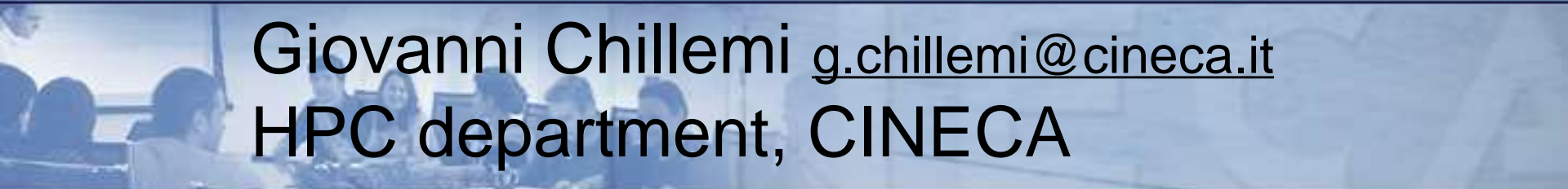
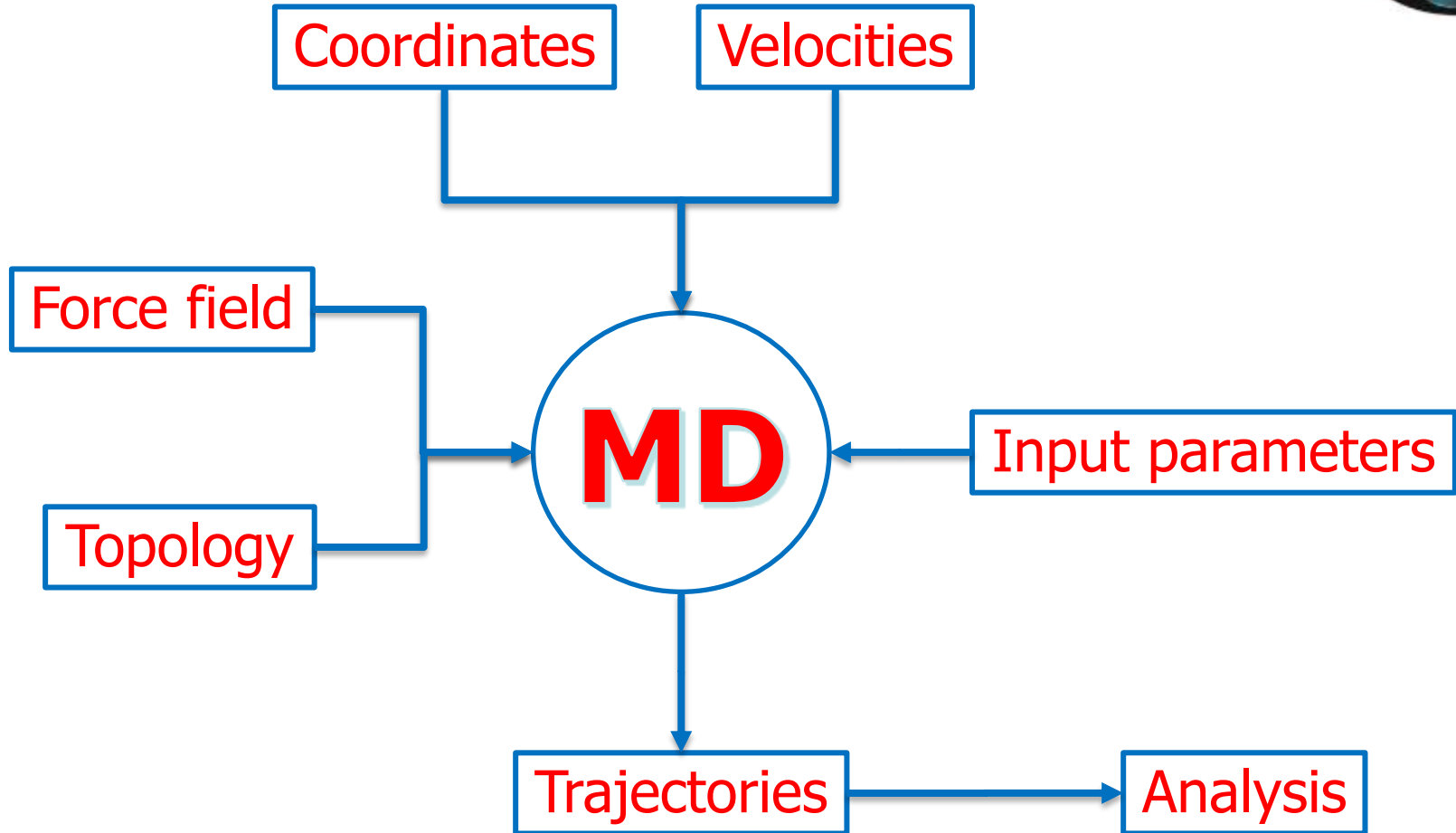
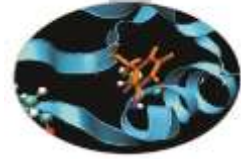


Introduction to Classical Molecular Dynamics

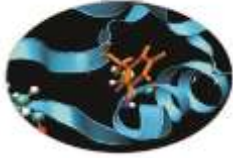


Giovanni Chillemi g.chillemi@ Cineca.it
HPC department, CINECA

MD ingredients



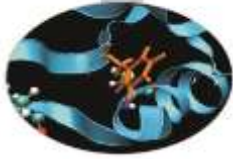
Equation of motion



The equations that describe the temporal evolution of a physical system is called **equation of motion**. There are several different equations of motions, which characterize the motion in different types of physical systems:

- Time-dependent Schrödinger's Equation
 - for quantum-mechanical system
- **Newton's Equation**
 - **for classical-mechanical system**
- Langevin's Equation
 - for stochastic system

Newton's Equation of motion



Molecules are quantum-mechanical systems whose motion should be described by Schrödinger's Equation. However, technical difficulties make solving Schrödinger's Equation for large systems impractical.

Therefore the motion of a molecule is usually approximated by the laws of classical mechanics and by Newton's equation of Motion. In its most simplistic form Newton's second law of motion states:

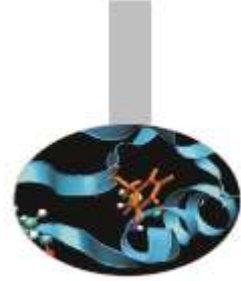
$$\mathbf{f}_i = m_i \cdot \mathbf{a}_i$$

where m_i is the mass of particle i , a_i is its acceleration. The force f_i is given as the derivative of the potential energy function V :

$$\mathbf{f}_i = -\frac{\partial V}{\partial \mathbf{r}_i}$$

where r_i is the position of particle i

Potential energy function



$$\begin{aligned}
 V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = & \sum_{bond} \frac{1}{2} k_{b_n} (b_n - b_{0_n})^2 + \sum_{angle} \frac{1}{2} k_{\theta_n} (\theta_n - \theta_{0_n})^2 + \\
 & + \sum_{\substack{improper \\ dihedral}} \frac{1}{2} k_{\xi_n} (\xi_n - \xi_{0_n})^2 + \sum_{dihedral} k_{\phi_n} [1 + \cos(m_n \phi_n - \delta_n)] + \\
 & + \sum_{\substack{nonbonded \\ pairs(ij)}} \left(\left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r_{ij}} \right)
 \end{aligned}$$



non bonded interactions



bonded interactions

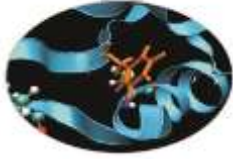


Model	Degree of freedom		Example of predicted properties
	Considered	Removed	
Quantum mechanic	Nucleus, electrons	Nucleons	Chemistry reaction
Polarizable atoms	Atoms, dipoles	Electrons	Binding of charged substrates
Non polarizable atoms	Solute atoms, solvent atoms	Dipoles	Conformational transitions Hydration
Implicit solvent	Solute atoms	Solvent atoms	Folding topology of macromolecules



Classical Molecular Dynamics

A Brief History



Year	System
1964	Liquid Argon (Rahman Phys Rev)
1974	Water (Rahman J. Chem Phys.)
1977	Small protein in vacuo (Mc Cammon Karplus Nature)
1988	First Protein in explicit water (Levitt PNAS)
From 1995	Protein-DNA Complexes – Membrane Proteins- Complex Systems

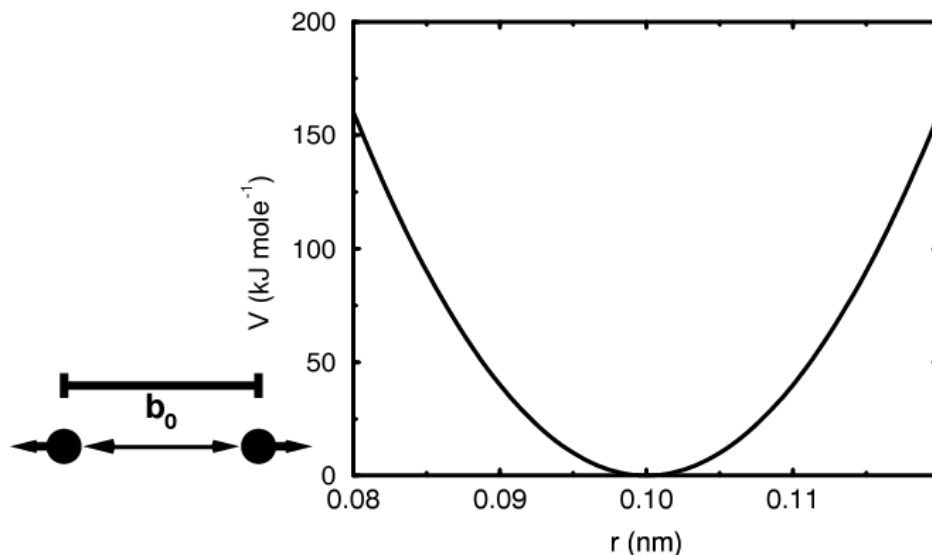
Bond Stretching Energy



$$= \sum_{bond} \frac{1}{2} k_{b_n} (b_n - b_{0_n})^2 + ..$$

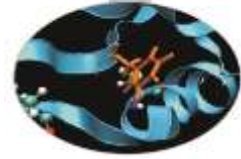
k_b is the spring constant of the bond
 b_0 is the bond length at equilibrium

Unique k_b and b_0 assigned for each bond pair, i.e. C-C, O-H



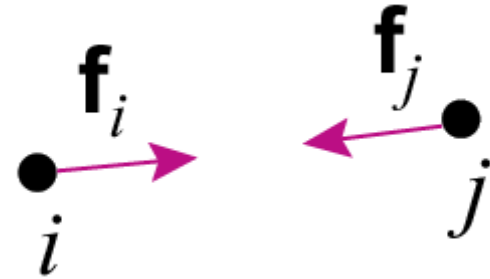
Principle of bond stretching (left), and the bond stretching potential (right).

Bond Stretching Force



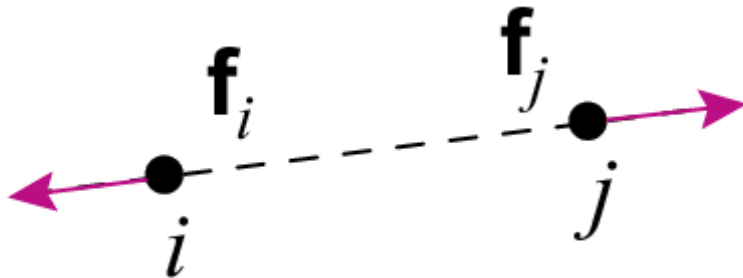
$$\mathbf{f}_i = -\frac{\partial V^{bond}}{\partial \mathbf{r}_i} = -\frac{\partial V^{bond}}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \mathbf{r}_i} = k_b (r_{ij} - b_0) \frac{\mathbf{r}_{ij}}{r_{ij}}$$

$$\mathbf{f}_j = -\mathbf{f}_i$$

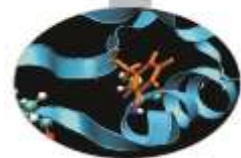


If atom i and j are closer than b_0 ,
the bond force separates them

If atom i and j are farther
than b_0 , the bond force
draws them nearer

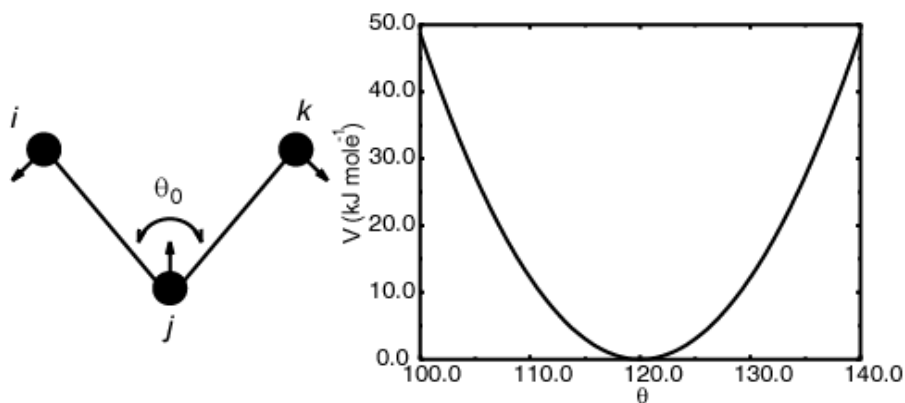


Bending Energy



$$+ \sum_{\text{angle}} \frac{1}{2} k_{\theta_n} (\theta_n - \theta_{0n})^2 +$$

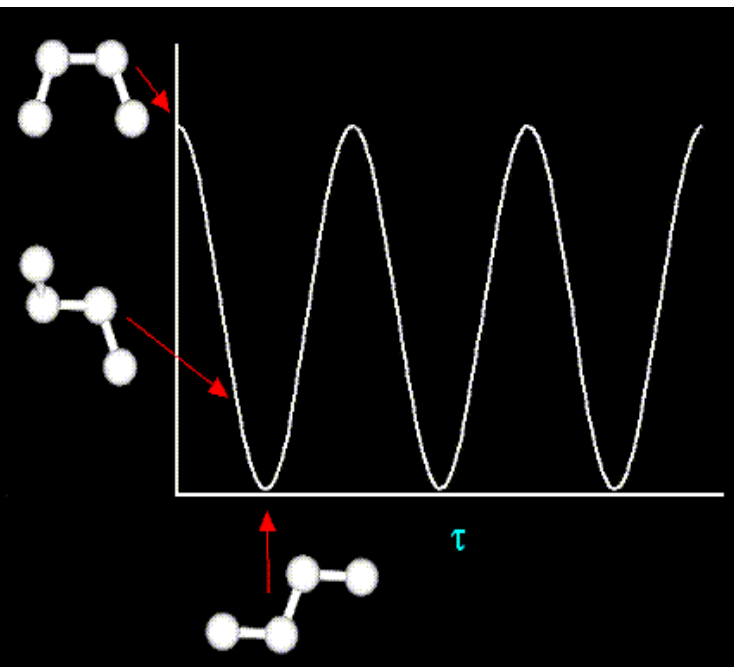
k_{θ} is the spring constant of the bending.
 θ_0 is the angle bending at equilibrium.



Principle of angle vibration (left) and the bond angle potential (right).

Unique parameters for angle bending are assigned to each bonded triplet of atoms based on their types (e.g. C-C-C, C-O-C, C-C-H, etc.)

Torsional or Dihedral Energy

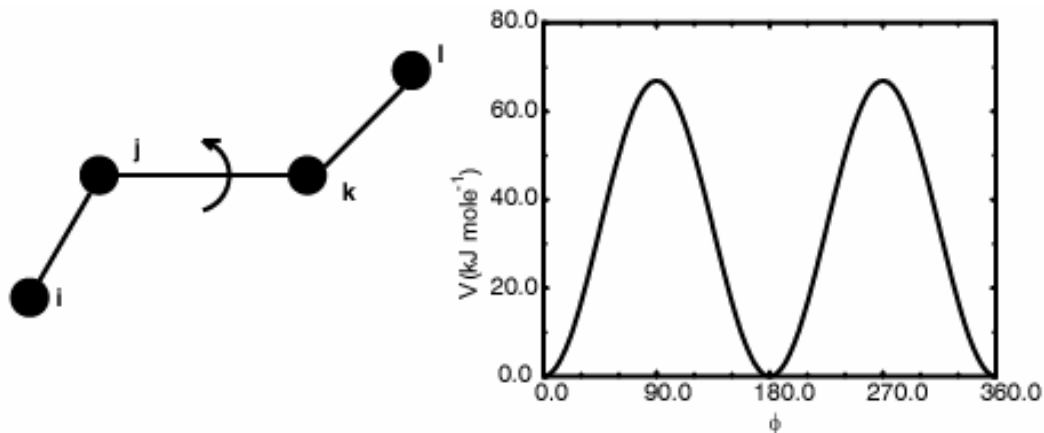


$$+ \sum_{dihedral} k_{\phi_n} [1 + \cos(m_n \phi_n - \delta_n)] +$$

ϕ = angle

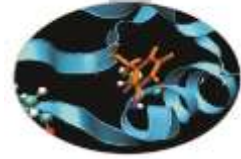
δ = phase

m = number of peaks in a full rotation

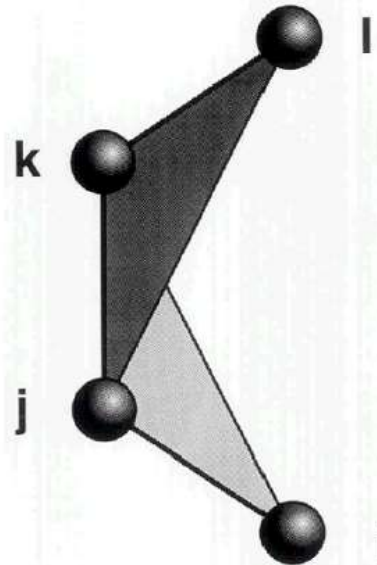
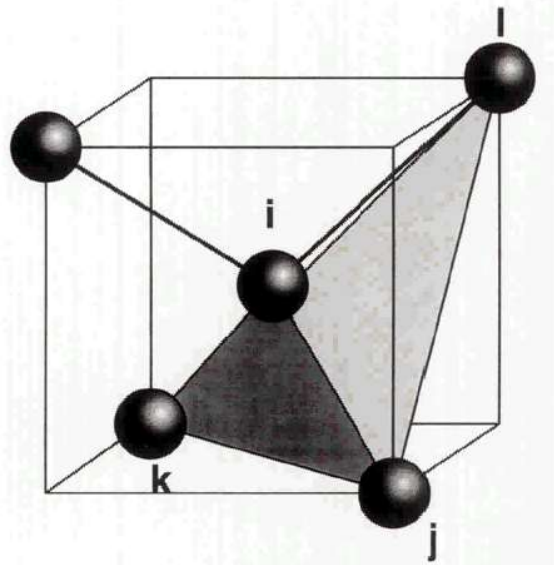


Principle of proper dihedral angle (left, in *trans* form) and the dihedral angle potential (right).

Improper Dihedral Energy



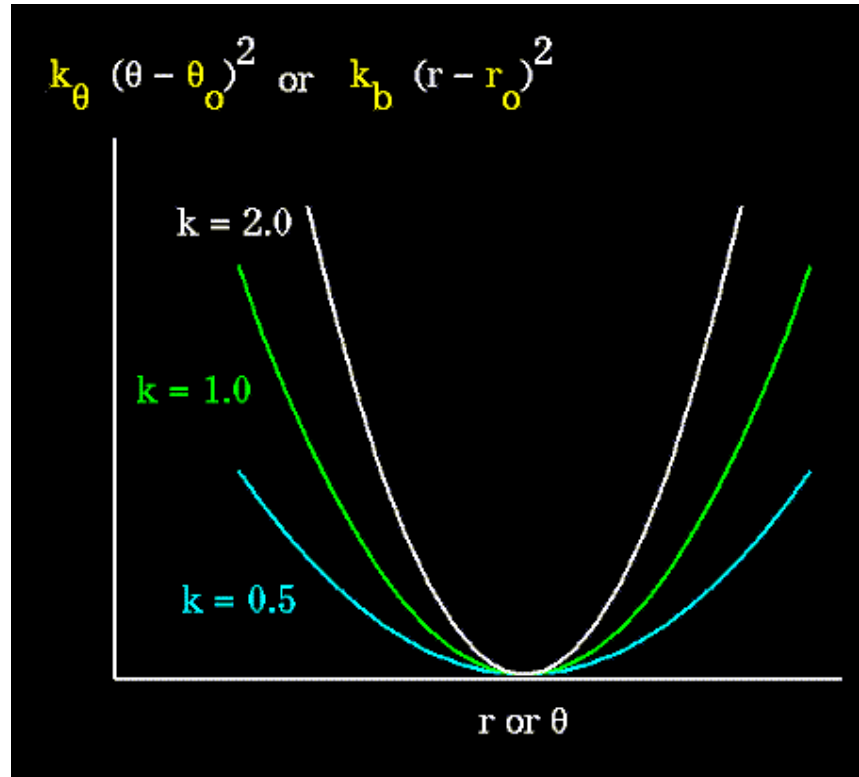
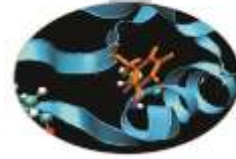
The energy required to deform a group of atoms from its equilibrium angle, x_0 . Used for tetrahedral or planar groups



$$+ \sum_{\text{improper dihedral}} \frac{1}{2} k_{\xi_n} (\xi_n - \xi_{0n})^2 +$$

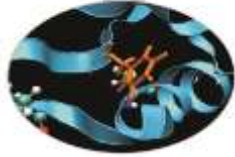
Again this system can be modeled by a spring, and the energy is given by the Hookean potential with respect to the planar angle

The “Hookean” potential

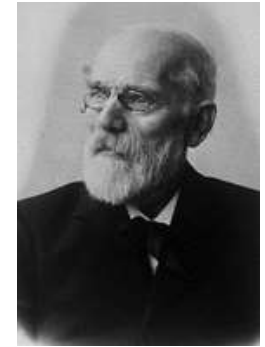
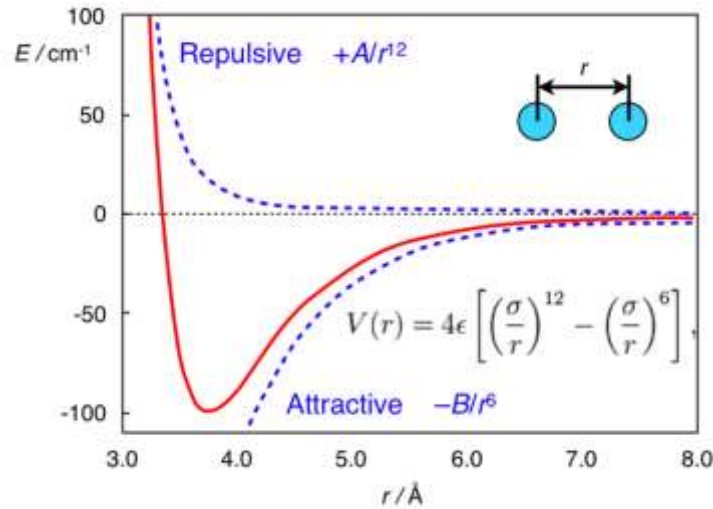


k_b and k_{θ} broaden or steepen the slope of the parabola
The larger the value of k , the more energy is required to deform an angle (or bond) from its equilibrium value

Lennard Jones (Van der Waals) interactions



Sir John Lennard Jones

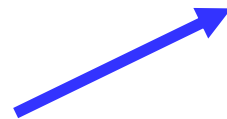


Johannes Diderik Van der Waals

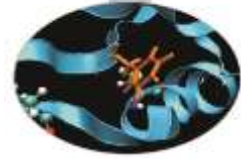
LJ interactions

$$+ \sum_{\text{nonbonded pairs}(ij)} \left(\left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r_{ij}} \right)$$

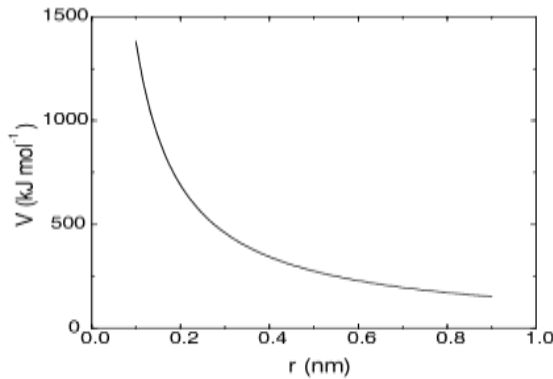
non bonded interactions



Electrostatic interactions



The q_i and q_j are the partial atomic charges for atoms i and j , separated by a distance r_{ij} . ϵ_r is the relative dielectric constant. For gas phase calculations ϵ_r is normally set to 1. Larger values of ϵ_r are used to approximate the dielectric effect of intervening solute ($\epsilon \sim 60-80$) or solvent atoms in solution



The Coulomb interaction (for particles with equal signed charge)

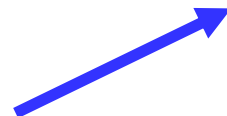
Electrostatic

$$+ \sum_{\text{nonbonded pairs}(ij)} \left(\left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r_{ij}} \right)$$

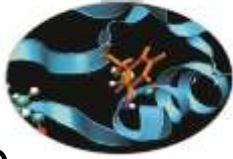


Charles Augustin de Coulomb

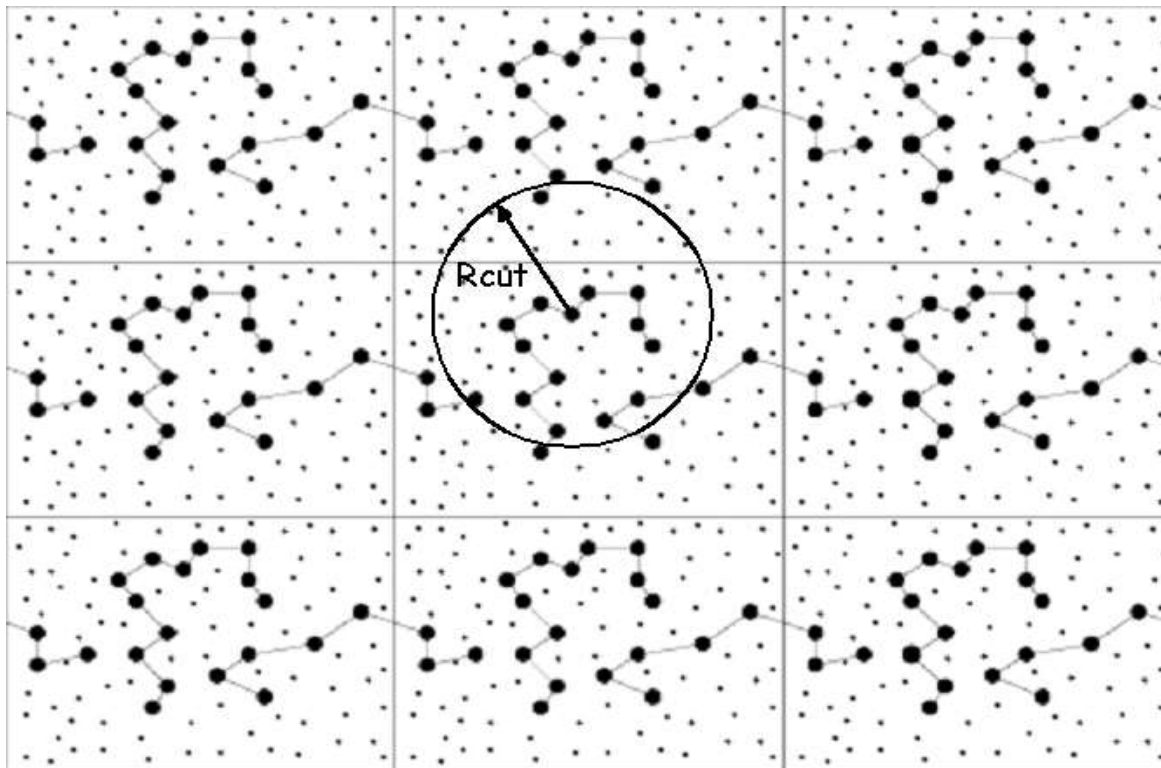
non bonded interactions



Periodic boundary conditions and cut-off radius



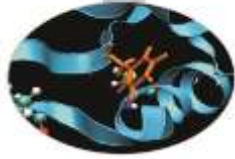
To simulate our finite system in liquid conditions, we apply the pbc: i.e. the system box is virtually surrounded in all directions by copy of itself. An atom close to a box border interacts with the atoms in another pbc image. The non-bonded interactions are only calculated between atom pairs closer than a spherical cut-off



BOX dimension

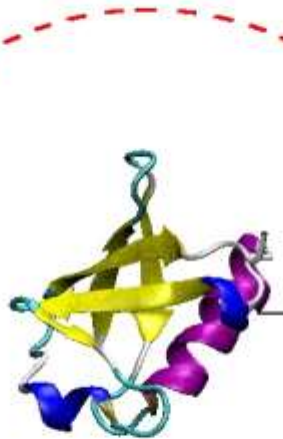
The edge of cubic box must be large enough to avoid interactions of the solute with itself.

Its minimal dimension therefore depends by the chosen cut-off for the non bonded interactions



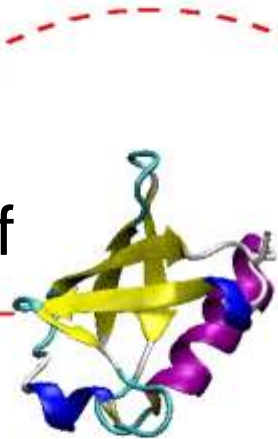
Original Protein

Periodic Image



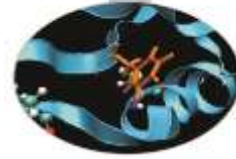
cut-off

d = 0.8 nm



Edge of the box = 3.2 nm

Electrostatic interactions: Particle Mesh Ewald (PME)



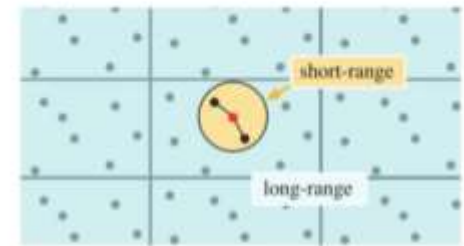
- Short range in the real space
- Long range in the Fourier space

Coulomb forces on N particles, charges q_i , positions \mathbf{r}_i , box length L , periodic b.c.

- electrostatic potential

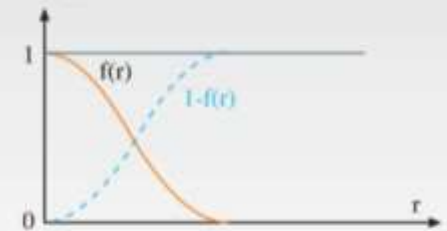
$$V = \frac{1}{2} \sum_{i,j=1}^N \sum_{\mathbf{n} \in \mathbb{Z}^3} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|}$$

straightforward summation
impracticable



- Trick 1:** Split problem into 2 parts with help of:

$$\frac{1}{r} = \underbrace{\frac{f(r)}{r}}_{\text{short range}} + \underbrace{\frac{1-f(r)}{r}}_{\text{long range}}$$



$$V = \underbrace{V_{dir}}_{\text{direct space}} + \underbrace{V_{rec}}_{\text{fourier space}}$$

The cut-off radius method for electrostatic interactions is particularly inaccurate for charged molecules such as DNA or for dipolar groups such as alpha helices

PME corrects these errors and it helps maintaining short the cut-off in the real space: i.e. the number of atom pairs is reduced

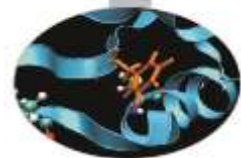
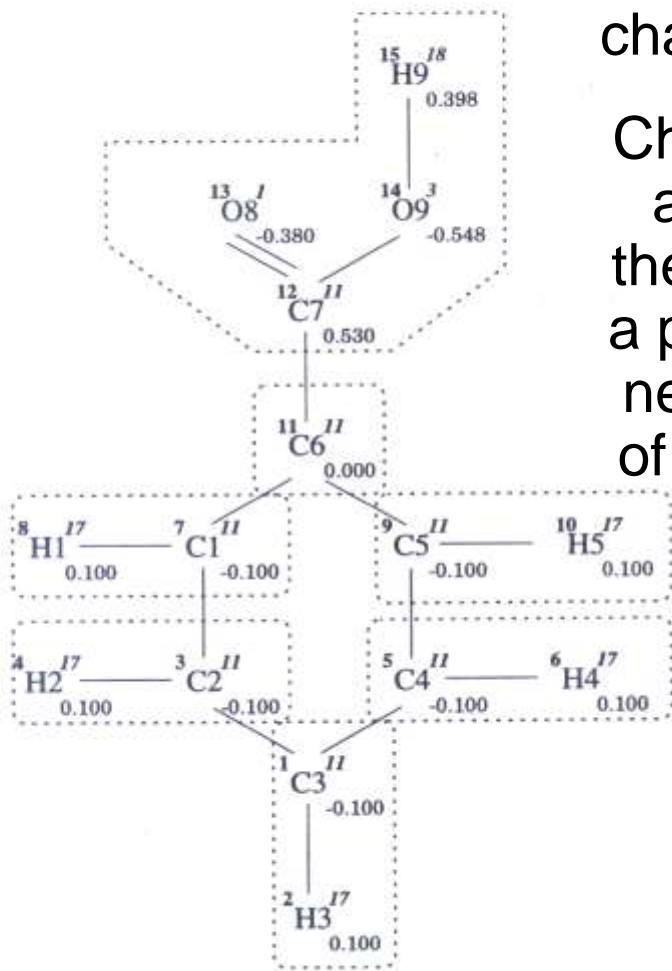


Fig. 4.2.70 Molecular topology building block definition

Solute building block: Benzoic acid (neutral)
Name: BA

a. Atoms

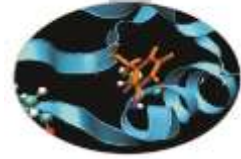


A **charge group** is a neutral charge group composed by several partially charged atoms of a chemical group. Electrostatics can be calculated between charge groups instead that atom pairs

Charge groups were first introduced to reduce artifacts in the electrostatics calculation but they can also speed up the calculations; given a pair of water molecules for instance, we only need to determine one atom distance instead of nine (or sixteen for a four-site water model)

Note that an **atom type** is not a physical feature. O8 is defined with a different atom type than O9. In fact, their bond constants with C7 and atomic charges are different

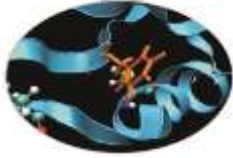
Force field



$$\begin{aligned}
 V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = & \sum_{bond} \frac{1}{2} k_{b_n} (b_n - b_{0_n})^2 + \sum_{angle} \frac{1}{2} k_{\theta_n} (\theta_n - \theta_{0_n})^2 + \\
 & + \sum_{\substack{improper \\ dihedral}} \frac{1}{2} k_{\xi_n} (\xi_n - \xi_{0_n})^2 + \sum_{dihedral} k_{\phi_n} [1 + \cos(m_n \phi_n - \delta_n)] + \\
 & + \sum_{\substack{nonbonded \\ pairs(ij)}} \left(\left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^6} \right) + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_r r_{ij}} \right)
 \end{aligned}$$

The potential energy function, together with the parameters required to describe the behavior of different kinds of atoms and bonds (k_b , k_θ , k_ξ , C_{ij} , ...), is called a **force field**.

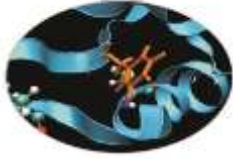
Several force fields are currently used and the choice depends from the studied system. Some force field are better suited for nucleic acids, for example, while others for membrane proteins



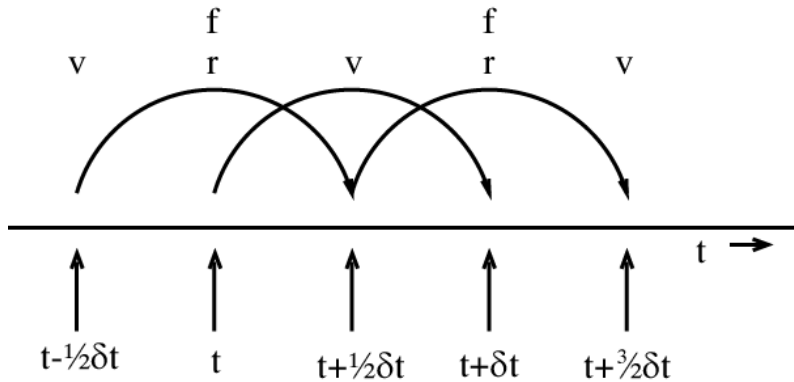
Numeric integration of Newton's equation of motion is typically done step by step using methods that are called **Finite Difference** methods.

These methods use the information available at time t to predict the system's coordinates and velocities at a time $t + \delta t$, where δt is a short time interval and are based on a Taylor expansion of the position at time $t + \delta t$

$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \dots$$



$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \quad \blacklozenge \text{ Verlet integrator}$$



$$r(t + \delta t) = r(t) - v(t + \frac{1}{2} \delta t) \delta t$$

$$v(t + \frac{1}{2} \delta t) = v(t - \frac{1}{2} \delta t) + a(t) \delta t$$

\blacklozenge Leap-frog integrator

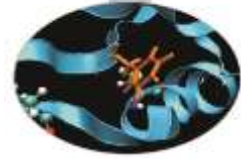
$$r(t + \delta t) = r(t) + v(t)\delta t + \frac{1}{2} a(t)\delta t^2$$

$$v(t + \delta t) = v(t) + [a(t) + a(t + \delta t)] \frac{\delta t}{2}$$

\blacklozenge Velocity Verlet



Choice of the timestep

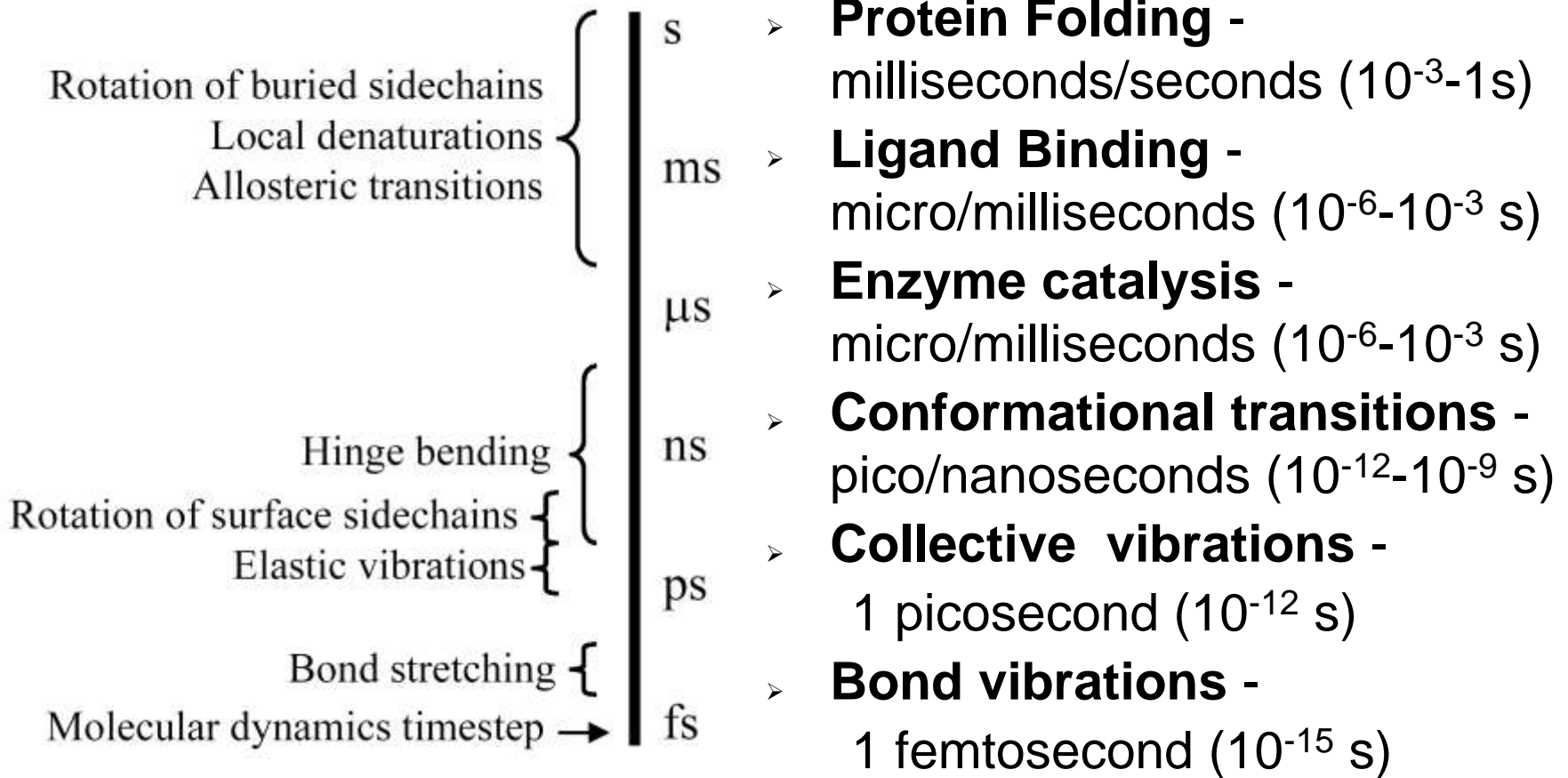
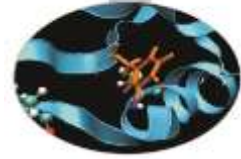


The length of the timestep must be small compared to the period of the highest frequency motions being simulated

Force characteristics	Relaxation time (fs)	Time step (fs)
High frequency motion: bond stretching vibrations	10	0.5
Medium frequency motion: angle bending, proper and improper dihedral angle deformation, LJ and short range Coulombian interactions	40	2
Low frequency motion: long range coulombian interactions	1000	20

The bond stretching vibrations are generally of minimal interest in the study of biomolecular structure and function. Therefore this degree of freedom is usually kept frozen with constraint algorithms such as Shake, Settle, Lincs and a typical timestep is 2 fs (2×10^{-15} s)

Timescale



Topology



The topology file describes the atoms composing a molecule and their bond connections

Es: flexspc.itp in gromacs

```
[ moleculetype ]
```

```
; molname      nrexcl
SOL            2
```

```
[ atoms ]
```

; id	at type	res nr	res name	at name	cg nr	charge	mass
1	OW_spc	1	SOL	OW	1	-0.82	15.99940
2	HW_spc	1	SOL	HW1	1	0.41	1.00800
3	HW_spc	1	SOL	HW2	1	0.41	1.00800

```
[ bonds ]
```

; i	j	funct	length	force.c.
1	2	1	0.1	345000 0.1 345000
1	3	1	0.1	345000 0.1 345000

```
[ angles ]
```

; i	j	k	funct	angle	force.c.
2	1	3	1	109.47 383	109.47 383

Topology

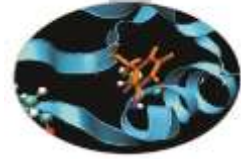
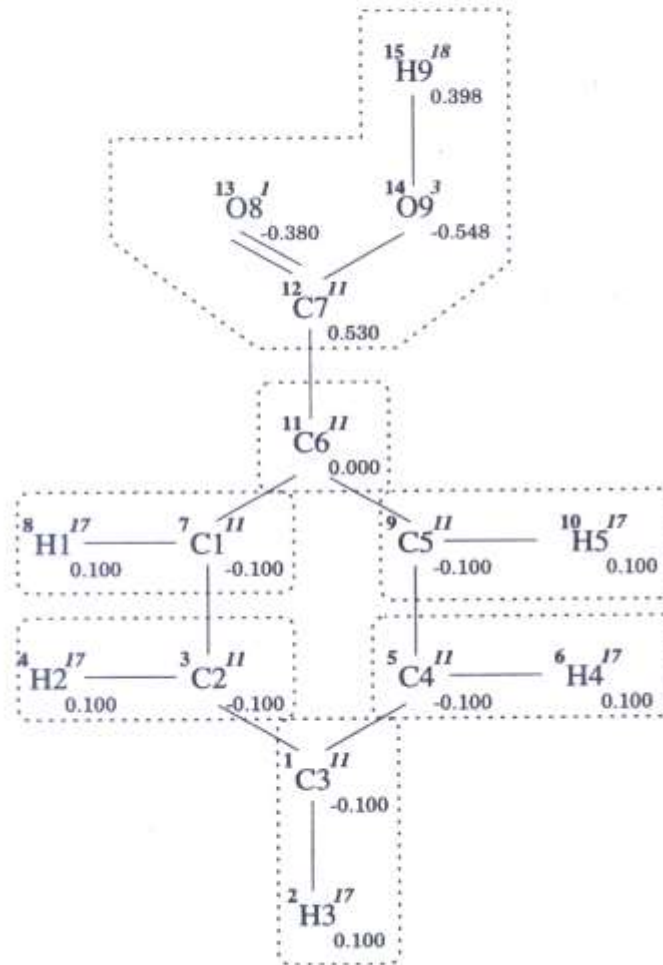


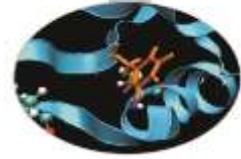
Fig. 4.2.70 Molecular topology building block definition

Solute building block: Benzoic acid (neutral)
Name: BA

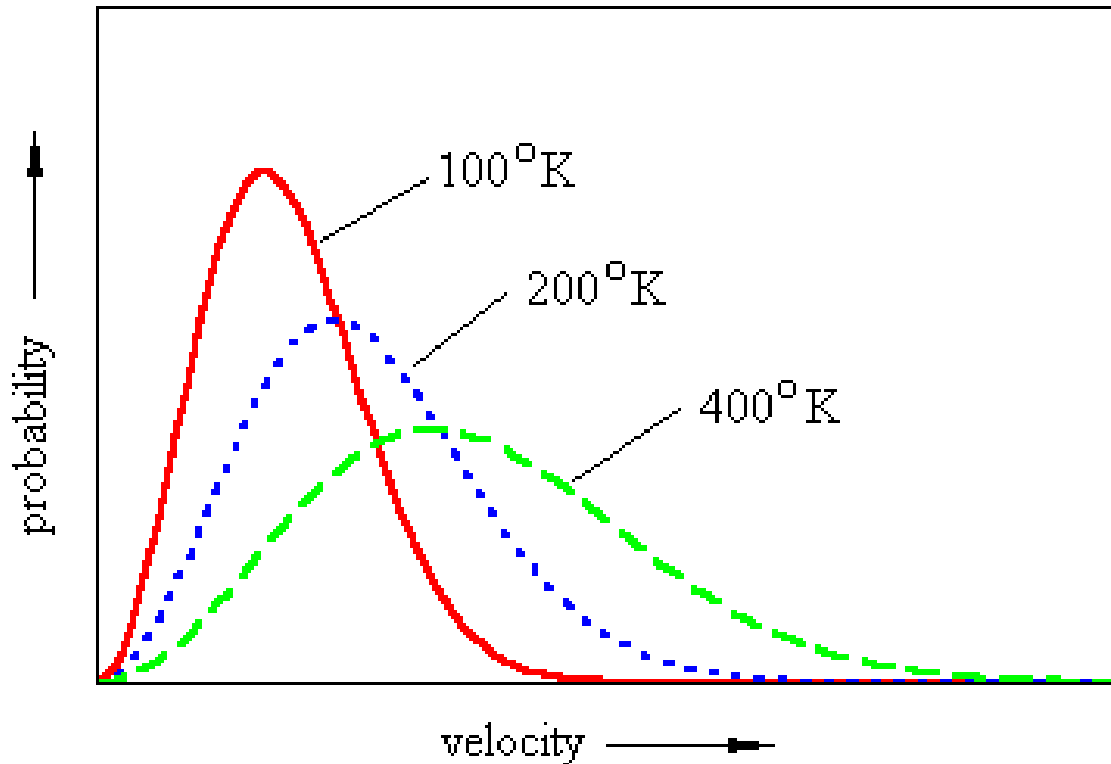
a. *Atoms*



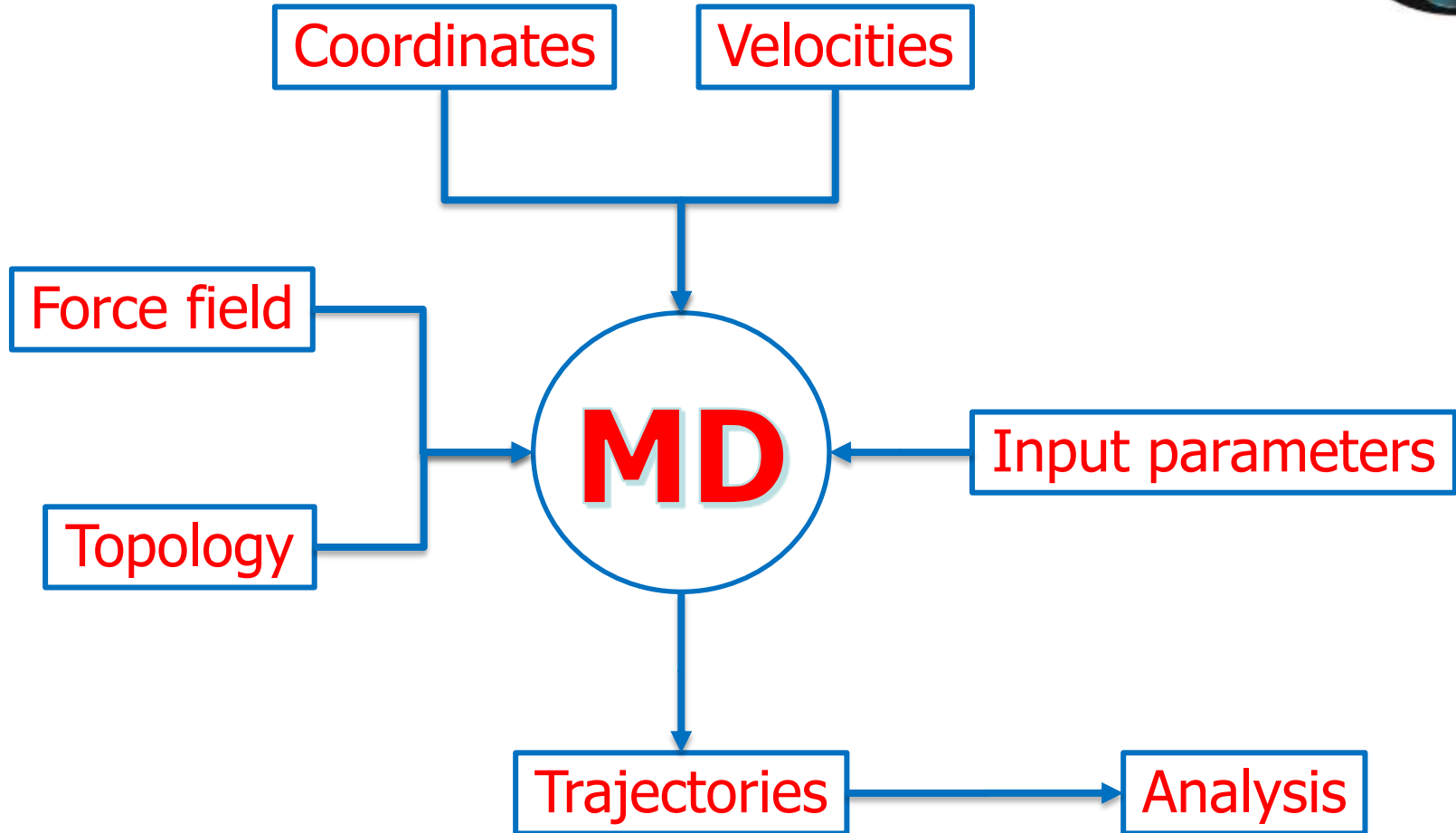
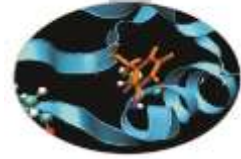
Initial velocities



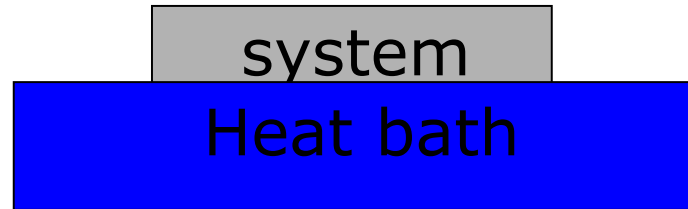
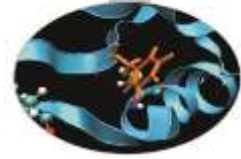
The initial velocity of each atom is random assigned through a **Maxwell-Boltzmann** distribution that is function of the temperature



To recapitulate..



Simulating at constant T: the Berendsen scheme



Bath supplies or removes heat from the system as appropriate

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T}$$

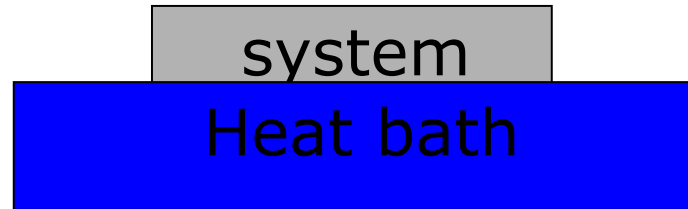
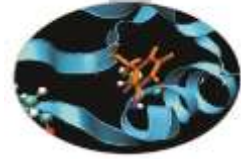
where τ determines how strong the bath influences the system

Exponentially scale the velocities at each time step by the factor λ :

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left(\frac{T_0}{T(t)} - 1 \right) \right]^{\frac{1}{2}}$$

T: “kinetic” temperature

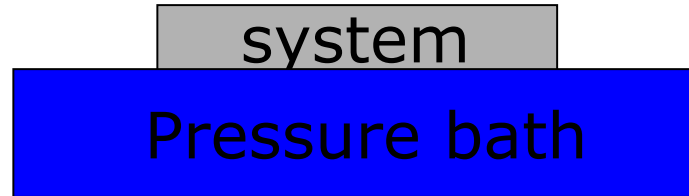
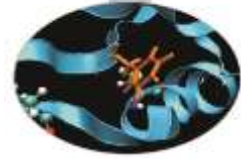
Simulating at constant T: the Berendsen scheme



A small τ , close to the timestep (strong thermostat), is useful in the equilibration phase, when the quick decreasing of the potential energy could increase too much the kinetic energy of the protein

A bigger τ , e.g. equal to ten times the timestep (weak thermostat), is useful in the production phase, when we want to keep at minimum the perturbation to the conformational sampling

Simulating at constant P: the Berendsen scheme



Couple the system to a pressure bath

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P}$$

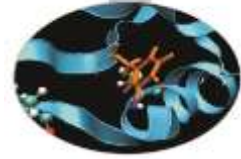
A change in pressure P is related to a change in volume V

To regulate pressure: exponentially scale the volume of the simulation box at each time step by a factor μ

$$\mu(t) = \left[1 - k_T \frac{\Delta t}{\tau_P} (P_0 - P(t)) \right]^{\frac{1}{3}}$$

where k_T : isothermal compressibility
 τ_P : coupling constant

Sample input file of gromacs



<http://manual.gromacs.org/current/online/mdp.html>

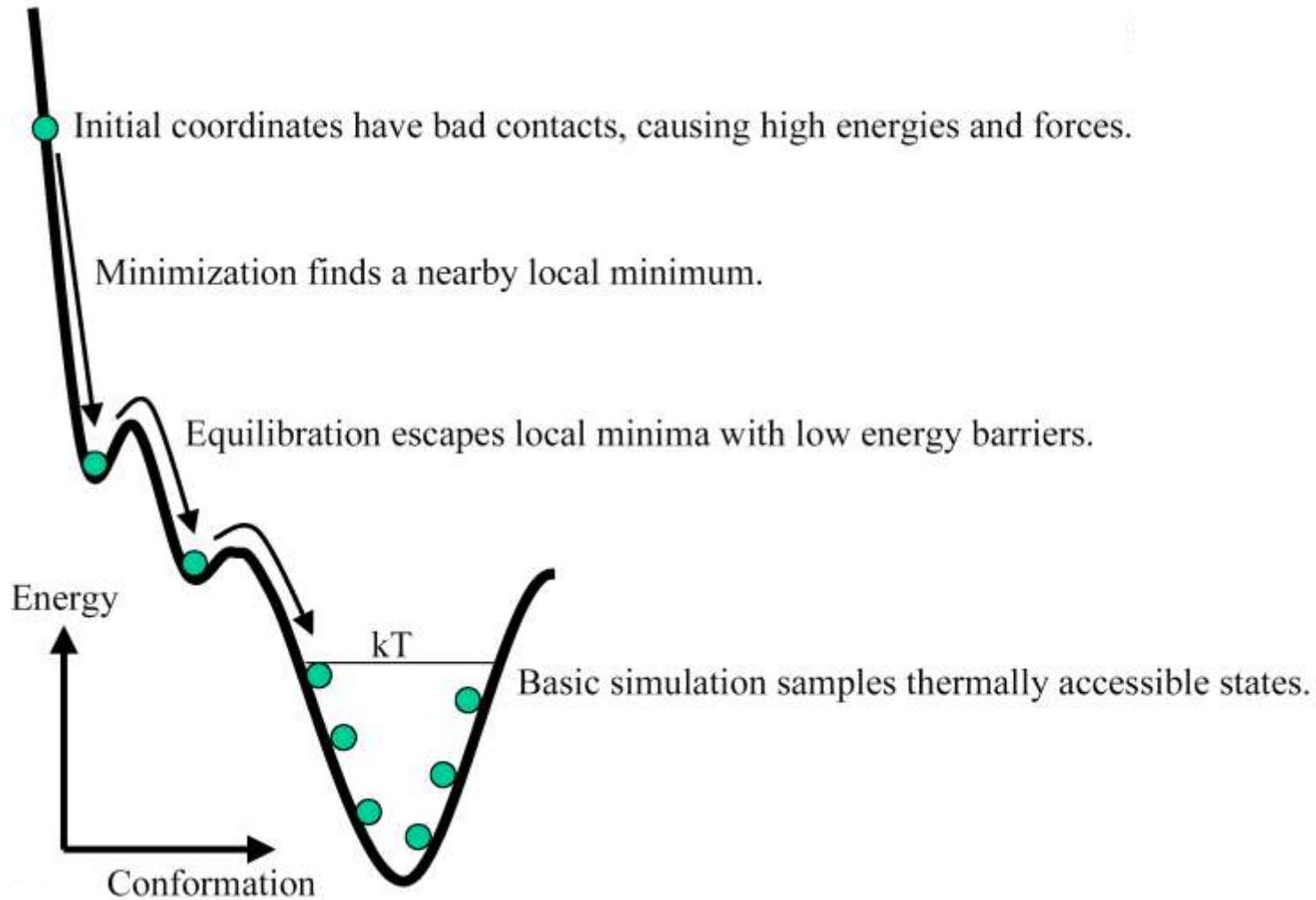
```

title                = Yo
cpp                  = /lib/cpp
include              = -l../top
define               =
integrator           = md
dt                   = 0.002
nsteps               = 500000
nstxout              = 5000
nstvout              = 5000
nstlog               = 5000
nstenergy            = 250
nstxout-compressed   = 250
compressed-x-grps    = Protein
energygrps           = Protein SOL
nstlist              = 10
ns-type              = grid
rlist                = 0.8
coulombtype          = cut-off
rcoulomb             = 1.4
rvdw                 = 0.8

tcoupl               = Berendsen
tc-grps              = Protein SOL
tau-t                = 0.1 0.1
ref-t                = 300 300
Pcoupl               = Berendsen
tau-p                = 1.0
compressibility      = 4.5e-5
ref-p                = 1.0
gen-vel              = yes
gen-temp             = 300
gen-seed             = 173529
constraints           = all-bonds
  
```



Conformational sampling



1-4 interactions

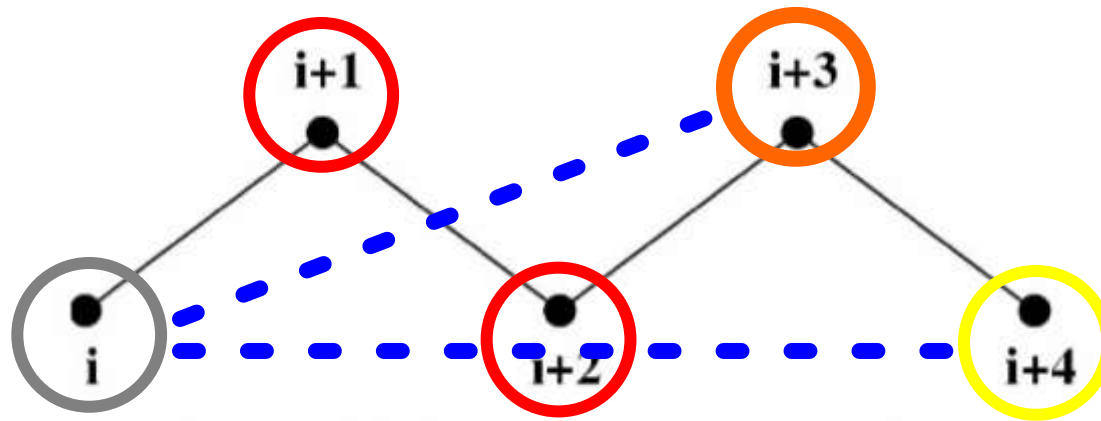
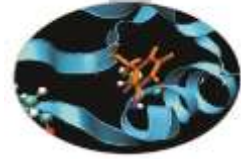


Figure 4.15: Atoms along an alkane chain.

Atoms covalently bound are defined as first neighbours second neighbours and so on....

LJ and electrostatic interactions **are not calculated among** first and second neighbours since they are considered in the stretching (first) or in the bending potential (second)

The standard non-bonding interactions are too strong for the third neighbours and are reduced (interactions 1-4; list 1-4)