

Introduction to Classical Molecular Dynamics

Alessandro Grottesi a.grottesi@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 different equations of motions, which characterize the motion with different levels of approximation:

- > 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 <u>classical mechanics</u> and by Newton's equation of Motion. In its most simplistic form Newton's second law of motion states:

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

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





non 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
mplicit solvent	Solute atoms	Solvent atoms	Folding topology of macromolecules

Classical Molecular Dynamics





F

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)
rom 1995	Protein-DNA Complexes – Membrane Proteins- Complex Systems





CA



Bond Stretching Energy





Unique k_b and b_0 assigned for each bond pair, i.e. C-C, O-H k_{b} is the spring constant of the bond b_{0} is the bond length at equilibrium







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 b₀, the bond force draws them nearer





Bending Energy

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

 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 o Dihedral Energy



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

 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



+ $\sum_{\substack{improper \\ j \neq i}} \frac{1}{2} k_{\xi_n} (\xi_n - \xi_{0_n})^2 +$

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_c , 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



SuperComputing Applications and Innovation Lennard Jones (Van der Waals) interactions







```
Johannes Diderik Van der Waals
```







Electrostatic interactions



 q_i and q_j are the partial atomic charges for atoms *i* and *j*, separated by a distance r_{ii} and ϵ_r is the relative dielectric constant:



The Coulomb interaction (for particles with equal signed charge)



Electrostatic interactions: Particle Mesh Ewald (PME)



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

The cut-off radius method for electrostatic interactions is particularly inaccurate for charged molecules such as DNA of 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



Darden et al., Particle mesh Ewald-an NlogN method for Ewald sums in large systems. J. Chem. Phys, 98 (1993), 10089



Charge groups and atom types

SuperComputing Applications and Innovation

Fig. 4.2.70 Molecular topology building block definition





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



IV-242



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







Force field



$$V(\mathbf{r}_{1},\mathbf{r}_{2},...,\mathbf{r}_{n}) = \sum_{bond} \frac{1}{2} k_{b_{n}} (b_{n} - b_{0_{n}})^{2} + \sum_{angle} \frac{1}{2} k_{\theta_{n}} (\theta_{n} - \theta_{0_{n}})^{2} + \sum_{bond} \frac{1}{2} k_{\xi_{n}} (\xi_{n} - \xi_{0_{n}})^{2} + \sum_{dihedral} k_{\phi_{n}} [1 + \cos(m_{n}\phi_{n} - \delta_{n})] + \sum_{dihedral} \frac{1}{2} k_{\xi_{n}} (\xi_{n} - \xi_{0_{n}})^{2} + \sum_{dihedral} k_{\phi_{n}} [1 + \cos(m_{n}\phi_{n} - \delta_{n})] + \sum_{dihedral} \sum_{dihedral} \frac{1}{2} \left[\frac{1}{2} \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \right] + \frac{1}{4\pi\varepsilon_{0}} \frac{q_{i}q_{j}}{\varepsilon_{r}r_{ij}} \left[\frac{1}{2} \frac{1}{2} \frac{C_{ij}^{(12)}}{\varepsilon_{r}r_{ij}} - \frac{C_{ij}^{(6)}}{\varepsilon_{r}r_{ij}} \right] + \frac{1}{4\pi\varepsilon_{0}} \frac{1}{\varepsilon_{r}r_{ij}} \left[\frac{1}{2} \frac{1}{2} \frac{1}{\varepsilon_{r}r_{ij}} \frac{1}{\varepsilon_{r}r_{ij}} \right]$$

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: **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





Available forcefield in Gromacs (5.1.2)



- 1: AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
- 2: AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)
- 3: AMBER96 protein, nucleic AMBER94 (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)
- 4: AMBER99 protein, nucleic AMBER94 (Wang et al., J. Comp. Chem. 21, 1049-1074, 2000)
- 5: AMBER99SB protein, nucleic AMBER94 (Hornak et al., Proteins 65, 712-725, 2006)
- 6: AMBER99SB-ILDN protein, nucleic AMBER94 (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010)
- 7: AMBERGS force field (Garcia & Sanbonmatsu, PNAS 99, 2782-2787, 2002)
- 8: CHARMM27 all-atom force field (CHARM22 plus CMAP for proteins)
- 9: GROMOS96 43a1 force field
- 10: GROMOS96 43a2 force field (improved alkane dihedrals)
- 11: GROMOS96 45a3 force field (Schuler JCC 2001 22 1205)
- 12: GROMOS96 53a5 force field (JCC 2004 vol 25 pag 1656)
- 13: GROMOS96 53a6 force field (JCC 2004 vol 25 pag 1656)
- 14: GROMOS96 54a7 force field (Eur. Biophys. J. (2011), 40, 843-856
- 15: OPLS-AA/L all-atom force field (2001 aminoacid dihedrals)



SCA Integration of the equation of motion



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 + δt , where δt is a short time interval and are based on a Taylor expansion of the position at time t + δt

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



SCAI Integration of the equation of motion

SuperComputing Applications and Innovation



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

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

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 (2 x 10⁻¹⁵ s)





Constraints



The application of geometrical constraints to maintain fixed all the bond lengths, during a simulation, allows the use of a time step up to 2 fs

In SHAKE (the first constraints algorithm to be implemented in a MD code) changes a set of unconstrained coordinates to a set of coordinates that fulfill a list of distance constraints, solving a set of **Lagrange multipliers** in the constrained equations of motion (iterative)





Constraints



Figure 3.9: The three position updates needed for one time step. The dashed line is the old bond of length d, the solid lines are the new bonds. $l = d \cos \theta$ and $p = (2d^2 - l^2)^{\frac{1}{2}}$.

The LINCS algorithm (developed in 1997, twenty years after SHAKE) solves the bond length constraints, always in two steps (non-iterative)





Constraints



constraints:

none

No constraints except for those defined explicitly in the topology, *i.e.* bonds are represented by a harmonic (or other) potential or a Morse potential (depending on the setting of morse) and angles by a harmonic (or other) potential.

h-bonds

Convert the bonds with H-atoms to constraints.

all-bonds

Convert all bonds to constraints.

h-angles

Convert all bonds and additionally the angles that involve H-atoms to bond-constraints.

all-angles

Convert all bonds and angles to bond-constraints.





Timescale









Topology

The topology file describes the atoms composing a molecule and their bond connections Es: flexspc.itp in gromacs



[atoms]

; id	at type	res nr	res nam	ne at na	ame	cg nr cha	arge 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







Constraints in Topology

Only in case of water, the constraint algorithm can be selected in the topology file

[bo	onds]								
; i	j	func	t le:	ngth	forc	e.c.			
1	2	1	(0.1	3450	000			
1	3	1	(0.1	345(000	flexspc.itp		
[an	gles]								
; i	j	k	fun	ct an	gle t	force.c			
2	1	3	1	10	9	383			
[se	ttles]								
; OV	V fu	nct	doh	dhh			spc.itp		
1		1	0.1	0.163	33				
								****	1



PRACE



Topology









Initial velocities



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







To recapitulate..





THE GLOBAL MD ALGORITHM

1. Input initial conditions

Potential interaction V as a function of atom positions Positions r of all atoms in the system Velocities v of all atoms in the system \Downarrow

repeat 2,3,4 for the required number of steps:

2. Compute forces

The force on any atom

$$F_i = -\frac{\partial V}{\partial r_i}$$

is computed by calculating the force between non-bonded atom

pairs:

$$F_i = \sum_j F_i$$

plus the forces due to bonded interactions (which may depend on 1, 2, 3, or 4 atoms), plus restraining and/or external forces.

The potential and kinetic energies and the pressure tensor are

computed. ∥

3. Update configuration

The movement of the atoms is simulated by numerically solving Newton's equations of motion

$$\frac{\frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \frac{\mathbf{F}_i}{m_i}}{\overset{\text{or}}{\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}} = \mathbf{v}_i; \quad \frac{\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t}}{\overset{\text{or}}{\frac{\mathrm{d}\mathbf{v}_i}{\mathrm{d}t}} = \frac{\mathbf{F}_i}{m_i}}$$

 if required: Output step write positions, velocities, energies, temperature, pressure, etc.







Molecular Dynamics ensembles



The method discussed above is appropriate for the microcanonical ensemble: constant N (number of particles), V (volume) and E_T (total energy = E + E_{kin})

Note that if time step is short enough, the system loses/gains no net energy (potential + kinetic) when running MD in the NVE ensemble.

When simulating biological macromolecules, it might be more appropriate to simulate under constant Temperature (T) or constant Pressure (P):

Canonical ensemble: NVT Isothermal-isobaric: NPT





Simulating at constant T: the Berendsen scheme



system Heat bath

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

Berendsen et al. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684 (1984)





Simulating at constant T: the Berendsen scheme



system Heat bath

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

Berendsen et al. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684 (1984)





Simulating at constant P: the Berendsen scheme

system

Pressure bath

Couple the system to a pressure bath:

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

where k_{τ} : isothermal compressibility $\tau_{\rm P}$: coupling constant

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

Berendsen et al. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684 (1984)





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

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



Sample input file of gromacs

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

title	=	Yo	
срр	=	/lib/cpp	
include	=	-I/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 S	SOL
nstlist	=	10	
ns-type	=	grid	
rlist	=	0.8	
coulombtype	=	cut-off	
rcoulomb	=	1.4	
rvdw	=	0.8	







Sample input file of gromacs



tcoupl	=	Berendsen
tc-grps	=	Protein
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



SOL



Conformational sampling









Energy minimization



The potential energy surface of a molecule is defined by only a global minumum and a great number of local minima: i.e. conformations where all the **first derivative** of the potential energy function with respect to the coordinates are **zero** and all **second derivatives** are **non-negative**







The energy minimization algorithms finds the nearest local minimum, i.e. the minimum that can be reached by systematically moving down the steepest local gradient

Usually they cannot found the global minumum

Two algorithms very used in MD codes are

- steepest descent
- conjugate gradient

Both of the first order: they use the first derivative of the energy potential function with respect to the coordinates





The steepest descent uses only the gradient of the potential energy function in the local position to calculate the coordinate movement



The *conjugate gradient* uses also the gradient of the potential energy function in the previous step

It is more accurate in finding the local minumum but it is slower than the steepest descent. Therefore is usually applied after some steps of steepest descent









Starting from A, the *steepest descents* goes through A-B-C (or A-D-F) While the *conjugate gradient*, weighting the A-B and B-C gradients, goes through A-B-O





GROMACS USER MANUAL



steep

A steepest descent algorithm for energy minimization. The maximum step size is **emstep** [nm], the tolerance is **emtol** [kJ mol⁻¹ nm⁻¹].

cg

A conjugate gradient algorithm for energy minimization, the tolerance is **emtol** [kJ $mol^{-1} nm^{-1}$]. CG is more efficient when a steepest descent step is done every once in a while, this is determined by **nstcgsteep**.

emtol: (100.0) [kJ mol $^{-1}$ nm $^{-1}$]

the minimization is converged when the maximum force is smaller than this value

emstep: (0.01) [nm]

initial step-size

nstcgsteep: (1000) [steps]

frequency of performing 1 steepest descent step while doing conjugate gradient energy minimization.





1-4 interactions





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).

