

Molecular Dynamics Principles and Statistical Mechanics

Reference:

1. Wang, Y., McCammon, J.A. (2012). Introduction to Molecular Dynamics: Theory and Applications in Biomolecular Modeling. In: Dokholyan, N. (eds) Computational Modeling of Biological Systems. Biological and Medical Physics, Biomedical Engineering. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-2146-7_1
2. GROMACS documentation

Statistical Mechanics

1. Microstates:

- a. Collection of positions and momenta of all particles in the system
- b. A point in a 6 N-dimension space, where N is the number of particles
- c. The total energy of the system, given by Hamiltonian, is the sum of kinetic and potential energy. Momentum eigenstates are eigenfunctions of the kinetic energy operator, and position energy operator is a function of particle coordinates.

$$\mathcal{H} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + U(\mathbf{r}^N)$$

2. Ensemble

- a. A collection of microstates with same macroscopic properties, which have the same Hamiltonian
- b. The points in phase space form a hypersurface
- c. MD simulation generates the phase space distribution, from which the ensemble average can be calculated

$$\langle A \rangle_{\text{NVT}} = \sum_{\mathbf{r}} A(\mathbf{r}) P(\mathbf{r}) = \sum_{\mathbf{r}} \frac{A(\mathbf{r}) \exp(-\beta V(\mathbf{r}))}{Z_{\text{NVT}}}.$$

$$Z_{\text{NVT}} = \sum_{\mathbf{r}} \exp(-\beta V(\mathbf{r})).$$

where

- d. According to ergodic assumption, the time average of simulation trajectory is equal to the ensemble average
- e. Z_{NVT} cannot be evaluated directly, enhanced sampling methods are required

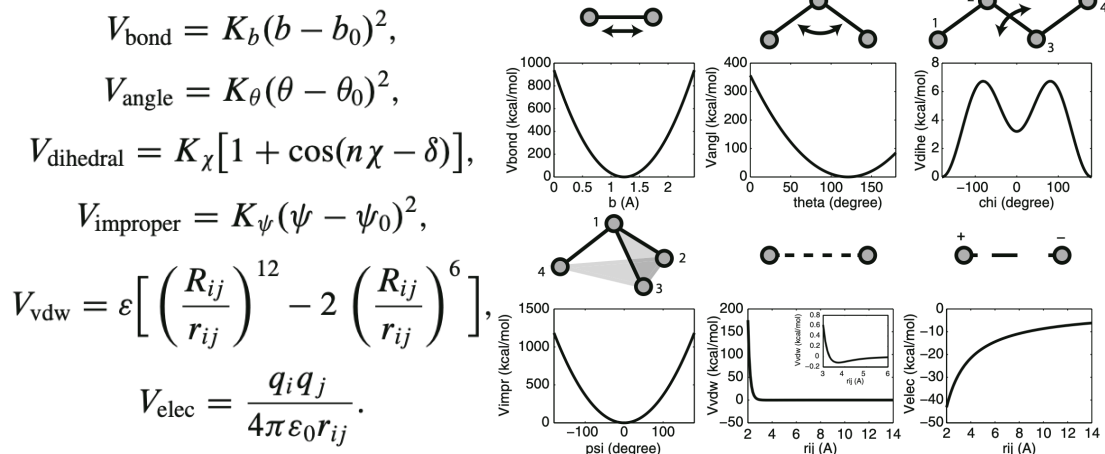
Molecular Dynamics

1. Evaluation of $V(\mathbf{r})$

- a. In a force field, V is broken down into (bonded terms) bond, angle, dihedral, improper, (nonbonded terms) vdW and electrostatic term

b. Examples of force fields: AMBER, CHARMM, GROMOS, OPLS

c. CHARMM



d. Nonbonded terms (vdW, electrostatics) are expensive to calculate

e. A potential is a short-range interaction if it decays faster than r^{-d} , where d is the dimensionality of system

f. Therefore, vdW (r^{-6}) is short-range and electrostatic (r^{-1}) is long-range

g. For short-range vdW, a cutoff distance 8-12Å. Beyond that, a correction term can be added

h. For long-range electrostatic potential, the Ewald summation is used. A neutralising charge distribution for every point charge is introduced, and it decays much faster than r^{-1} , so cutoff scheme can be used

i. Ewald summation requires PBC

2. Periodic boundary condition (PBC)

a. Cubic box containing the original simulation system is replicated throughout space to form an infinite lattice

b. Atoms leaving the box from one side will enter from the opposite side

c. Pro: Eliminates surface effect

d. Cons: Inhibits long-wavelength fluctuations and reduce magnitude of ionic solvation energy

e. Minimum image convention: Particles interact only with the closest periodic image of the other particles

3. Equations of motion

a. Velocity form of Verlet algorithm is commonly used to calculate the time evolution of positions, velocities and forces, where total energy is conserved (NVE ensemble)

$$V_{n+\frac{1}{2}} = V_n + \frac{\Delta t}{2} M^{-1} F_n,$$

$$X_{n+1} = X_n + \Delta t V_{n+\frac{1}{2}},$$

$$F_{n+1} = F(X_{n+1}),$$

$$V_{n+1} = V_{n+\frac{1}{2}} + \frac{\Delta t}{2} M^{-1} F_{n+1}.$$

- b. Newton equations are modified for different ensembles
- c. Timestep is usually 1 fs, corresponding to the vibration of the bond length
- d. To increase the timestep, constraint methods like SHAKE algorithm fix the bond lengths and increase timestep from 1 to 2 fs

Applications

1. Calculation of water diffusion
 - a. Diffusion coefficient of water (D) is dependent on the velocity or mean square displacement of the water molecule
 - b. Velocity form is not used as it's not stored in simulation
 - c. Slope mean square displacement against time is 1/6 of D
 - d. Used to study aquaporins (AQPs, a family of channel protein), which is responsible for conduction of water across lipid membrane. It can compare the water permeability coefficient obtained in experiments
2. Characterisation of receptor flexibility in virtual screening
 - a. In computer-aided drug design (CADD) virtual screening (VS), different compounds are docked into the active site of protein receptor (usually enzyme from a pathogenic organism), i.e. receptor-ligand binding, to identify molecules with high binding affinity, thus inhibitors that can block the active site of enzyme and kill the pathogen
 - b. Protein is usually crystal, i.e. kept rigid, not accounting for protein flexibility, which is the "included fit" mechanism
 - c. Relaxed complex scheme (RCS) accounts for flexibility by generating an ensemble of receptor structures. It uses RMSD or QR factorisation based clustering analysis method to create subset of the ensemble, used in VS. Then subset of structures is extracted from NVT or NPT
 - i. A new binding trench next to original active site of HIV-1 integrase contributed to the discovery of raltegravir (2 ns)
 - ii. A new opening in neuraminidase enzyme from avian influenza virus H5N1 was discovered from ensemble created by RMSD-based clustering, 7

inhibitors are selected using MD structures (40 ns)

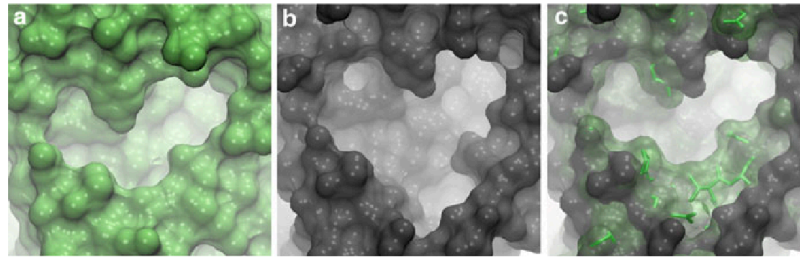


Fig. 5 Receptor flexibility revealed by MD simulations. (a) The neuraminidase active site in the crystal structure. (b) Representative snapshot from a 40 ns MD simulation. (c) The two structures are overlapped to highlight the difference in their active site conformations. Residues from the crystal structure are shown in *green stick* representation. Figure was created using structures from Cheng et al. [24]

Steps in MD Simulation

Choose initial structures	<ul style="list-style-type: none"> - Wild type vs mutant - High resolution
pKa calculation	<ul style="list-style-type: none"> - Determine protonation state, influenced by H-bonds, desolvation effect and Coulombic interactions - MCCE, MEAD, PROPKA, UHBD
Remove crystal water	
{pdb2gmx}	<ul style="list-style-type: none"> - .top (Topology, defines molecule), .itp (Include topology, defines restraints), .gro (post-processed structure, atoms defined in force field)
Define unit cell {editconf}	-
{grompp}	<ul style="list-style-type: none"> - Mdp + gro → top + tpr
Add water and ions (solvate) {genion}	<ul style="list-style-type: none"> - Explicit water - Number of molecules is determined by the size of simulation box - Simulation box has to be large enough so molecules can't see its periodic image, i.e. separation larger than cutoff distance - Need to account for large conformational changes - Layer of water 10-15 Å wide added around protein - Add ions to neutralise net charge of biomolecule - No need to add ions for Ewald summation method as the background charge is already neutralised
Energy minimisation	<ul style="list-style-type: none"> - Remove large steric clashes or close contacts - Check: final PE from log is negative, maximum force is smaller than that defined in mdp, energy converges {energy}
{mdrun}	<ul style="list-style-type: none"> - log (text-log), edr (energy), trr (trajectory), grp (structure)
Equilibration	<ul style="list-style-type: none"> - Allow environment to relax around molecule and bring to desired condition and configuration - Position restraints avoid drastic rearrangements of critical parts, critical for membrane protein - 500 ps for small protein, water travels 30 Å on average

Possible error	<ul style="list-style-type: none"> - No force field parameters for new residue or ligand: check MD package - Transformation matrix to generate oligomer of protein is entered in wrong order → resulting protein monomers are too close: check matrix and pdb file - Atom moving too fast by NAMD, due to repulsive vdW interaction: energy minimisation, slow motion
Analysis	- RMSD, radius of gyration, energy, temperature, pressure

Topology file (.top)

	<pre> ; Include forcefield parameters #include "toppar/forcefield.itp" #include "toppar/PROA.itp" #include "toppar/PROB.itp" #include "toppar/PROC.itp" #include "toppar/POPC.itp" #include "toppar/SOD.itp" #include "toppar/CLA.itp" #include "toppar/TIP3.itp" </pre>
<ul style="list-style-type: none"> • Force field • Molecule name • Atoms information <ul style="list-style-type: none"> • nr: Atom number • type: Atom type • resnr: Amino acid residue number • residue: The amino acid residue name. Note that this residue was "LYS" in the PDB file; the use of .rtp entry "LYSH" indicates that the residue is protonated (the predominant state at neutral pH). • atom: Atom name • cgnr: Charge group number. Charge groups define units of integer charge; they aid in speeding up calculations • charge: Self-explanatory • The "qtot" descriptor is a running total of the charge on the molecule • mass: Also self-explanatory • typeB, chargeB, massB: Used for free energy perturbation • Bonds, pairs, angles, dihedrals information 	

Include topology file (.itp)

<ul style="list-style-type: none"> • Position restraints <pre> [position_restraints] ; The backbone and side chains of the protein are restrained. ; i func fcx fcy fcz 1 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB 5 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB 7 1 POSRES_FC_SC POSRES_FC_SC POSRES_FC_SC 9 1 POSRES_FC_SC POSRES_FC_SC POSRES_FC_SC 13 1 POSRES_FC_SC POSRES_FC_SC POSRES_FC_SC 17 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB 18 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB 19 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB 21 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB 23 1 POSRES_FC_SC POSRES_FC_SC POSRES_FC_SC 26 1 POSRES_FC_SC POSRES_FC_SC POSRES_FC_SC 27 1 POSRES_FC_SC POSRES_FC_SC POSRES_FC_SC 28 1 POSRES_FC_SC POSRES_FC_SC POSRES_FC_SC 29 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB 30 1 POSRES_FC_BB POSRES_FC_BB POSRES_FC_BB </pre> • BB = backbone • SC = side chain 	
---	--

- Restraints for lipids: Position restraints only in z direction (only move in bilayer), two dihedral angles are typically restrained to preserve the cis-isomer

```
#ifdef POSRES
[ position_restraints ]
; Lipid P atoms are restrained in the xy plane:
; i      func      fcx      fcy      fcz
; 20      1      0.0      0.0      POSRES_FC_LIPID
#endif
#ifdef DIHRES
[ dihedral_restraints ]
; 25      36      28      30      1      -120.0      2.5      DIHRES_FC
; 60      63      65      67      1      0.0      0.0      DIHRES_FC
#endif
```

Molecular dynamics parameter (.mdp)

Run

- integrator: steep: or energy minimisation; md: leap-frog method
- dt: Time step
- nsteps: Maximum number of steps to integrate
- emtol: Minimization is converged when the maximum force is smaller than this value
- emstep: Initial step-size
- define: (equilibration) Position restraints for different groups, triggers the inclusion of prose.itp into topology

```
define          = -DPOSRES -DPOSRES_FC_BB=4000.0 -DPOSRES_FC_SC=2000.0 -
DPOSRES_FC_LIPID=1000.0 -DDIHRES -DDIHRES_FC=1000.0
```

Neighbour searching

- cutoff-scheme=Verlet: Interactions are only computed up to a fixed distance and buffer (skin)
- nstlist: Steps before updating list
- rlist: Cut-off distance without dynamics, Verlet buffer is used when there is dynamics

vdW

- vdWtype=Cut-off: Cut-off calculation for van der Waals force
- vdW-modifier=Force-switch: Smoothly switch the vdW to zero from `rvdw_switch` to `rvdw`
- rvdw_switch: Switch the LJ potential
- rvdw: Cut-off distance for LJ potential

Electrostatics

- coulombtype=pme: Fast smooth Particle-Mesh Ewald (SPME) electrostatics
- rcoulomb: Cut-off distance for coulomb force

Bonds

- constraints=h-bonds: Convert bonds with H-atoms to constraints
- constraint_algorithm=LINCS: LINCS is an algorithm that resets bonds to their correct lengths after an unconstrained update. In the first step, the projections of the new bonds on the old bonds are

set to zero. In the second step, a correction is applied for the lengthening of the bonds due to rotation

Temperature coupling (NVT)

- tcoupl = v-rescale: Modified version of Berendsen thermostat, uses velocity rescaling with stochastic term to ensure the correct conical ensemble is generated
- tc_groups: defines group according to the CHARMM-GUI index file
- tau_t: Time constant
- ref_t: Reference temperature

Velocity generation

- gen-vel: Assign velocities from maxwell distribution
- gen-temp: Temperature for maxwell distribiton
- gen-seed: Generate a random speed

Pressure coupling (NPT)

- pcoupl = C-rescale: Modified version of Berendsen thermostat
- pcoupltype = semiisotropic: Pressure coupling which is isotropic in the x and y direction, but different in the z direction. This can be useful for membrane simulations
- tau_p: Time constant
- ref_p: Reference pressure

Advanced simulation techniques

1. Accelerated MD

- a. Sample configurational space in non-Boltamann way
- b. Add bias potential (ΔV) when system PE (V) is lower than threshold (E)

$$\Delta V(\mathbf{r}) = \begin{cases} 0 & V(\mathbf{r}) \geq E \\ \frac{(E-V(\mathbf{r}))^2}{\alpha + E - V(\mathbf{r})} & V(\mathbf{r}) < E. \end{cases}$$

- c. Lower energy barriers between adjacent low-energy states, system can explore configurational space more efficiently
- d. Bias potential removed by

$$\langle A \rangle = \frac{\langle A(\mathbf{r}) \exp(\beta \Delta V(\mathbf{r})) \rangle^*}{\langle \exp(\beta \Delta V(\mathbf{r})) \rangle^*}$$

where * represents the biased ensemble

2. Free-energy perturbation

- a. Difference between Helmholtz or Glibbs free energy A of state 0 and 1

$$\begin{aligned} \Delta A &= -\beta^{-1} \ln \frac{\sum \exp(-\beta V_1)}{\sum \exp(-\beta V_0)} \\ &= -\beta^{-1} \ln \frac{\sum \exp[-\beta (V_1 - V_0)] \exp(-\beta V_0)}{\sum \exp(-\beta V_0)} \\ &= -\beta^{-1} \ln \langle \exp[-\beta (V_1 - V_0)] \rangle_0. \end{aligned}$$

, useful when studing phas transition

- b. System is frozen at a configuration in state 0 and 1 respectively to find the average exponential

$$\Delta A = -\beta^{-1} \ln Z_1 / Z_0; \quad \Delta A = \beta^{-1} \ln \langle \exp(\beta \Delta V) \rangle_1.$$

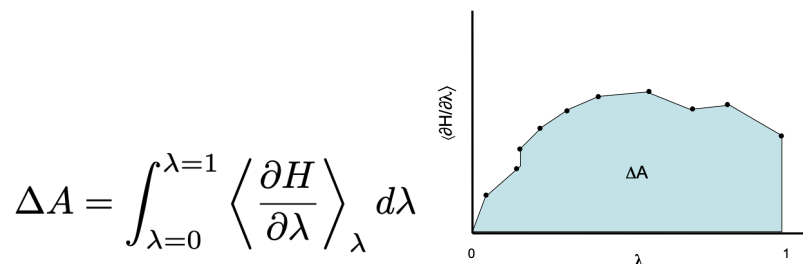
3. Thermodynamic integration

- a. $\frac{dA}{d\lambda} = -\frac{1}{\beta} \frac{1}{Z_{\text{NVT}}} \frac{\partial Z_{\text{NVT}}}{\partial \lambda}$, where lambda is the parameter to describe transition of system from state 0 to 1

$$\frac{dA}{d\lambda} = \frac{\sum \exp(-\beta V) \frac{\partial V}{\partial \lambda}}{\sum \exp(-\beta V)} = \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda}$$

$$\Delta A = \int_{\lambda_0}^{\lambda_1} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

- b. In alchemical transformation where ligand is annihilated, state 0 and 1 are system with or without ligand
c. Hamiltonian can be calculated, and free energy is the area under the curve



4. Umbrella sampling

- a. Divides transition into multiple overlapping windows, uses biasing potential to restrain each system, gives free energy profile using WHAM