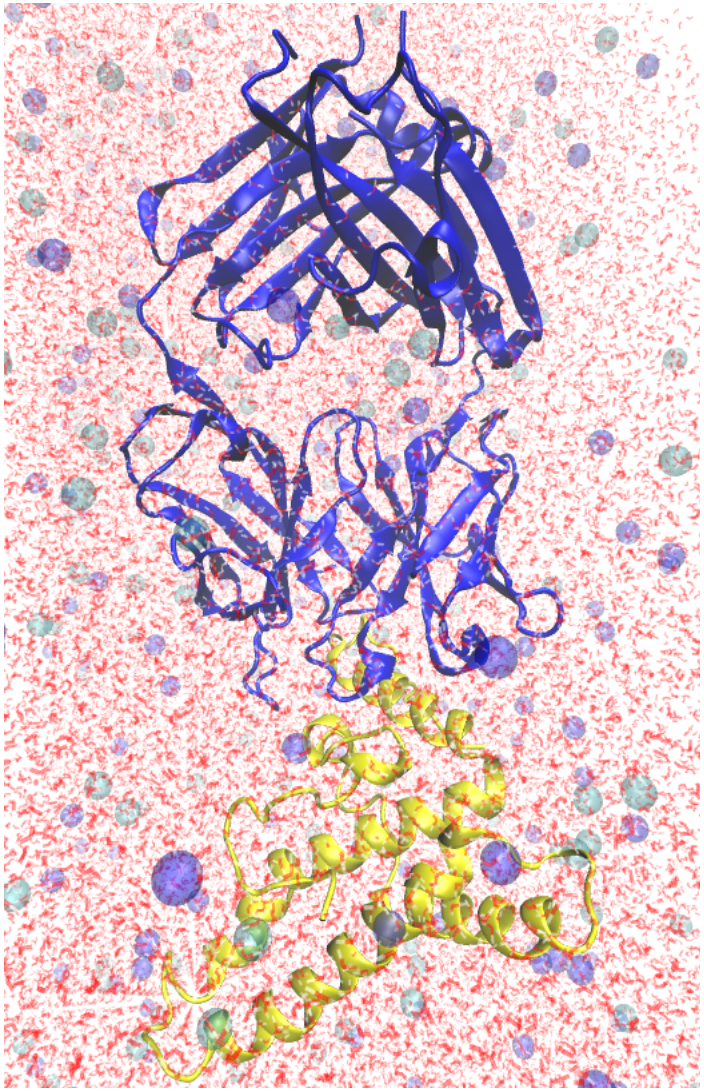
Introduction to MD simulations using GROMACS

Alessandra Villa

avilla@kth.se

How an antigen interacts with antibody



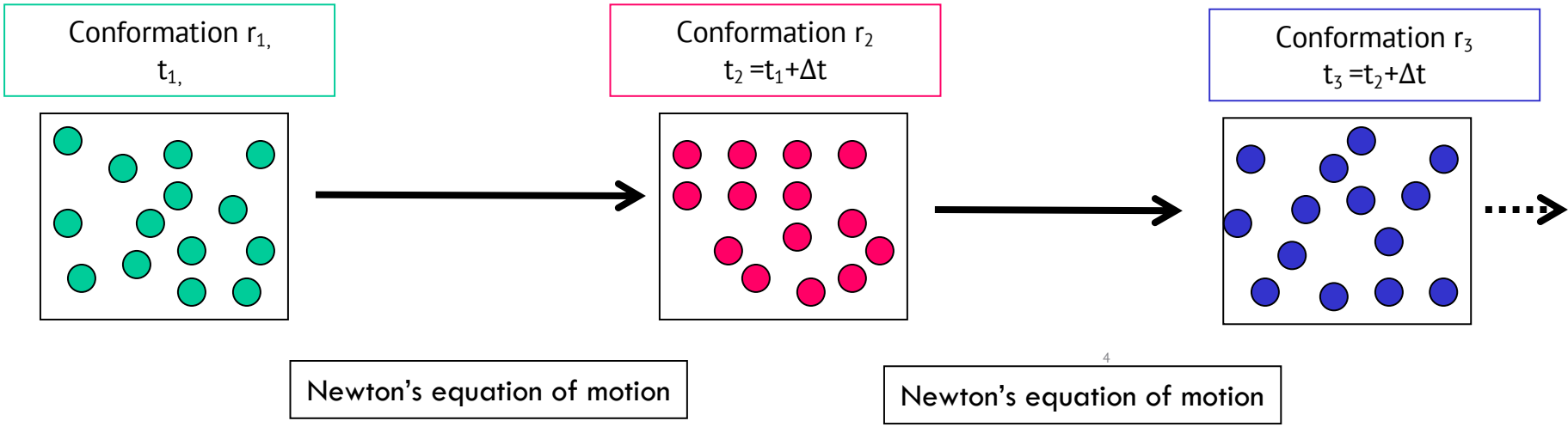
Molecular Simulation

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.

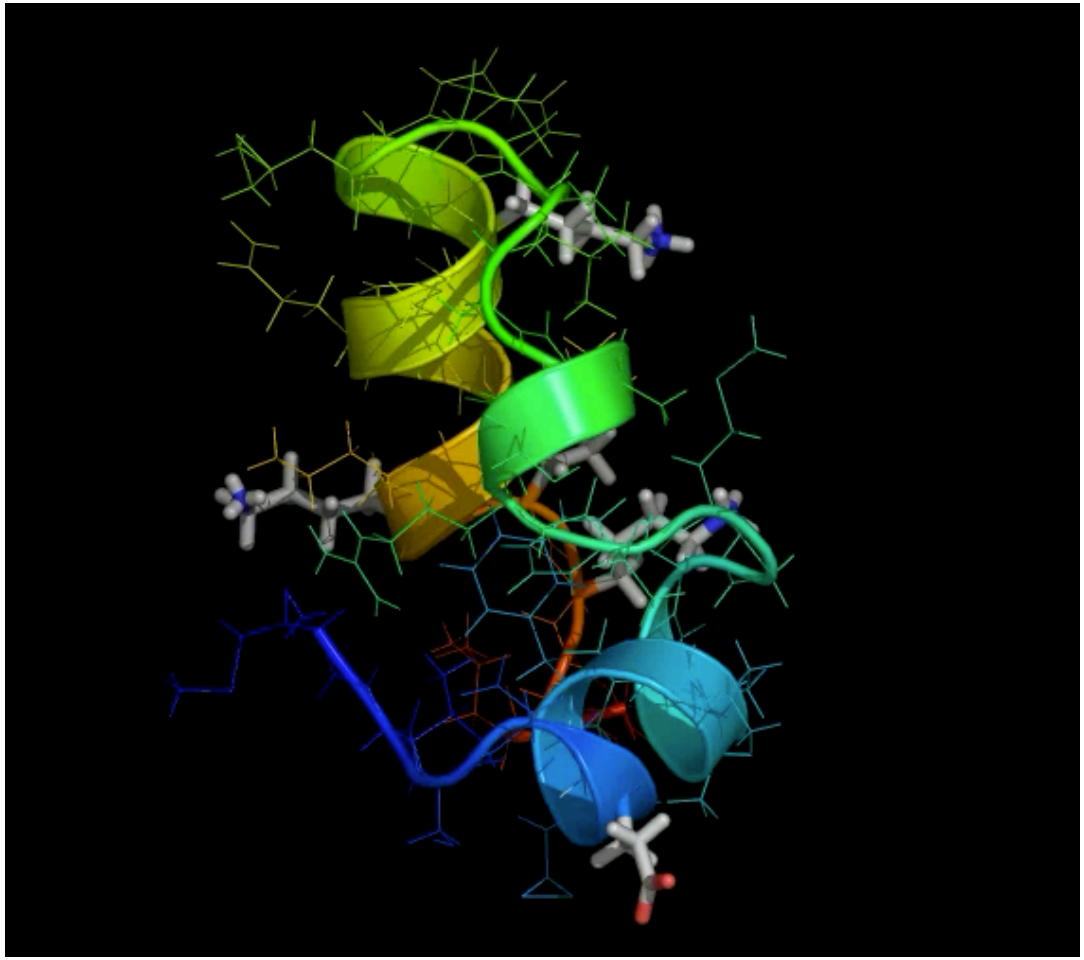
One method is Molecular Dynamics

Molecular Dynamics (MD) methods

generates configurations applying Nature's law of motion for atoms of a molecular system.



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

Challenge for biomolecular simulations

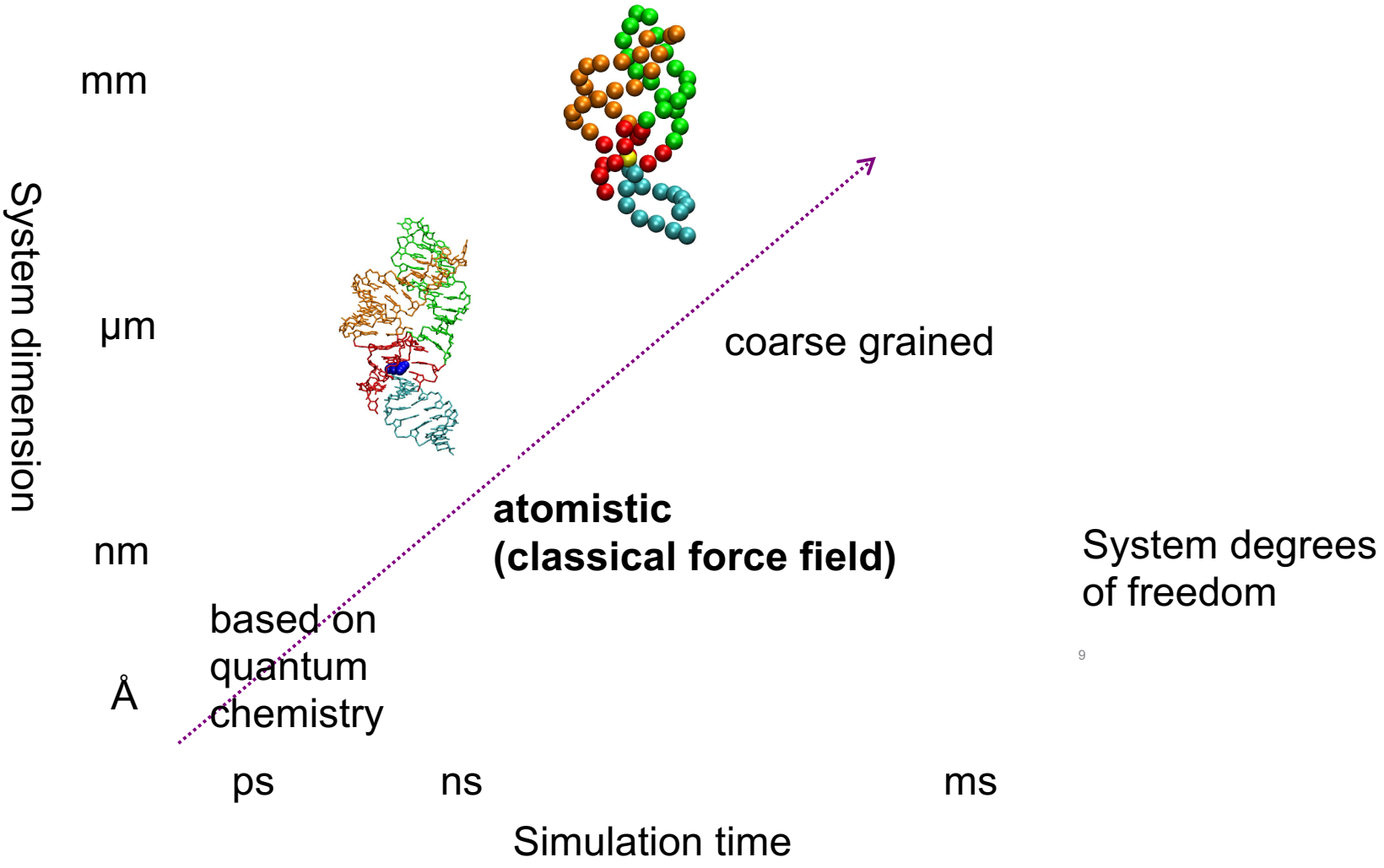
- Biophysical processes involve hundreds of thousands of atoms often in intricate interactions that are difficult to simply
- They span a wide range of time scales: primary events (e.i. photosynthesis) occur within ps, enzymatic and regulatory processes take milliseconds, and structural reorganizations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

What governs a MD simulation

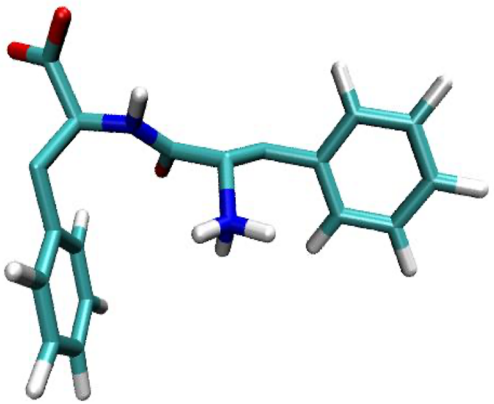
- choice of degrees of freedom
- Interaction potential (force field parameters)
- treatment of non-bonded interactions
- boundary conditions
- integration time step
- treatment of temperature and pressure
- starting configuration
- environment (solvation effects, ions and etc)

Degree of freedom

Molecular model and system dimension

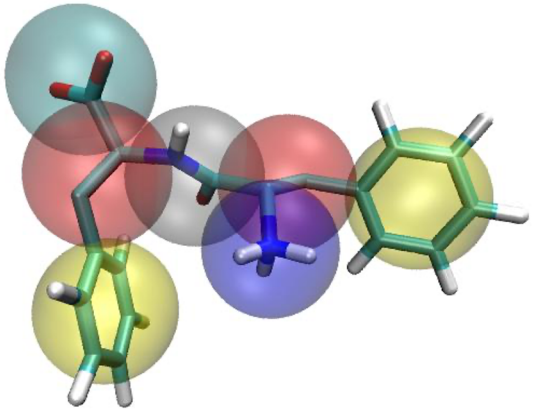


Degree of freedom



Degree of freedom reduction

Back-mapping to atom



ATOMISTIC
40 particles

COARSE GRAINED
7 particles

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

Interaction potentials

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)

Molecular mechanics force field

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

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

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

bonded-interactions

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

14

non-bonded interactions

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

Molecular mechanics force fields

- Usually based on atom types, often many for each element
- Parameters intimately co-dependent on functional forms and each other
- Most common biomolecular force field many families: AMBER, CHARMM, GROMOS, OPLS, Martini..
- Do not expect to mix and match force field parameters
- Addition of new parameters in line with parameterization strategy
- Online servers / offline tools where possible (SwissParam, ATB, Antechamber, LEaP, acpype, PRODRG, STaGE, MKTOP, CHARMM-GUI)

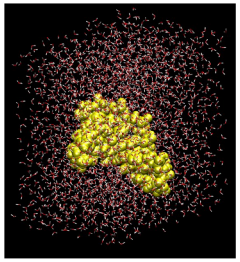
Force field parameters

- 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

Molecular simulation and GROMACS

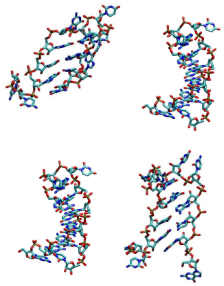
Input files

Structure
(*.gro , *pdb, etc)



Simulation parameter
file (*.mdp)

topology file (*.top)

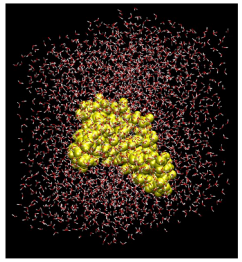


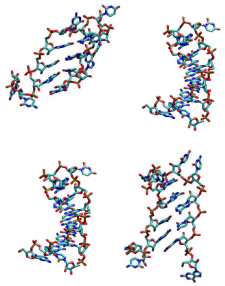
Molecular simulation and GROMACS

Input files

Structure
(*.gro , *pdb, etc)

Simulation parameter file
(*.mdp)





topology file (*.top)
Information on molecular model to describe the molecular system.
-> gmx pdb2gmx

Treatment of long-range interactions

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**

–Decay however relatively fast with the distance:

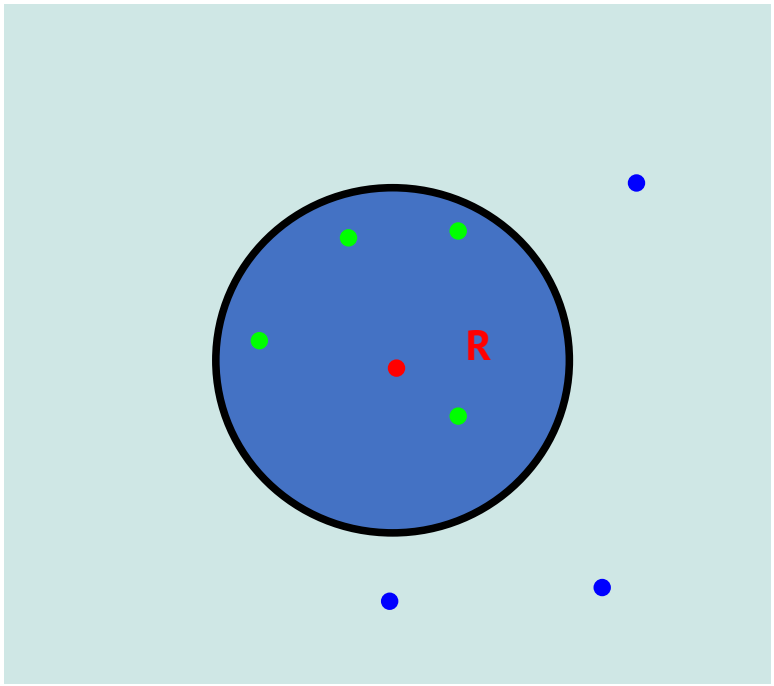
- Lennard-Jones ~ $1/r_{ij}^6$
- Coulomb ~ $1/r_{ij}$

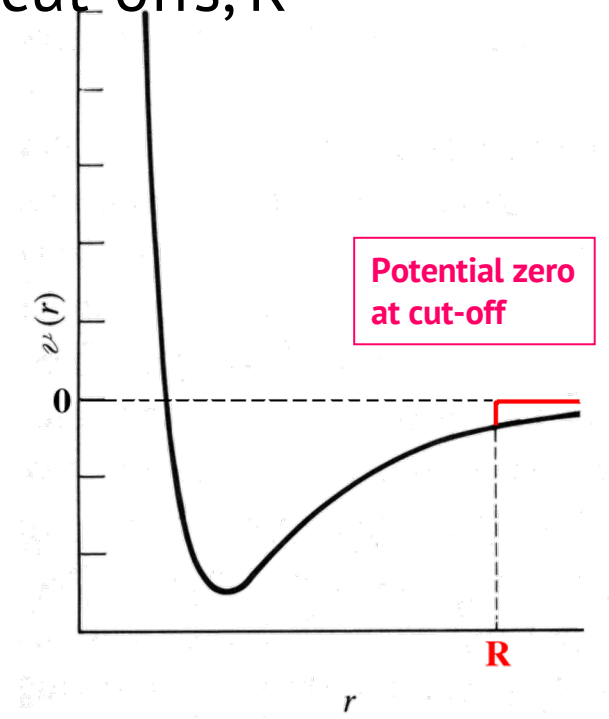
Instead of computing the interactions with all atoms in the system only consider those atoms that are closer than a cut-off distance R_c and reduce thereby the computing time

Non-bonded interactions

One solution:

to truncate the potential at given cut-offs, R





Truncating the potential

Problems:

- Charge-charge interactions decay slowly (~$1/r_{ij}$) than van der Waals interactions and long-range contributions are neglected.
- Non-bonded potential and forces are discontinuous at the cut-off

Possible solutions:

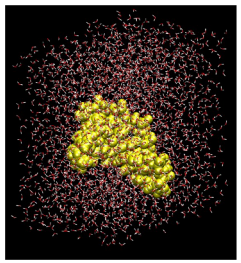
Particle Mesh Ewald

Reaction field

Molecular simulation and GROMACS

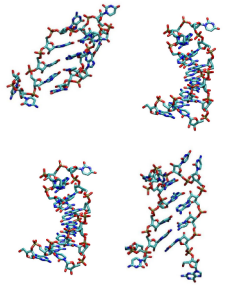
Input files

Structure
(*.gro , *pdb)



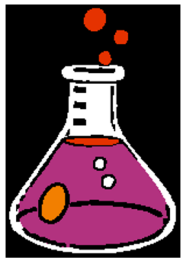
simulation parameter file
(*.mdp)
information on the treatment
of non-bonded interactions

topology file (*.top)
point charges and vdw
parametes

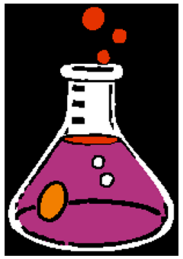


Q & A Break

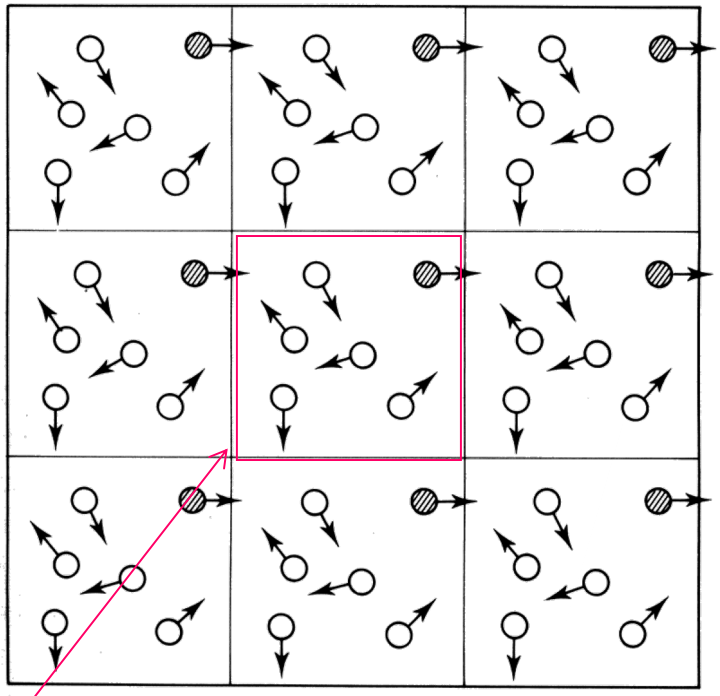
Boundary Conditions



Boundary Conditions



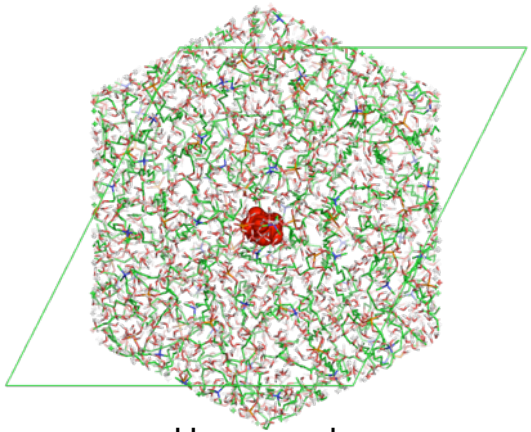
- Simulation box is replicated in all directions
- Particles leaving the central box are reintroduced on the opposite side
- Interactions calculated with particles in neighboring cells
- No vacuum interface anymore!



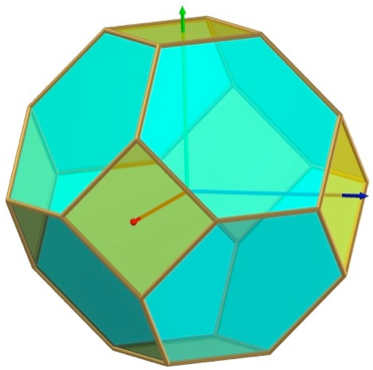
The actual box is surrounded by **26** images of its-self

Common periodic cell shapes

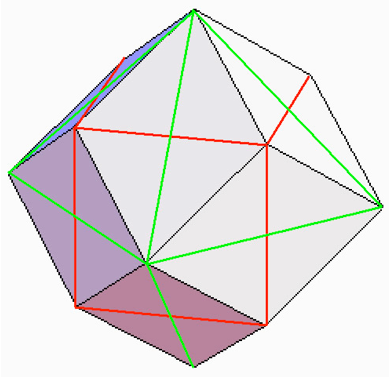
- Cubic / rectangular
- Hexagonal for membrane simulations (volume 87% of cube)
- Truncated octahedron (approximate sphere, volume 77% of a cube)
- Rhombic dodecahedron (most spherical cell, 71%)



Hexagonal



Truncated octahedron



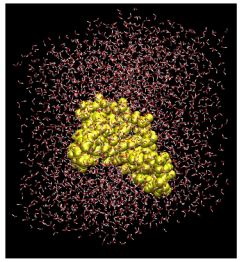
Rhombic dodecahedron

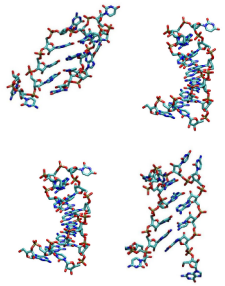
Molecular simulation and GROMACS

Input files

Structure
(*.gro , *pdb)
Box info
-> gmx editconf

simulation parameter file
(*.mdp)
periodic boundary condition





topology file (*.top)

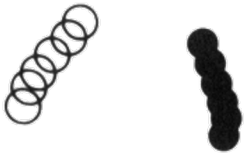
Integration time step

The time step
determines how much time can be simulated.

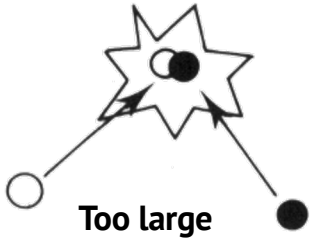
The smaller the time step
the more expensive the calculation

Choosing the time step

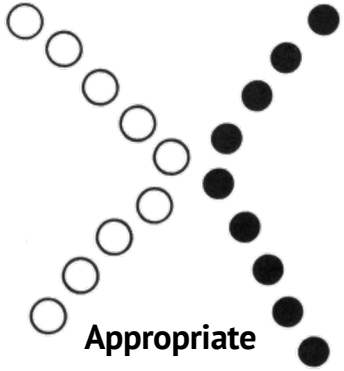
•What is an appropriate time step for the integration?



Too small
nothing happens



Too large
instabilities

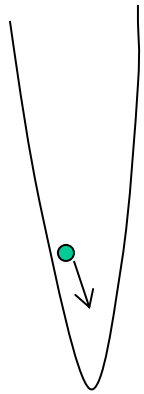


Appropriate

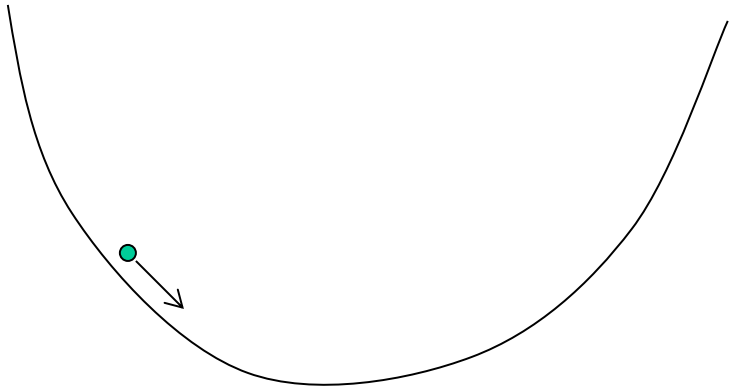
System	Types of motion	Suggested time step
Rigid molecules	translation, rotation	5 fs
Flexible molecules, rigid bonds	translation, rotation, torsions	2 fs
Flexible molecules, flexible bonds	Translation, rotation, torsions, vibrations	1 or 0.5 fs

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

31

Useful tricks

Δt limited by fast motions - 1fs

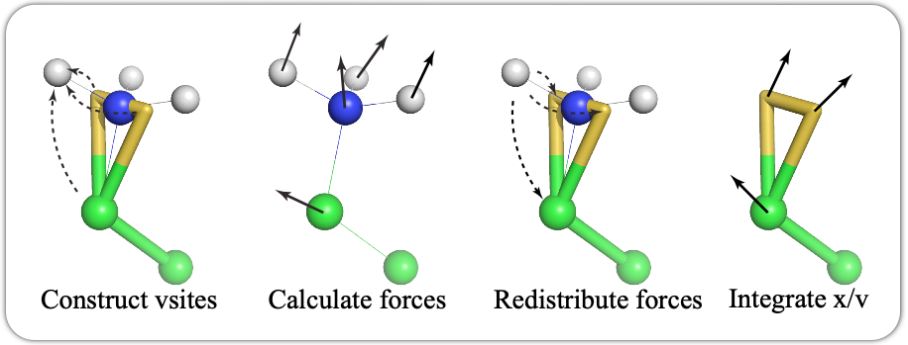
=> Remove bond vibrations => 2 fs

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

Next fastest motions is H-angle and rotations of CH3/NH2 groups

= > Remove them => upto 5fs

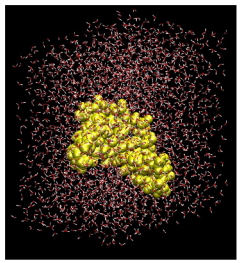
- Virtual interaction sites



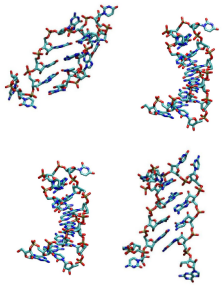
Molecular simulation and GROMACS

Input files

Structure
(*.gro , *pdb)



simulation parameter
file (*.mdp)
time step, constraint



topology file (*.top)
virtual site
-> gmx pdb2gmx

Temperature and Pressure

Thermodynamic ensembles

- ***NVE* ensemble**

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

- ***NVT* 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

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

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

Pressure in MD simulations

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

- Ideal gas: $PV = Nk_bT$
- Real system: $PV = Nk_bT - \frac{1}{3k_bT}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\vec{r}_{ij}\vec{f}_{ij}$

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

Molecular simulation and GROMACS

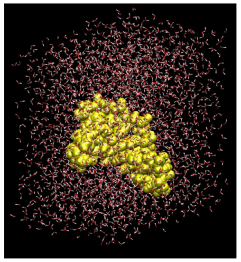
Input files

INPUT:

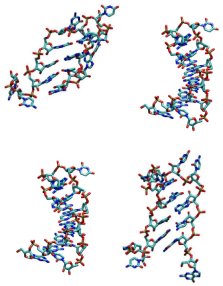
OUTPUT:

simulation parameter file (*.mdp)
temperature and pressure setting

Structure
(*.gro , *pdb)



topology file (*.top)



thermodynamic properties

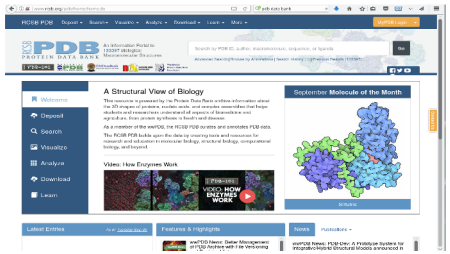
structural properties

kinetic properties

dynamic information

Starting configuration

Initial coordinates



- Experimental structures from database

(from X-ray crystallography, NMR spectroscopy, cryo-electron microscopy)

- Self-built model three-dimensional structure (e.i homology modelling, docking model)
- Pre-built solvent box

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 ?

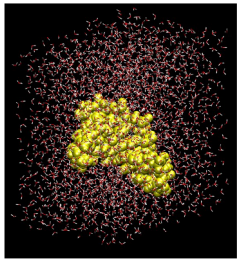
- Extra molecules (i.e. co-factors, ligands, surfactant, etc) or special conditions used to promote the observation of the experimental structure

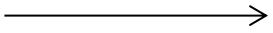
Molecular simulation and GROMACS

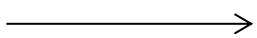
Input files

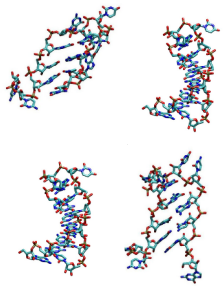
Structure
(*.gro , *pdb)
particles positions
-> gmx pdb2gmx

simulation parameter
file (*.mdp)



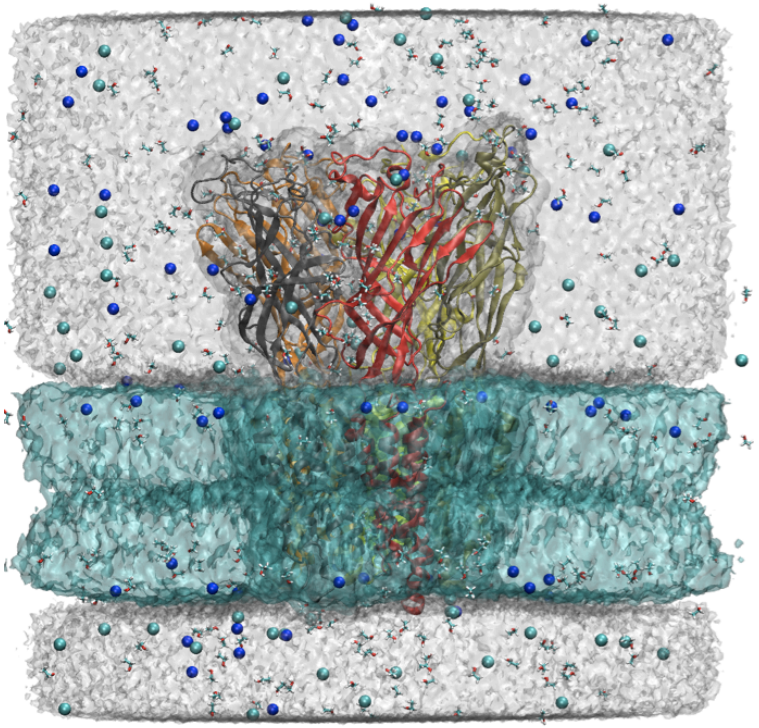






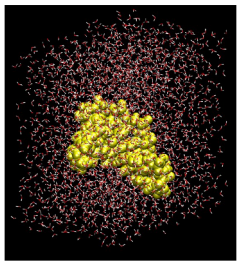
topology file (*.top)
protonation and tautomeric states, termini
-> gmx pdb2gmx

Environment – Solvent - Ions



Molecular simulation and GROMACS

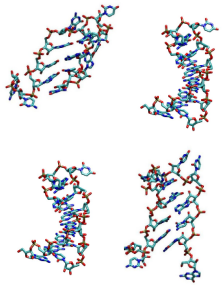
Structure
(*.gro , *pdb)
particles positions
-> gmx solvate
-> gmx genion



Input files

simulation parameter
file (*.mdp)

topology file (*.top)
Solvent and ion models
-> gmx solvate
-> gmx genion



Molecular simulation files in GROMACS

INPUT:

Structure
(*.gro, *.pdb)

Simulation parameter
(*.mdp)

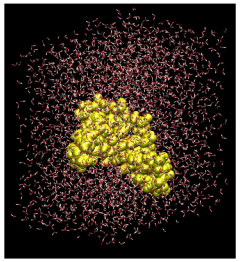
Topology (*.top)

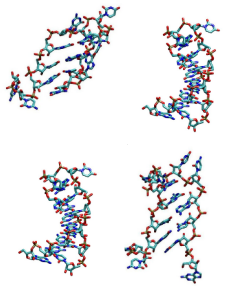
gmx grompp

*.tpr file

gmx mdrun

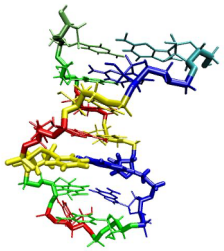
OUTPUT:
Ensemble of structures, called configurations
(*.xtc/*.trr trajectory, *.edr energy file, *.log log file)





System setting

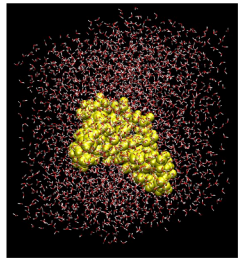
Starting structure



generate
topology file

define box
include environment
(ion, solvent)

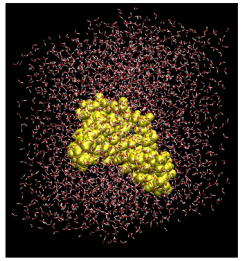
Starting configuration



Choose the initial coordinate, visualize the structure, add missing atoms

Choose potential function, solvent and ion models. Generate/Implement missing parameters

Simulation setting



*.mdp

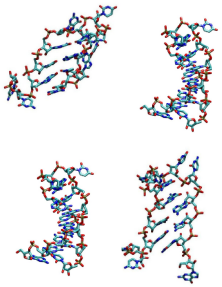
Energy minimization

*.mdp

Solvent and ion relaxation, system equilibration

Data production

*.mdp



Ensemble of structures (called configurations)

thermodynamic properties

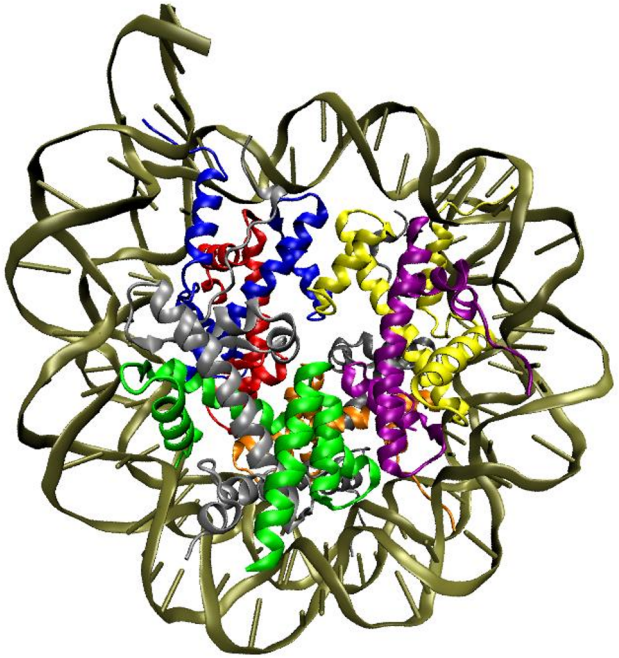
structural properties

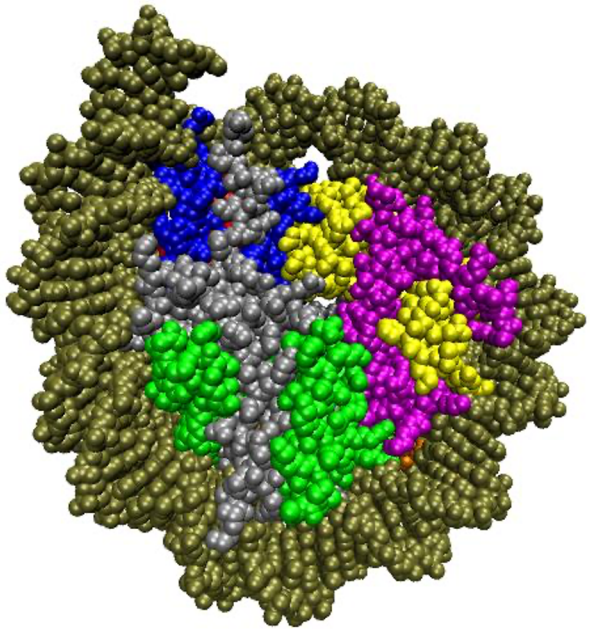
kinetic properties

dynamic information

For each step a dedicated simulation parameter file (*mdp) has to be created

Visualization





Some visualization softwares:

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

Rasmol http://www.openrasmol.org/

Pymol http://www.pymol.org/

Q & A

BioExcel Partners

Horizon 2020
European Union Funding
for Research & Innovation

BioExcel is funded by the European Union Horizon 2020 program under grant agreements 675728 and 823830.