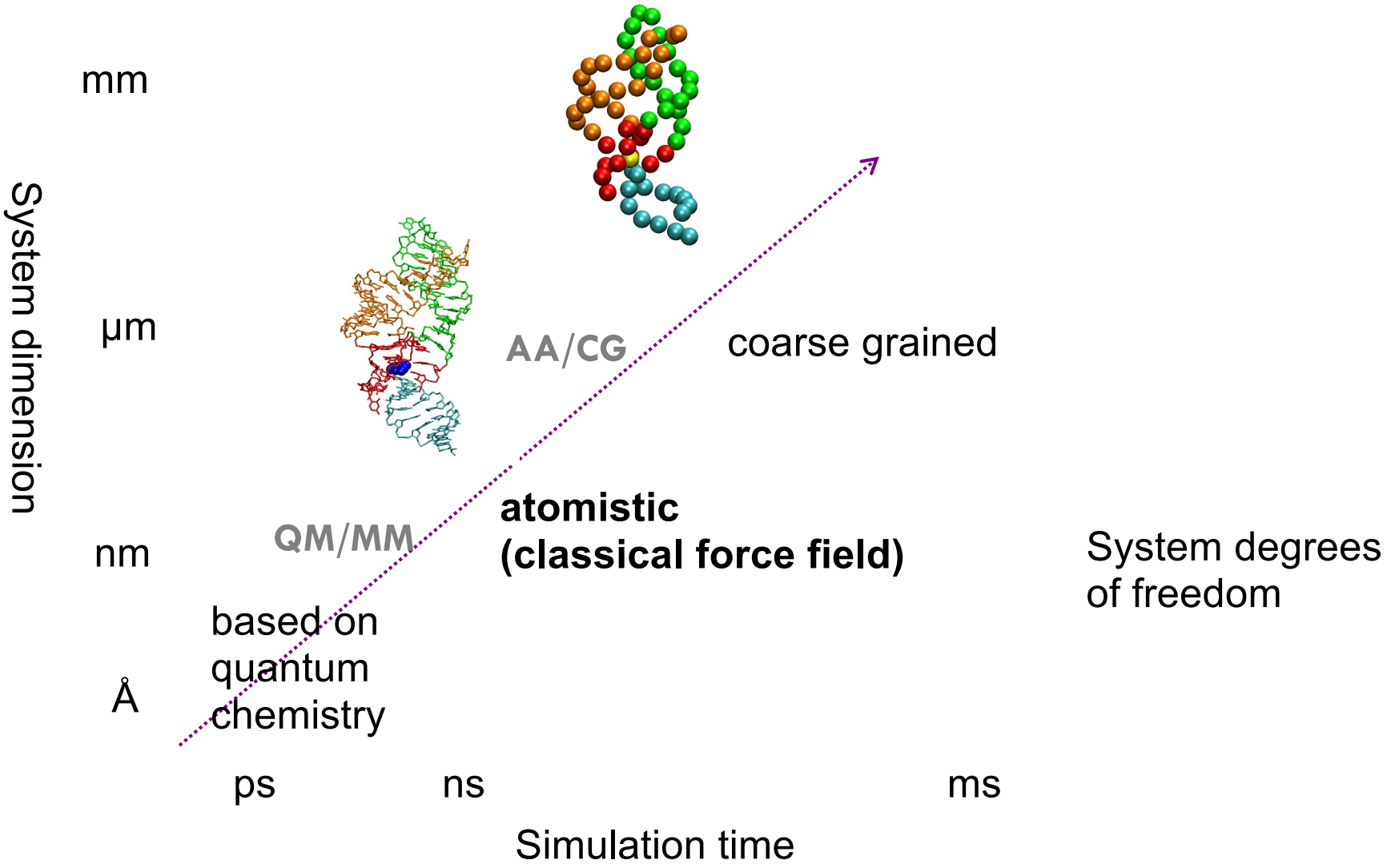
Reliability:

- Model must encompass the properties of interest

- Simulation time >> time scale of the process to be investigated

- Simulation size >> size of the simulated system

Molecular model and system dimension



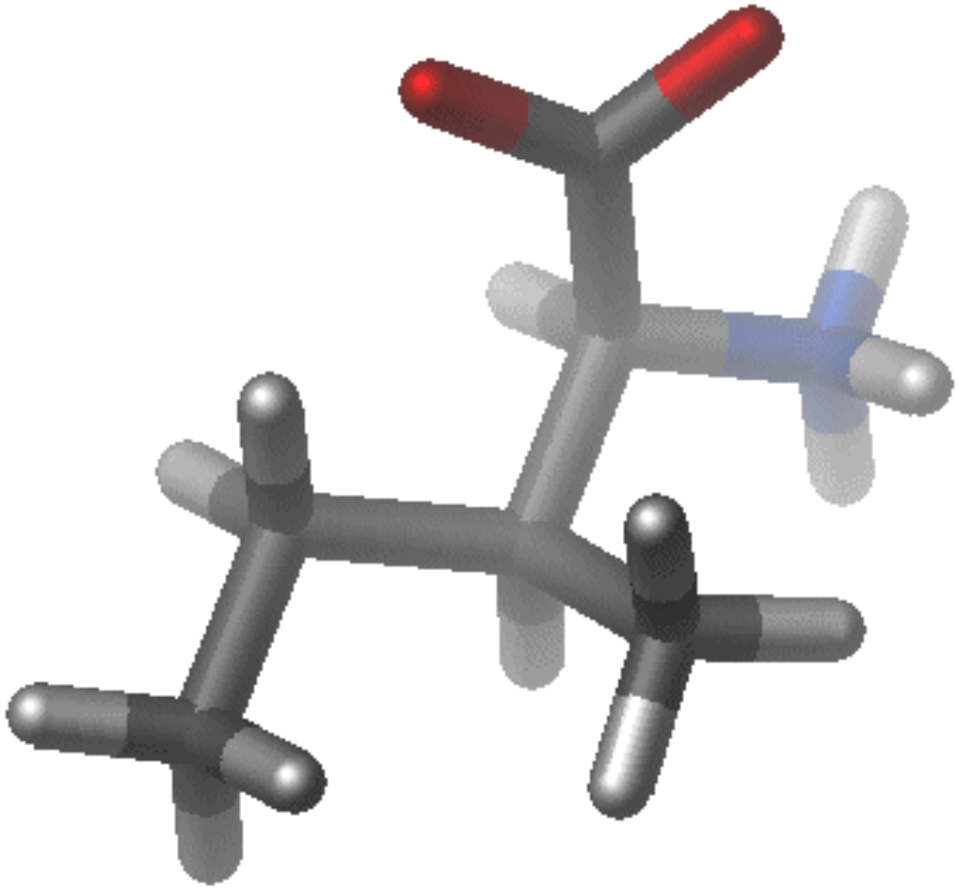
Pair approximation – effective potential

The pair approximation gives remarkably good description if the average of three-body effects can be partially included by defining an 'effective' pair potential

$$V(r) = V(r_1, r_2, \cdots, r_N)$$
$$= \sum_i V^{(1)}(r_1) + \sum_i Veff^{(2)}(r_1, r_2)$$

Consequence of this approximation: effective potential depends on density and temperature, while the true two-body potential does not.

Molecular “movements”



Force field and analytical functions

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

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

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

bonded-interactions

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

non-bonded interactions

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

Force fields and parameters

- Usually based on atom types, often many for each element
- Parameters intimately co-dependent on functional forms and each other
- Do not expect to mix and match force field parameters
- Addition of new parameters in line with parameterization strategy

- Most common biomolecular force field families: AMBER, CHARMM, GROMOS, OPLS, Martini..
- Generalized force field
- Open force field initiative (https://openforcefield.org/)

- AI-based force field
- Polarizable force field

- Online servers / offline tools where possible (SwissParam, ATB, Antechamber, LEaP, acpype, PRODRG, STaGE, MKTOP, CHARMM-GUI)

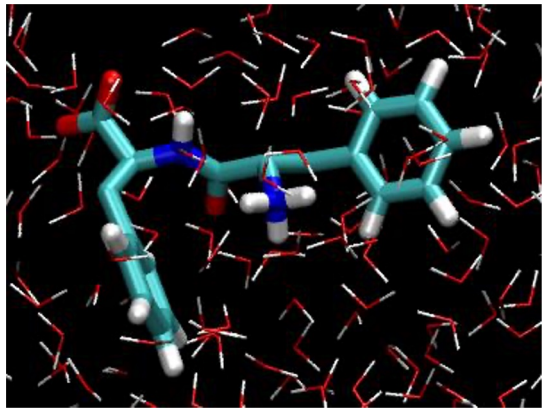
Parameter dependencies

Type of interaction	
Bonded	Bonds weakly dependent geometry
	Angles weakly dependent on bonds
	Dihedral Angles **strongly dependent on angles, vdw, charges, ...**
Non-bonded	vdw parameters highly correlated with charge distribution
	Partial charges Dependent on dielectric, vdw's, treatment of long-range, all-atom/united atom, ...

Force field parameters

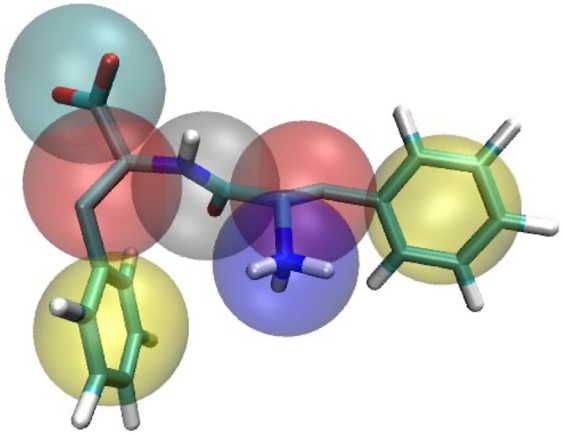
- From experimental and *ab initio* studies of small model compounds
- Crystal data (bond lengths, angles)
- Spectroscopy (eg, IR, Raman) can give force constants for bonds, angles
- Charges usually by comparison with QM electrostatic potential;
- Values corresponding to observed data (e.i thermodynamic/kinetic properties) are computed with trial energy function and parameters are varied until a reasonable fit is obtained
- One hopes then that the parameters can be transferred to the large molecules of interest

Atomistic vs Coarse grained model



Degree of freedom reduction

Back-mapping to atom



ATOMISTIC
40 particles + water molecules

COARSE GRAINED
7 particles

time mapping?

Villa et al. PCCP 2009

What governs a MD simulation

- choice of degrees of freedom
- force field parameters
- **treatment of non-bonded interactions**
- **egration time step**
- **boundary conditions**
- **treatment of temperature and pressure**
- **environment (solvation effects, ions and etc)**
- **starting configuration**

Non-bonded long-range 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**

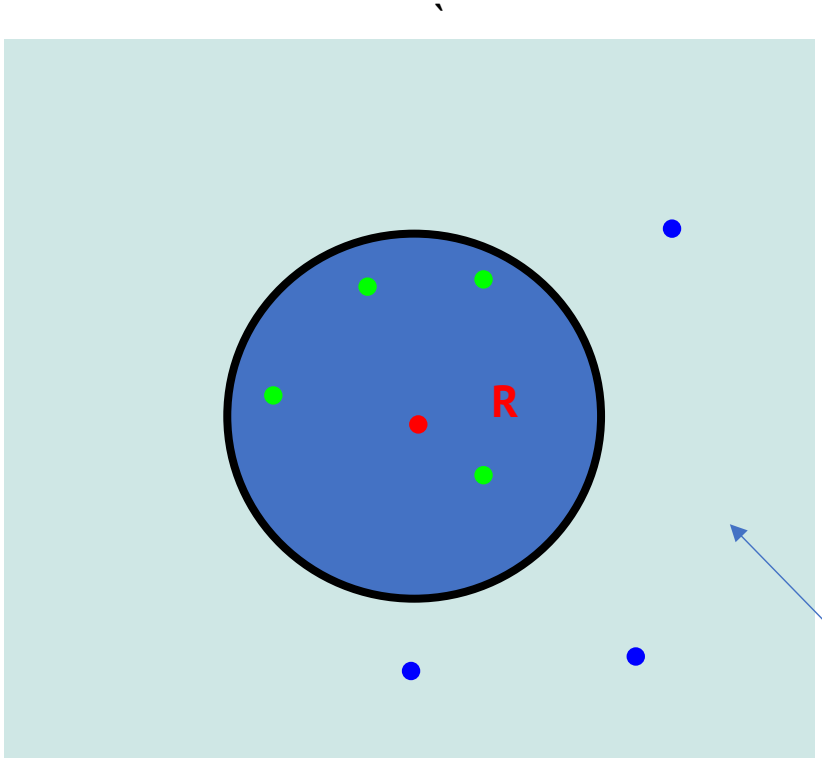
–Decay however relatively fast with the distance:

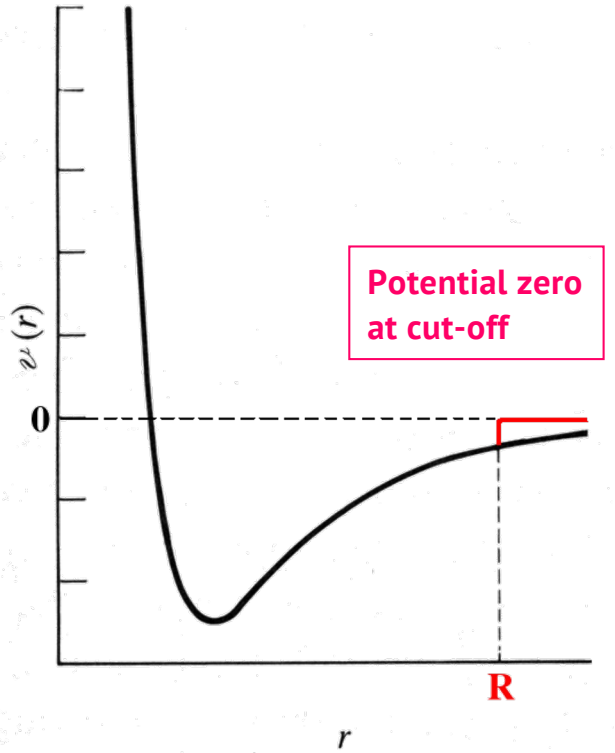
•Lennard-Jones ~ $1/r_{ij}^6$

•Coulomb ~ $1/r_{ij}$

Instead of computing the interactions with all atoms in the system only consider those atoms that are closer than a cut-off distance R_c and reduce thereby the computing time

Treatment of long-range interactions





Particle Mesh Ewald

Reaction field

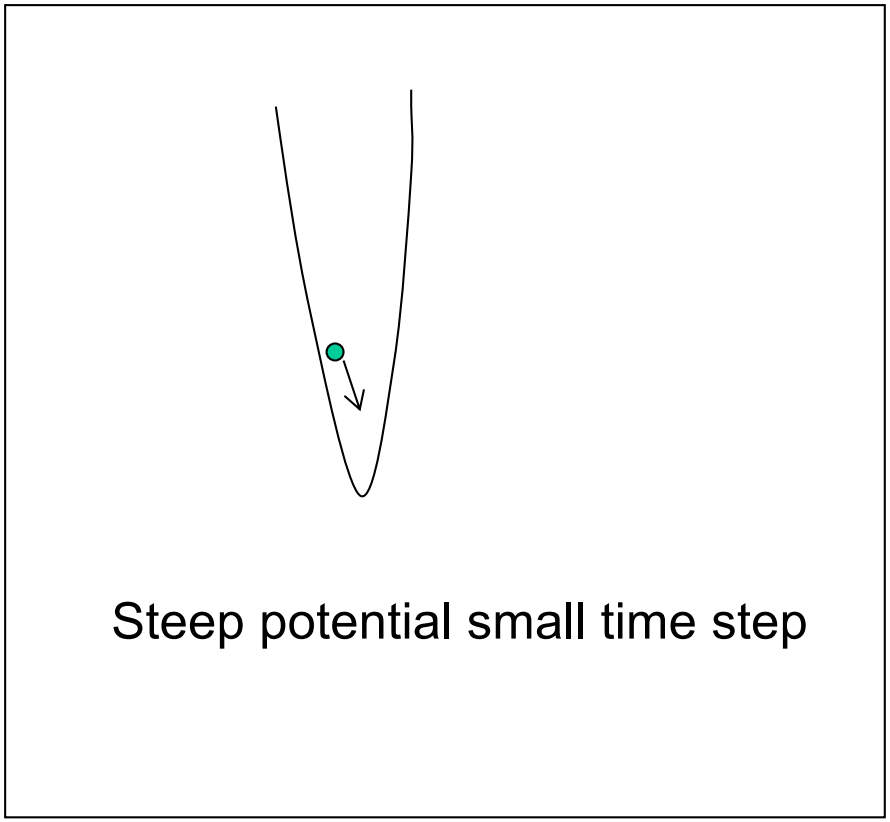
Integration time step

The time step
determines how much time can be simulated.

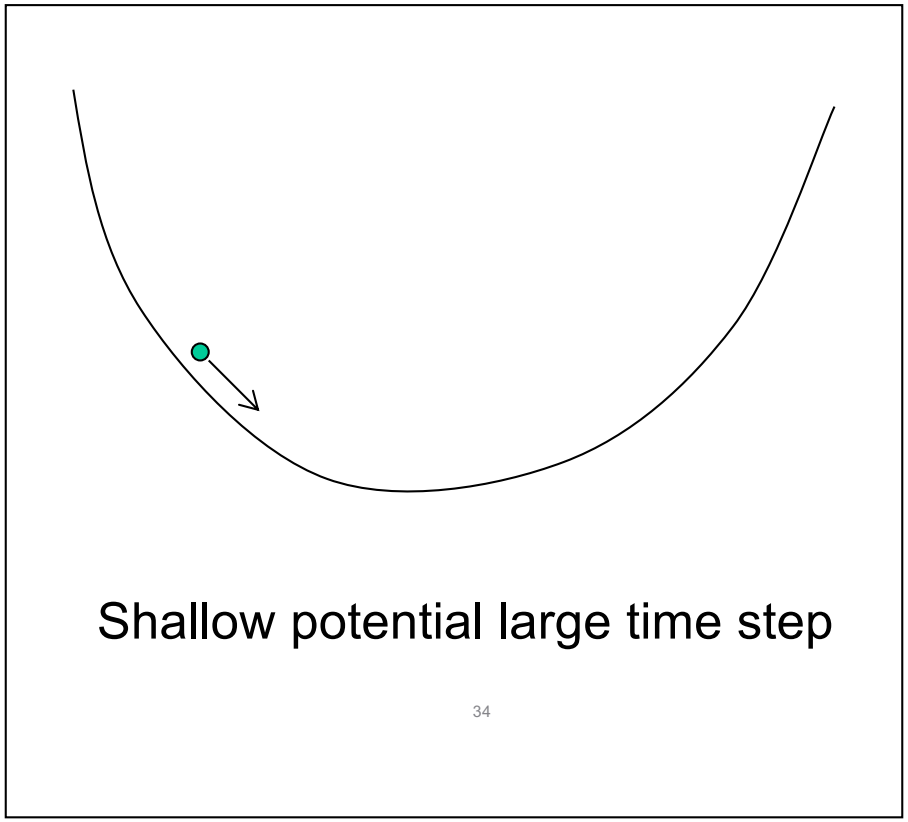
The smaller the time step
 the more expensive the calculation

Choosing the time step

The maximum time step determined by the curvature of the potential



Steep potential small time step



Shallow potential large time step

34

Useful tricks

Δt limited by fast motions - 1fs

=> Remove bond vibrations => 2 fs

- Constraint algorithms (LINCS, P-LINCS, SHAKE)

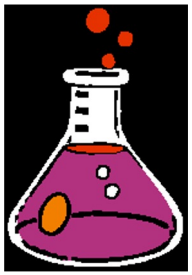
- Multi-time step

  => different forces are evaluated at different time steps

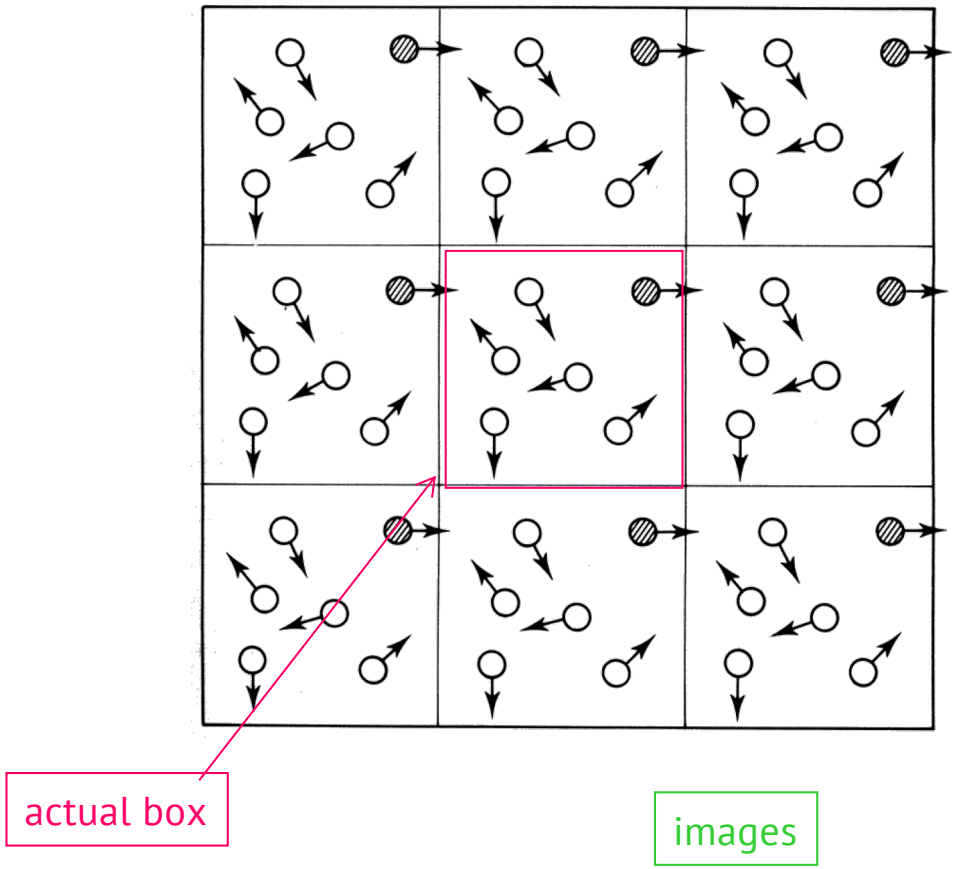
- Mass-repartition

  => typically hydrogen masses are scaled (h-bonds constraints a factor of 3 will enable a time step of 4fs)

Boundary Conditions



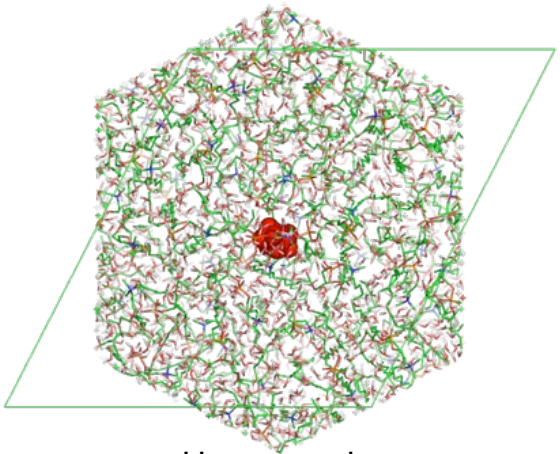
- Simulation box is replicated in all directions
- Particles leaving the central box are reintroduced on the opposite side
- Interactions calculated with particles in neighboring cells
- No vacuum interface anymore!



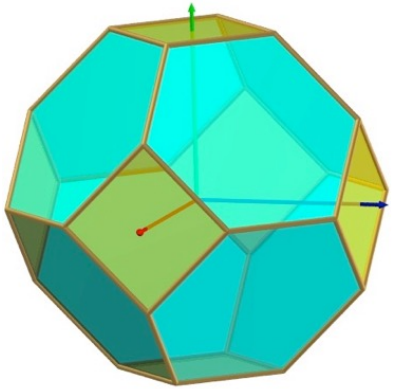
The actual box is surrounded by **26** images of its-self

Common periodic cell shapes

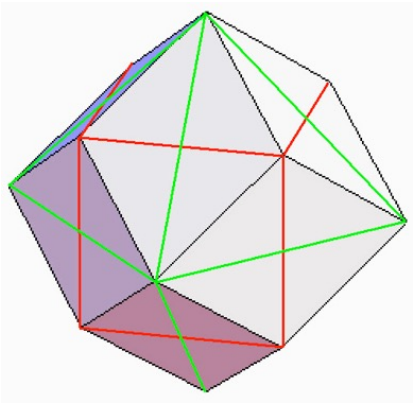
- Cubic / rectangular
- Hexagonal for membrane simulations (volume 87% of cube)
- Truncated octahedron (approximate sphere, volume 77% of a cube)
- Rhombic dodecahedron (most spherical cell, 71%)



Hexagonal



Truncated octahedron



Rhombic dodecahedron

Thermodynamic ensembles

- ***NVE* ensemble**

number of particles *N*, volume *V* and total energy of the system are constant

- ***NVT* ensemble (canonical ensemble)**

number of particles *N*, volume *V* and temperature T of the system are constant

- ***NPT* ensemble**

number of particles *N*, pressure p and temperature T of the system are constant

Temperature in MD simulation

Temperature *T* related to the kinetic energy *K* of the system and therefore to the velocities

$$E_{kin} = \sum_{i=1}^{N} m_i \frac{v_i^2}{2} = \frac{N_{df} k_B T}{2} = \frac{(3N - Nc\ - 3) k_B T}{2}$$

N_{df} is the number of degrees of freedom in the system, N is the number of particles and N_c is the number of constraints on the system

The temperature can thus be controlled by modifying the velocities of particles in the system

Thermostat

The role of a thermostat is to ensure that we have

- the correct average temperature
- fluctuations of the correct size.

Examples of thermostat

Berendsen weak-coupling (Berendsen,1981) -> efficient for relaxing a system to the target temperature but do not correctly reproduce the fluctuations of the kinetic energy.

Velocity-rescaling temperature coupling (Bussi, 2007) -> a Berendsen thermostat with an additional stochastic term that ensures a correct kinetic energy distribution

Nose'-Hoover temperature coupling (Nose', 1984; Hoover, 1985) -> a thermal reservoir and a friction term in the equations of motion is added to system Hamiltonian .

Pressure in MD simulations

The pressure is related to the volume of the system and the interaction between particles

$$P = N_{df} k_b T / V - \frac{1}{3 \mathrm{k}_b V T} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \vec{r}_{ij} \vec{f}_{ij}$$

The second term is the virial describing the contribution due to the forces between the particles and it can easily be calculated in MD when calculating the forces.

Pressure can be controlled by varying the volume of the system and scaling the position of the molecules

Barostat

Pressure coupling can be isotropic, semi-isotropic or anisotropic.

Surface-tension coupling is also possible.

Examples of barostat

Berendsen (Berendsen,1984) -> scales coordinates and box vectors every step

Stochastic cell rescaling (Bernetti, 2020) -> add a stochastic term to Berendsen algorithm

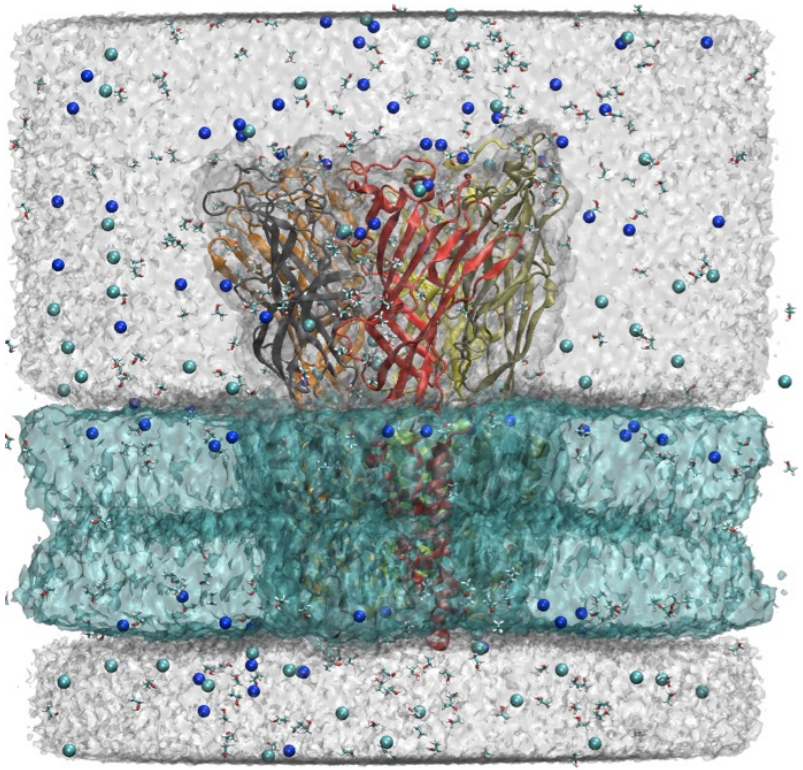
Parrinello-Rahman (Parrinello, 1981; Nose', 1984) -> similar to the Nosé-Hoover temperature coupling

What governs a MD simulation

- choice of degrees of freedom
- force field parameters
- treatment of non-bonded interactions
- integration time step
- boundary conditions
- treatment of temperature and pressure
- **environment (solvation effects, ions and etc)**
- **starting configuration**

System setting-up

Environment – Solvent - Ions

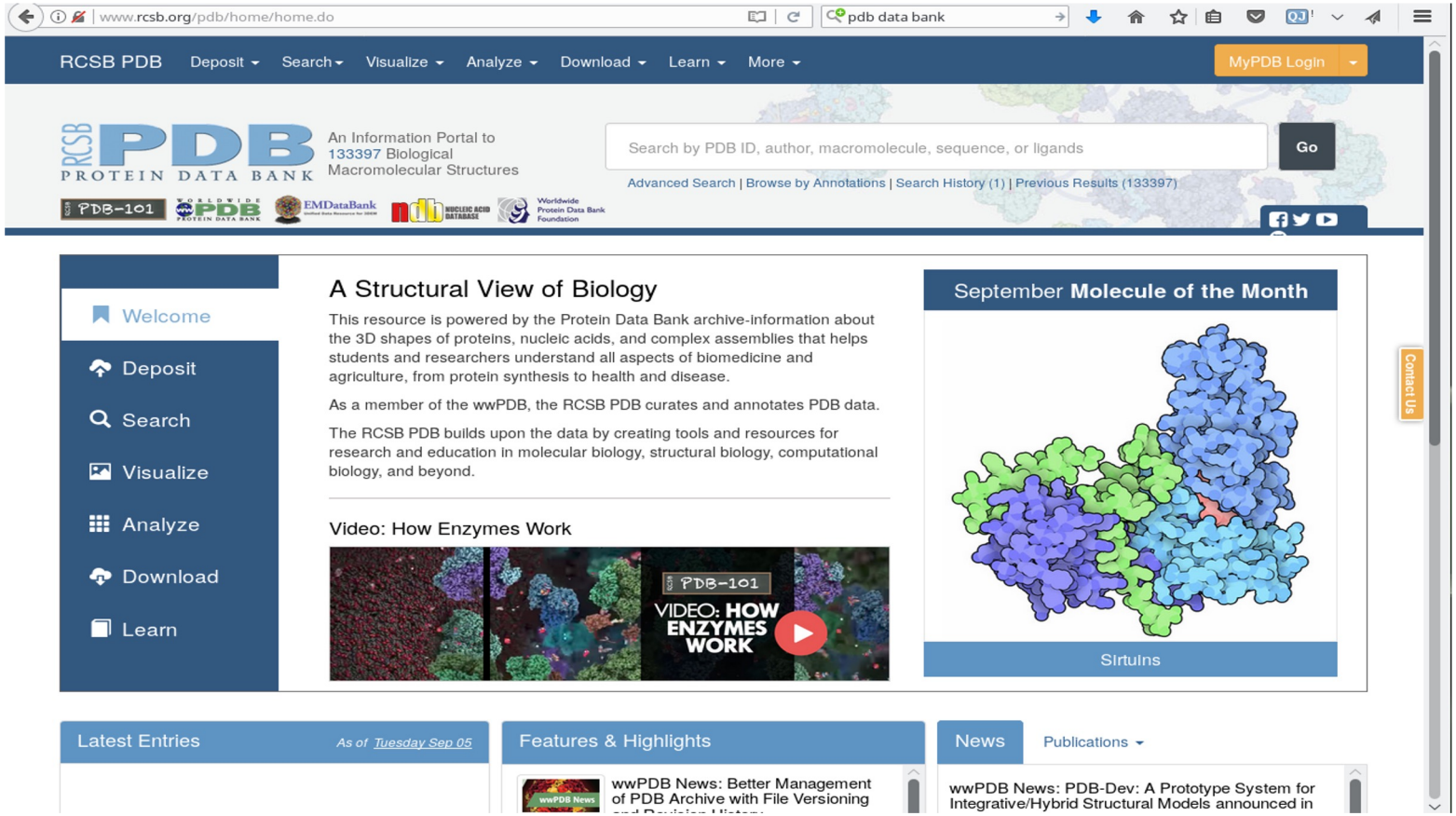


Starting configuration

Initial coordinates

- Experimental structures from database (coming from X-ray crystallography, NMR spectroscopy, cryo-electron microscopy)
- Self-built model three-dimensional structure (e.i homology modelling, docking model, Alpha fold)
- Pre-built solvent box

3-D structures Database



http://www.rcsb.org/

Possible issues

- Not all the atoms are available in the experimental structure (i.e. flexible loops)

- H positions:

- not always experimentally available
- pKa shift due to pocket environment
- account for possible tautomeric states

- Water positions (kinetically trapped water). Water or ion ?

- Extra molecules (i.e. co-factors, ligands, surfactant, etc) or special conditions used to promote the observation of the experimental structure

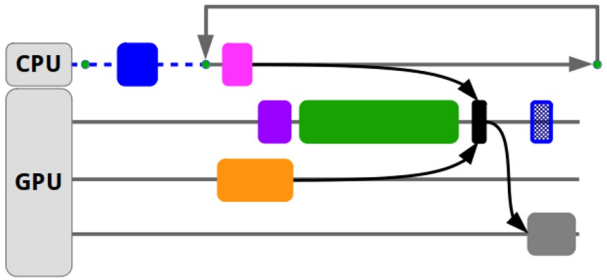
Molecular Dynamics Simulations and

www.gromacs.org

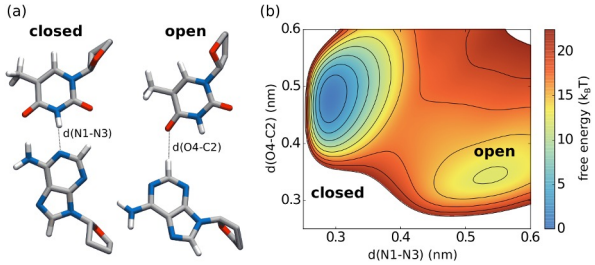
What is the name of the bird in GROMACS logo?

Highlights

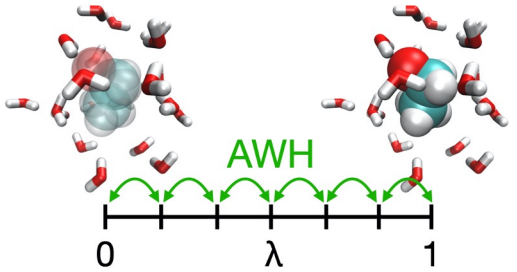
Heterogeneous parallelization and GPU acceleration



Accelerating sampling with the AWH method



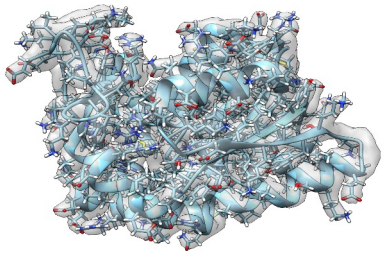
Applying the Accelerated Weight Histogram method to alchemical transformations



Interfaces
-hybrid QM/MM with CP2K
- COLVARS -Collective variable simulations with Colvars

gmxapi

Density-fit simulations



Ions in Action! Studying ion channels by Computational Electrophysiology

https://www.gromacs.org/highlights.html

Not only molecular dynamics simulation

GROMACS

- Simulated annealing
- Walls
- COM pulling
- AWH adaptive biasing
- Enforced rotation
- Free energy calculations
- Expanded ensemble calculations
- Non-equilibrium MD
- Electric fields
- Computational Electrophysiology
- Density-guided simulations
- QM/MM simulations with CP2K interface
- Collective variable simulations with Colvars

Work on progress

- pH simulations
- Implementation of AI force field

https://manual.gromacs.org/current/user-guide/mdp-options.html

Restraints and simulations

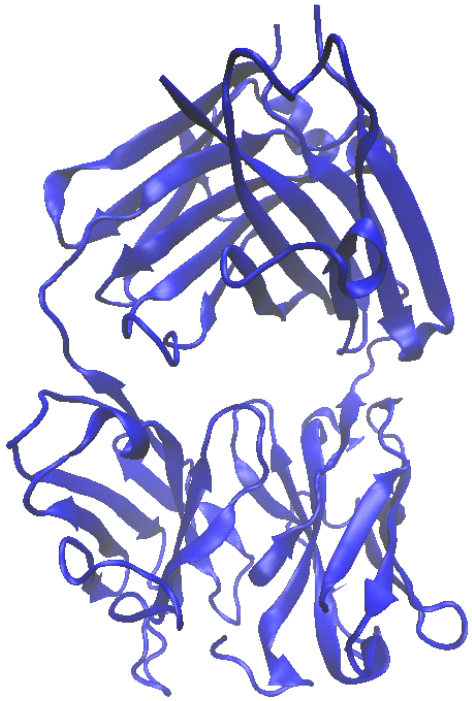
Here special potentials are used for imposing restraints on the motion of the system

Type of restraints :

- **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

The tool gmx genrestr may be useful here

Visualization and Analysis



Most used GROMACS analysis tools:

gmx energy – to process energy

gmx trjconv – to convert trj files

gmx select / gmx make_ndx – for selection

gmx rms

gmx cluster

gmx distance/gmx mindist /gmx pairdist

for trajectory analysis

Example of visualization software:

VMD http://www.ks.uiuc.edu/Research/vmd/

Rasmol http://www.openrasmol.org/

Pymol http://www.pymol.org/

Chimera https://www.cgl.ucsf.edu/chimera/

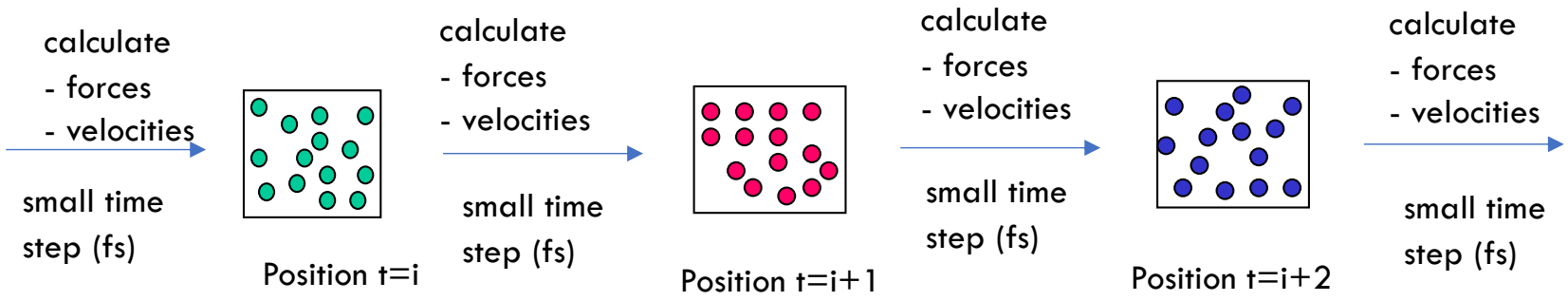
All GROMACS commands are command-line and require an option before any arguments.

Use option –h to show help

Why good performance

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.

How? by iteratively solving equations of motion



Some interactions take more time than other

High computational cost

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)

How to get good performance

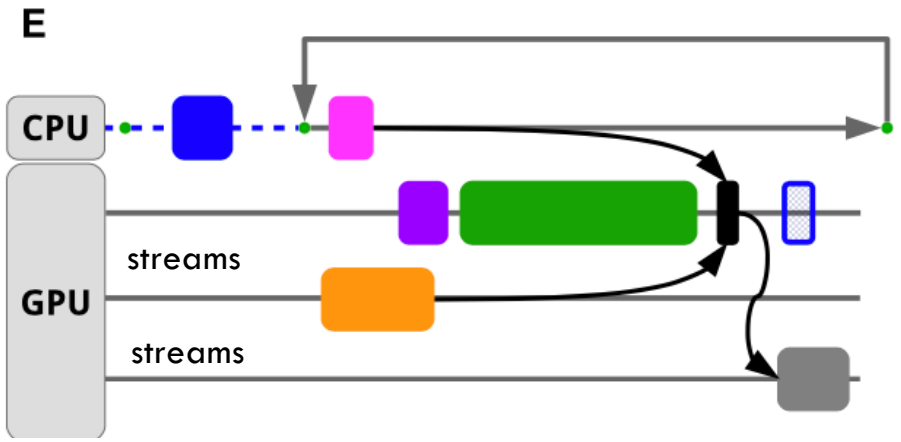
- Optimal mdp parametes
  - Currently most mdp parameters do not affect performance much (except PME order and grid).
  - Automated PME tuning optimises the Coulomb cut-off and PME grid size (in GROMACS)

- 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

What on CPU and GPU?

CPU is used for scheduling work, transferring data, and launching computation on the accelerator, as well as inter- and intra-node communication.
Accelerator tasks are launched asynchronously using APIs to allow concurrent CPU–GPU execution



Non bonded forces
PME forces
Bonded forces

dynamic list pruning
integration, constrains

J. Chem. Phys.. 2020;153(13). doi:10.1063/5.0018516

Thermostats & barostats

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

Good performance

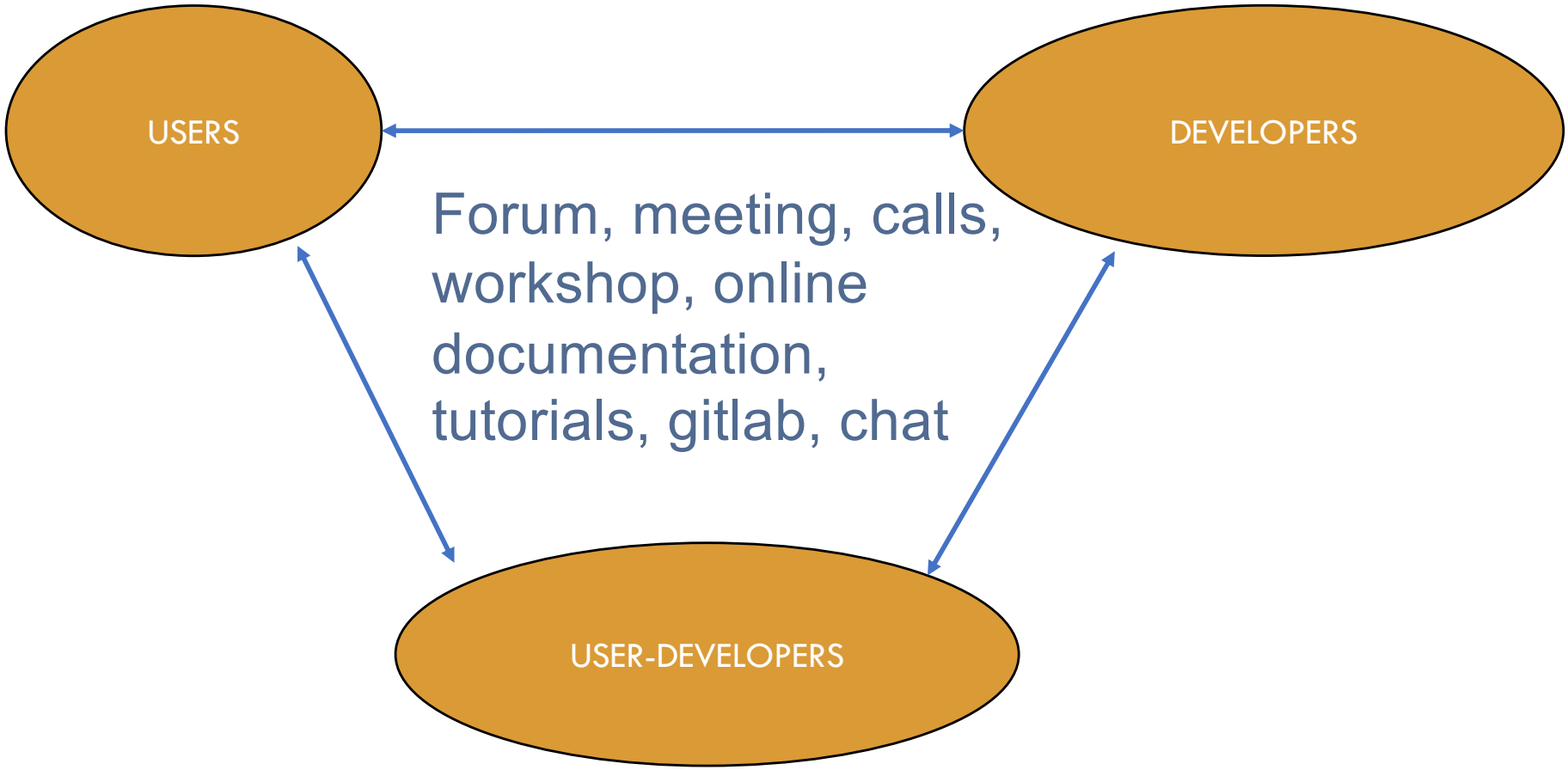
Important for everybody => but there is not recipe

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.

Material from a workshop how to run GROMACS efficiently on the LUMI supercomputer

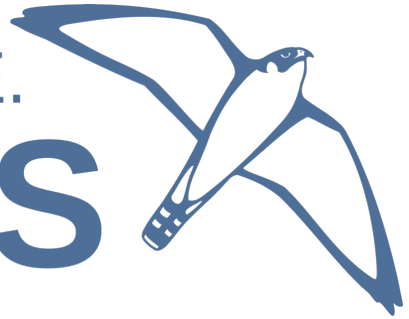
doi:10.5281/zenodo.10683366

FAST. FLEXIBLE. FREE.
GROMACS community



FAST. FLEXIBLE. FREE.

GROMACS



GROMACS versions are always defined by two numbers:
Year.Release => version 2024.1 2023.4

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

Interested in coding in GROMACS: register to the workshop
learn to code in GROMACS – online – 10-12 September 2024
www.gromacs.org/workshop

GROMACS 2024

Project leaders

- Berk Hess (KTH-Royal Institute of Technology, SE)
- Erik Lindahl (Stockholm University and KTH-Royal Institute of Technology, SE)

Current contributors

Mark Abraham, Andrey Alekseenko, Vladimir Basov, Cathrine Bergh, Eliane Briand, Ania Brown, Mahesh Doijade, Giacomo Fiorin, Stefan Fleischmann, Sergey Gorelov, Gilles Gouaillardet, Alan Gray, M. Eric Irrgang, Farzaneh Jalalypour, Joe Jordan, Carsten Kutzner, Justin A. Lemkul, Magnus Lundborg, Pascal Merz, Vedran Miletic, Dmitry Morozov, Julien Nabet, Szilárd Páll, Andrea Pasquadibisceglie, Michele Pellegrino, Hubert Santuz, Roland Schulz, Tatiana Shugaeva, Alexey Shvetsov, Philip Turner, Alessandra Villa, Sebastian Wingbermuehle.

THANK YOU !

survey on the core software is online soon

Help us and fill it !

Co-funded by
the European Union

EuroHPC
Joint Undertaking

Utrecht
University