ÚOCHB AV ČR
IOCB PRAGUE

RECEPTOR.AI

Univerzita Palackého
v Olomouci

Introduction to Molecular Dynamics

Semen Yesylevskyy

- Receptor.AI LTD
- IOCB Prague
- Palacký University Olomouc

Plan

1. A bit of philosophy and general bla-bla.
2. What is classical MD?
3. Theory of MD without formulas (almost)
4. Which phenomena could be simulated?
5. Example applications of MD in drug discovery

The place of simulations in modern science

Experiment

(+) ground truth data
(-) expensive, complex, error prone, time consuming

Theory

(+) Reveals the most fundamental relations; cheap; compact; predictive.
(-) Too general; complex math; hard to apply to real world; limited area of validity

Simulations

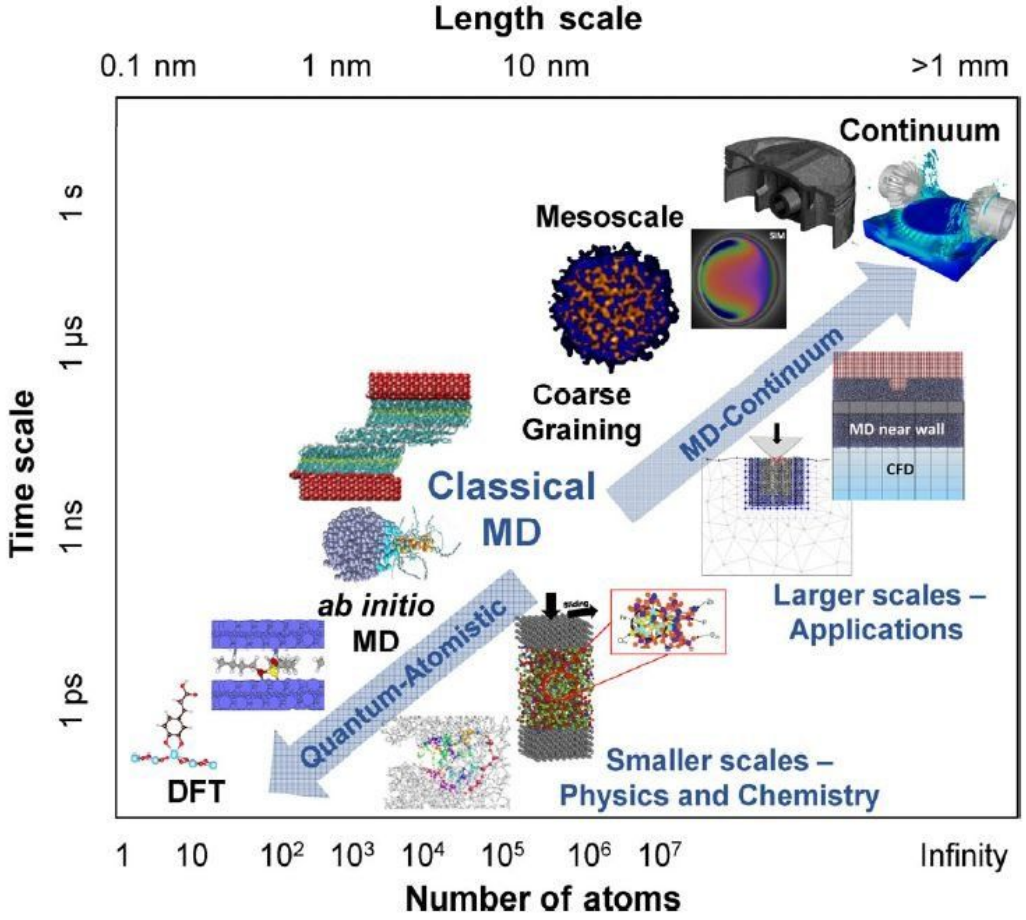
Applies general theory to concrete systems using computer power. Systems are too large or too complex for analytical theory. Theory sets the rules and we compute how the system evolves. It is impossible to predict the result.

Simulations are not theory

- Based on theory
  - Historically it was considered as theory
  - Now it is treated more like experiments: *in vivo*, *in vitro* and *in silico*
- Workflow in simulations is like in experiments:
  - Prepare the system, set all conditions and external parameters
  - Account for side factors and try to eliminate them
  - Let it go and see what happens. In contrast to theory you never know what exactly will happen!
  - Analyze the results and plan the next round of study
- When running simulation is a black box – you don't understand all processes which happen inside.
  - Like in football: you know the rules perfectly, but can't predict the course of the game.

Hierarchy of simulations

- Quantum mechanics (Schrödinger equation)
  - Quantum chemistry
- Classical mechanics of material points (Newton laws)
  - Molecular dynamics (MD)
- Classical mechanics of continuous media (Navier–Stokes equations)
  - Fluid dynamics and aerodynamics
- Electrodinamics (Maxwell's equations)
  - Computational optics and radiophysics



What is classical MD?

- The system is modeled on atomistic level
- Each atom moves according to the Newton's mechanics due to the forces from other atoms
- Equations of motions are integrated numerically by a special software called an MD engine
- Trajectory of each atom in time is recorded and then analyzed and visualized → fancy movies and illustrations
- Atom is a material point with some mass and charge
- No electrons! No orbitals! No quantum effects! ~~No problems!~~

The (hard) live without electrons

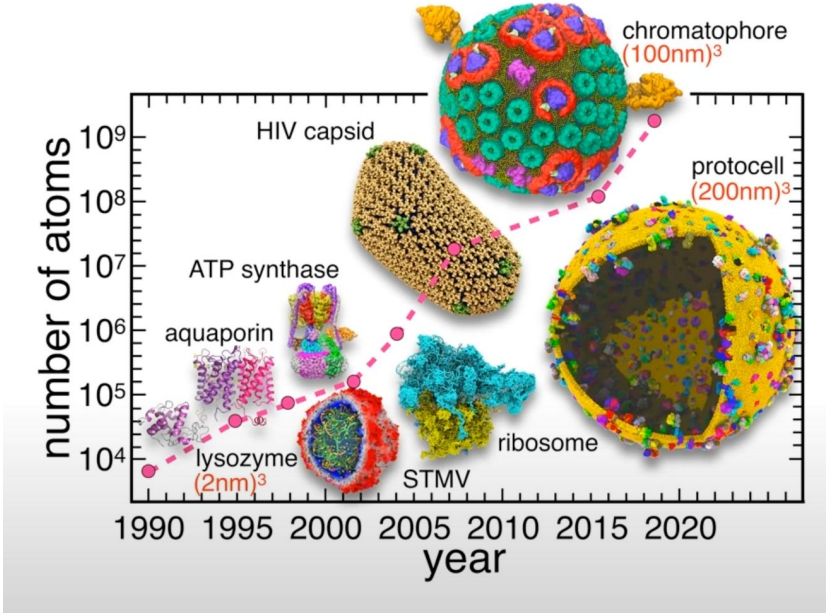
Absence of electrons sets some limitations:

- No chemical reactions involving covalent bonds – the bonds can't be created or destroyed (but ionic and hydrogen bonds are Ok)
- No electronic excitations – can't model photochemistry, fluorescence, etc.
- No charge transfer – atomic charges are fixed
- No funny quantum things like umbrella undulations of ammonium

How long and how large?

Method	System size	Time scale
Quantum chemistry	~100 atoms, 1 nm	~1 ps
Molecular dynamics	~10^6 atoms, 100 nm	~1 μs
Fluid dynamics	> 1 μm up to infinity	any

- MD sits in what is called nanoscopic scale
- Ideal for the systems which are already large on quantum scale but still not macroscopic (biological macromolecules, supramolecular chemistry).



MD is complementary to biological experiments

- Provides atomic resolution
- Allows picking any desired atom or group and seeing all its interactions in details
- All conditions are 100% controllable
- Very easy to play with the system – change atoms, mutate residues in protein, etc.
- Provides quantitative results comparable to experiments (NMR, X-ray, viscosity, diffusion coefficients, etc.)

Theory of MD without formulas (almost)

General MD algorithm

- Classical Newton equations of motion are integrated numerically
  - Different integration schemes exist
- For N atoms a system of 3N differential equations should be integrated
- Explicit exact forces within given cut-off
- Special long-range forces are approximated
- Initial velocities from random distribution
- Special methods for maintaining pressure and temperature

1. Input initial conditions

Potential interaction V as a function of atom positions
Positions $\boldsymbol{r}$ of all atoms in the system
Velocities $\boldsymbol{v}$ of all atoms in the system

⇓

---

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

---

2. Compute forces

The force on any atom

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i}$$

is computed by calculating the force between non-bonded atom pairs:

$$\boldsymbol{F}_i = \sum_j \boldsymbol{F}_{ij}$$

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

$$\frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = \frac{\boldsymbol{F}_i}{m_i}$$

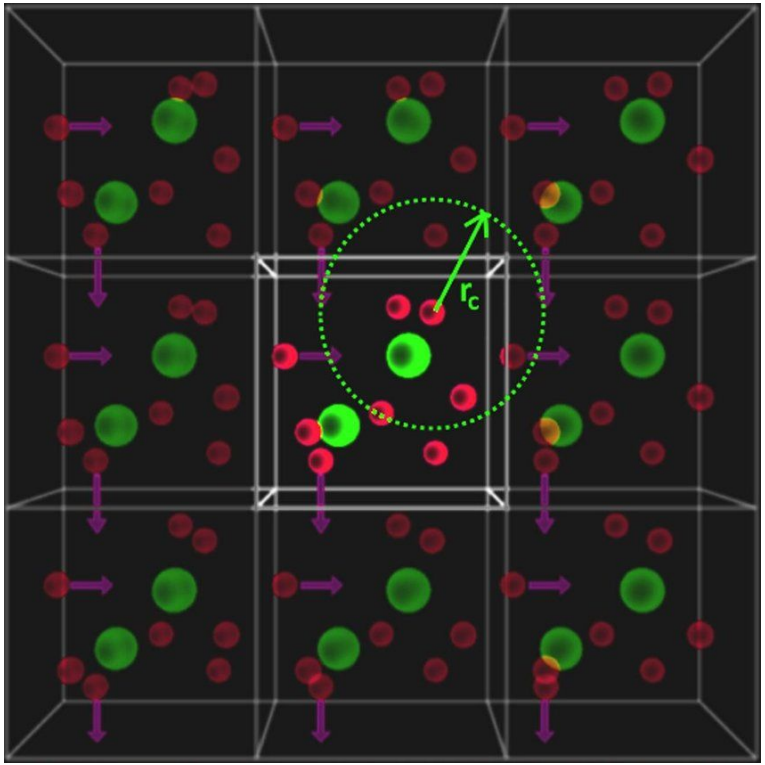
or

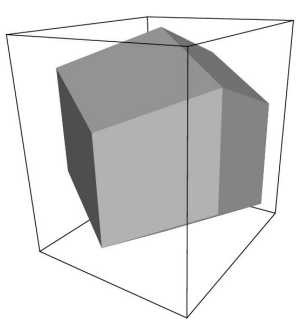
$$\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} = \boldsymbol{v}_i; \quad \frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = \frac{\boldsymbol{F}_i}{m_i}$$

⇓

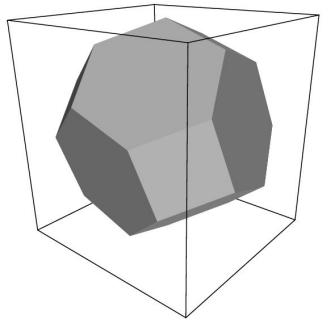
4. if required: **Output step**
write positions, velocities, energies, temperature, pressure, etc.

Periodic boundary conditions





rhombic dodecahedron

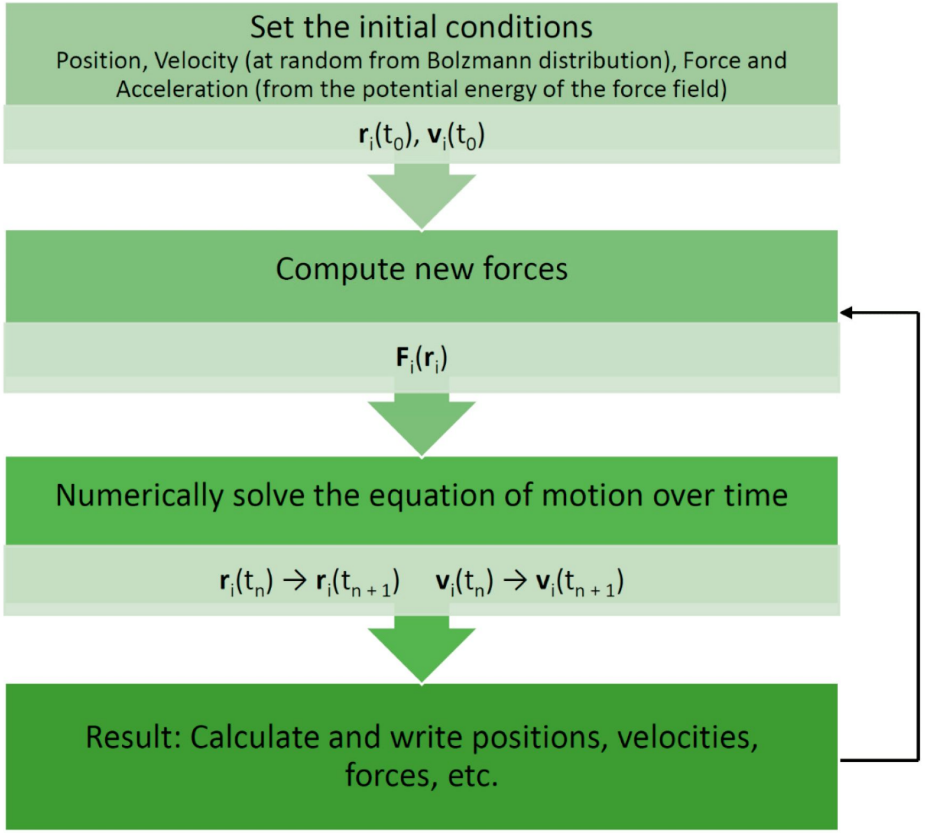


truncated octahedron

- Infinite volume is modelled as periodically replicated box.
- Atoms exiting the box enter it from the other side.
- Minimum image convention
  - Single periodic image of each particle considered
- Box could be rectangular or triclinic

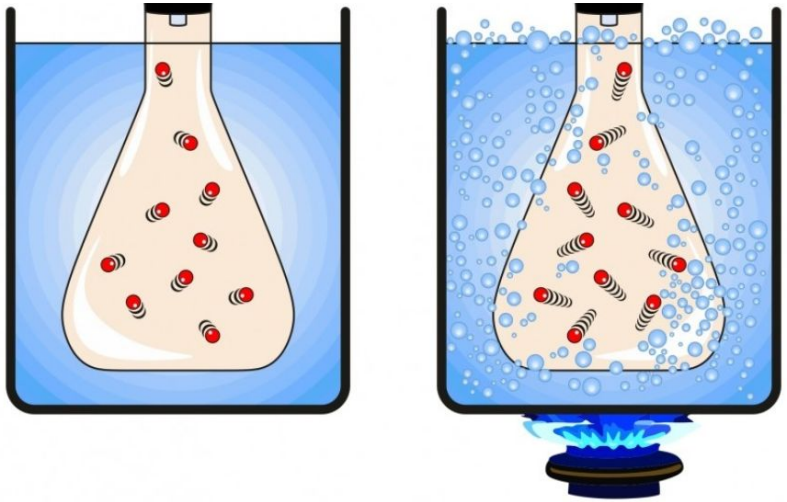
Equations of motion

- Classical Newton equations of motion are integrated numerically
  - Different integration schemes exist
- For N atoms a system of 3N differential equations should be integrated
- MD is among the most computationally expensive techniques in life sciences



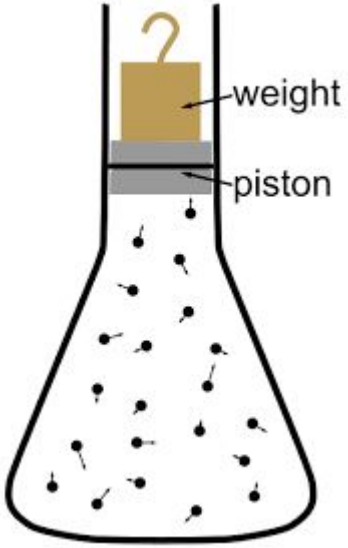
Temperature coupling

- Simulations are not exact due to numerical errors
  - The energy is “drifting” leading to cooling or heating of the system.
- Mitigated by coupling to an external “heat bath” - a thermostat with given temperature.
  - Acts by rescaling velocities so that the average temperature fluctuates around desired value.
  - Many different thermostat algorithms exist, used in different cases.



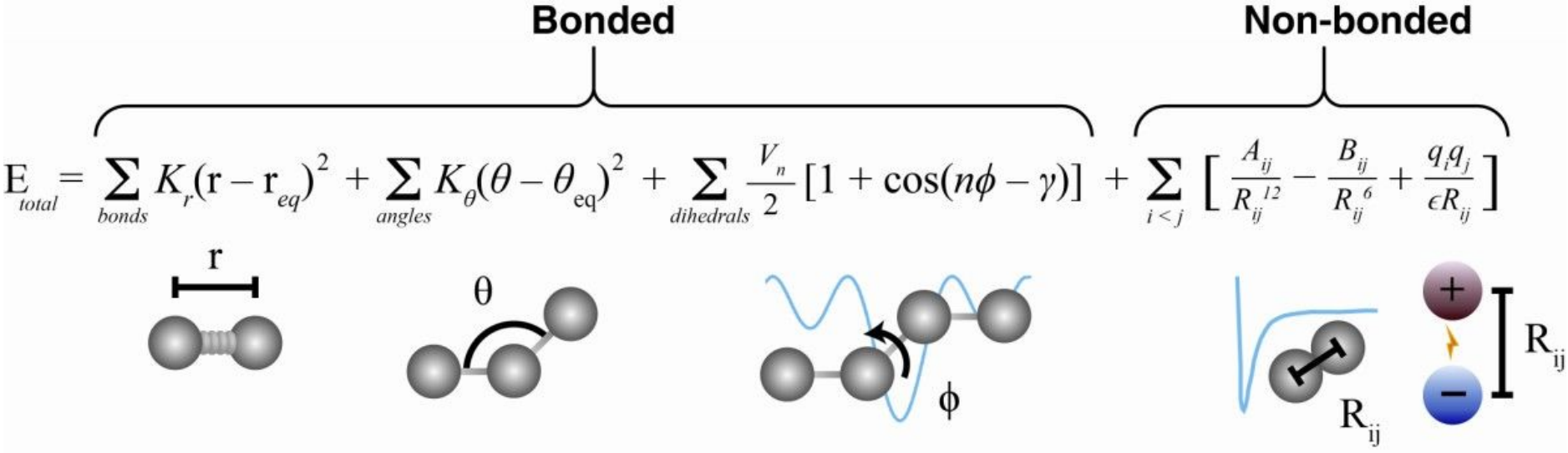
Pressure coupling

- Volume of the simulation box determines which pressure is inside.
- To have a correct pressure the box have to adjusted and must fluctuate according to random fluctuations of velocities.
- This is achieved by coupling the system to external barostat - a virtual piston that acts on the simulation box.
- Pressure could be isotropic, semi-isotropic (for membranes) or anisotropic.
- Different coupling algorithms exist.



Empirical potential functions

- Bonded
  - Bonds
  - Angles
  - Dihedrals
  - Improper dihedrals
- Non-bonded
  - Coulombic
  - Van-der-Waals



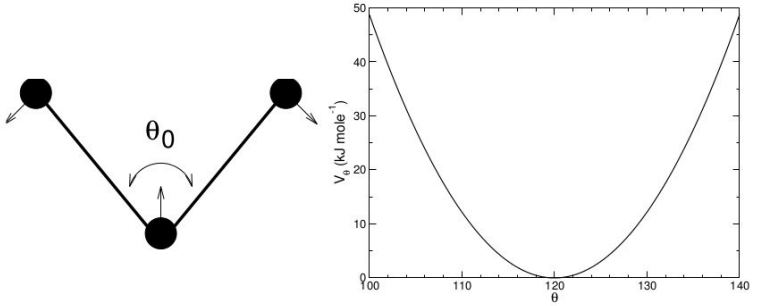
Bonds

- Harmonic bond

- There also other less common types

Angles

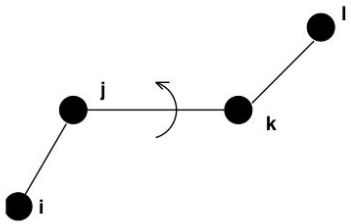
- Harmonic angle

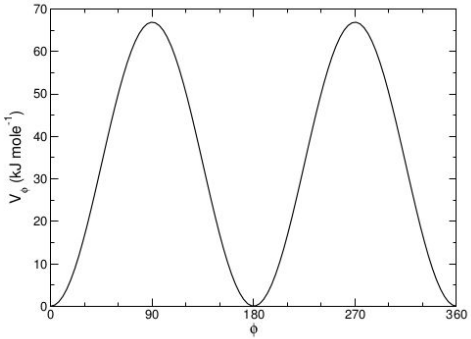


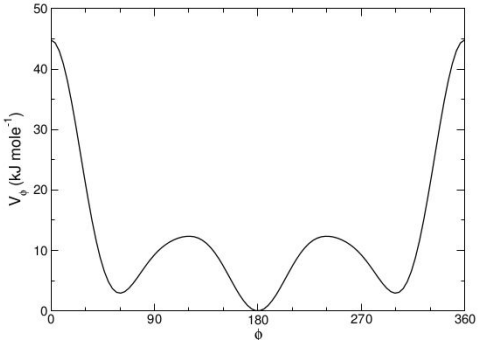
- Urey-Bradley potential

- There are also other less common functional forms

Proper dihedrals (torsion anles)



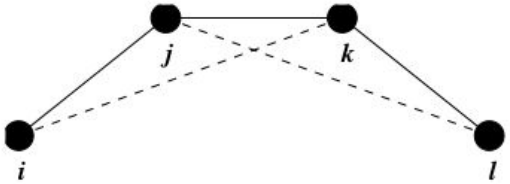


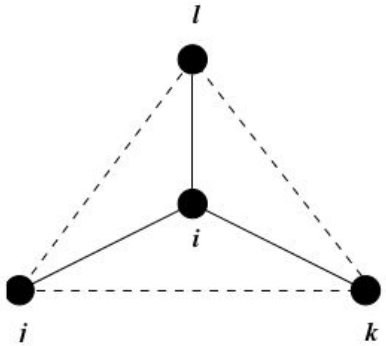


- Simple periodic
- Ryckaert-Bellemans
- One dihedral angle may be presented as a sum of several potentials and thus may have very complex form

Improper dihedrals

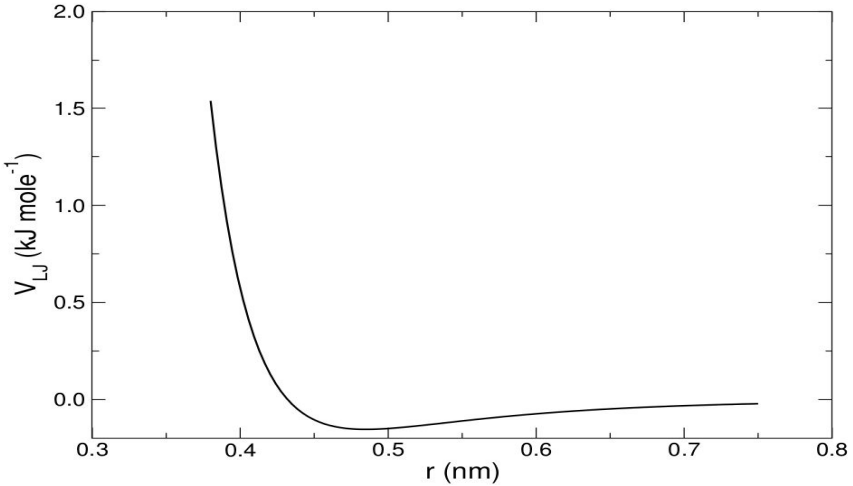
- Used to keep planarity of chemical groups (rings and and other planar geometries)





Van der Waals interactions

Lennard-Jones potential



$$V_{LJ}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}}$$

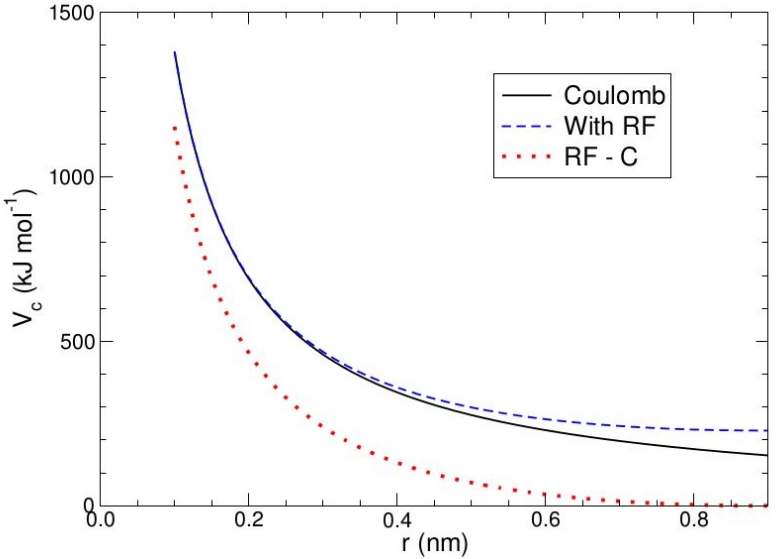
$$V_{LJ}(\boldsymbol{r}_{ij}) = 4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)$$

Combination rules:

$$\begin{aligned} \sigma_{ij} &= \tfrac{1}{2}(\sigma_{ii} + \sigma_{jj}) \\ \epsilon_{ij} &= (\epsilon_{ii}\,\epsilon_{jj})^{1/2} \end{aligned}$$

$$\begin{aligned} \sigma_{ij} &= (\sigma_{ii}\,\sigma_{jj})^{1/2} \\ \epsilon_{ij} &= (\epsilon_{ii}\,\epsilon_{jj})^{1/2} \end{aligned}$$

Electrostatics (short range, explicit)



- Only within the cut-off
- Derivative must be zero at cut-off distance
- Modified by shift of switch functions to ensure this

Electrostatics (long range, epproximate)

- Ewald sum over periodic images

- Special method with summation in reciprocal space to improve convergence and speed → PME (Particle-Mesh Ewald)
- Applied *beyond* the cut-off
- Approximate energy but *much better* than nothing

Force fields

Force field is a set of constants in empirical potentials that determined from experiments, computed or guessed in some way.

Force field is the broad sense also includes:

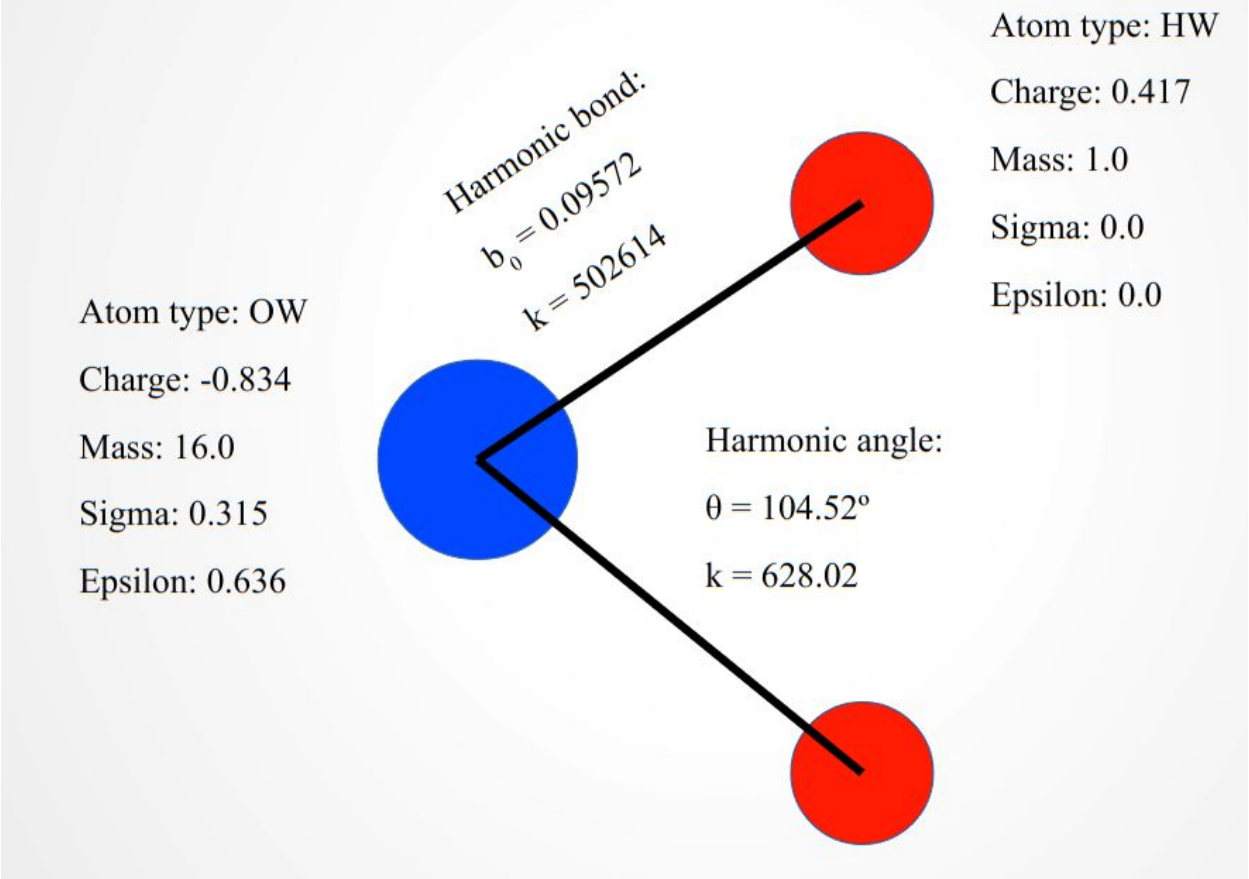
- Functional forms of bonded and non-bonded interactions
- Non-bonded combination rules
- Rules for assigning angles and dihedrals (exhaustive, minimal)
- Pre-defined topologies for standard molecules and building blocks
- Rules for parameterizing new compounds
- Recommended MD parameters

Common FFs: Amber, CHARM, GROMOS, OPLS…

Topology and atom types

- Topology is scheme of molecular connectivity
- Explicitly lists bonds, angles, torsions and improper dihedrals
- Assigns an *atom type* for each atom
  - Atom type is a distinct “chemical identity” of the atom like “carbon in the 6-member ring with non-polar substituent” OR “hydrogen of the hydroxyl group” OR “non-polar carbon in sp3 form”.
  - Also define atom mass, default charge (overridable) and non-bond sigma-epsilon pair
  - Atom types are arbitrary and assigned individually for each force field
  - There are 50-100 atom types in most common force fields
  - You can add your own atom types if needed
  - In some MD programs (Gromacs) you can also override atom properties on top of atom types.

Example topology: water in Amber FF



Example topology: water in Amber FF

```
[ moleculetype ]
; molname       nrexcl
SOL         2

[ atoms ]
; id  at type     res nr  res name  at name  cg nr  charge    mass
  1   OW         1       SOL       OW       1      -0.834    16.00000
  2   HW         1       SOL       HW1      1       0.417     1.00800
  3   HW         1       SOL       HW2      1       0.417     1.00800

[ bonds ]
; i     j       funct   length  force_constant
1       2       1       0.09572 502416.0   0.09572        502416.0
1       3       1       0.09572 502416.0   0.09572        502416.0

[ angles ]
; i     j       k       funct   angle   force_constant
2       1       3       1       104.52  628.02      104.52  628.02
```

MD packages

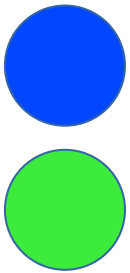
There are several commonly used MD packages for different domains

- For biological and organic molecules:
  - Gromacs
  - NAMD
  - OpenMM
- For solid state physics and material science:
  - Dl_poly
  - LAMMPS
- All modern MD packages are parallelized and GPU-accelerated
- Marvels of software optimization but usually examples of *bad* software engineering

Combination rules

- If you have N atom types there are N^2 parameters for non-bonded interactions
  - In practice N=50, N^2=2500
  - This is way too many for any kind of rational parametrization procedure
  - We need to decrease this number somehow
- Combination rule is a simple arithmetics to deduce non-bonded parameters for interaction A-B from properties of atoms A and B.
- Example:

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \qquad \sigma_{ij} = (\sigma_{ii}\,\sigma_{jj})^{1/2}$$
$$\epsilon_{ij} = (\epsilon_{ii}\,\epsilon_{jj})^{1/2} \qquad \epsilon_{ij} = (\epsilon_{ii}\,\epsilon_{jj})^{1/2}$$



Atom type: OW
σ_1 = 0.315
ε_1 = 0.636

Atom type: CA
σ_2 = 0.339
ε_2 = 0.359

σ_{12} = (0.315+0.339)/2 = 0.327
ε_{12} = (0.636*0.359)1/2 = 0.477

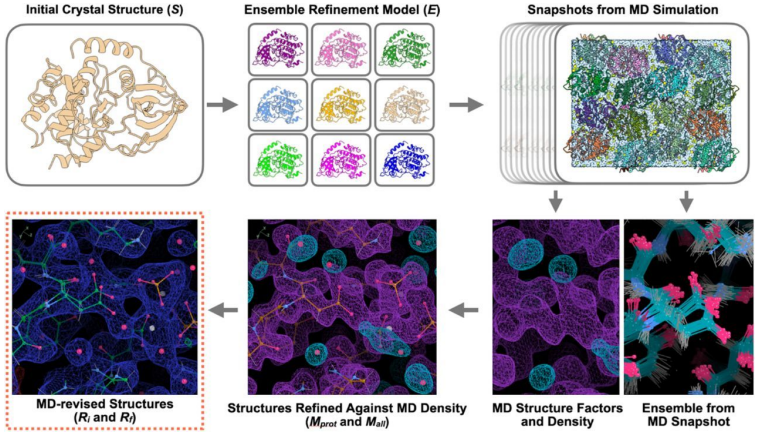
There is much-much more...

- Coarse-grained MD
- Non-equilibrium MD and external potentials
- Hybrid QM/MM
- Enhanced sampling techniques and collective coordinates
- ML potentials
- Multiple time stepping
- Non-standard topology: virtual atoms, restraints, heavy hydrogen...
- Special purpose techniques: stochastic dynamics, brownian dynamics, normal modes, free energy estimation, replica exchange, essential dynamics, expanded ensembles...
- Polarizable FFs, dispersion correction
- Trajectory analysis and visualization
- ...

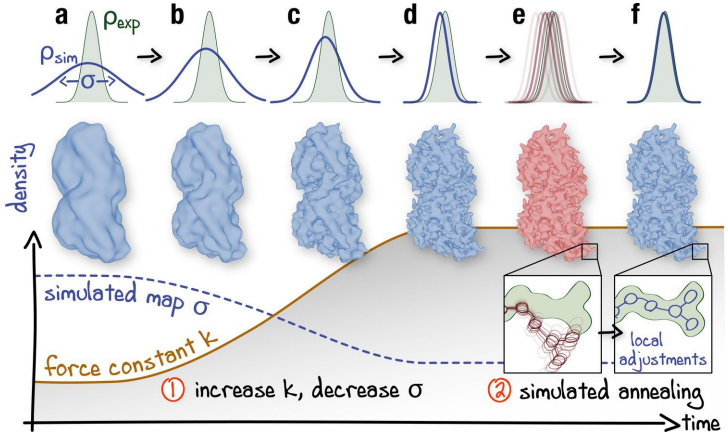
Applications of MD in drug discovery

Protein structure prediction

- X-ray and CryoEM structure refinement
- Relaxation and stability assessment after homology modeling
- AlphaFold-like models use MD energy minimization on the last step of atomic coordinate generation to remove steric clashes



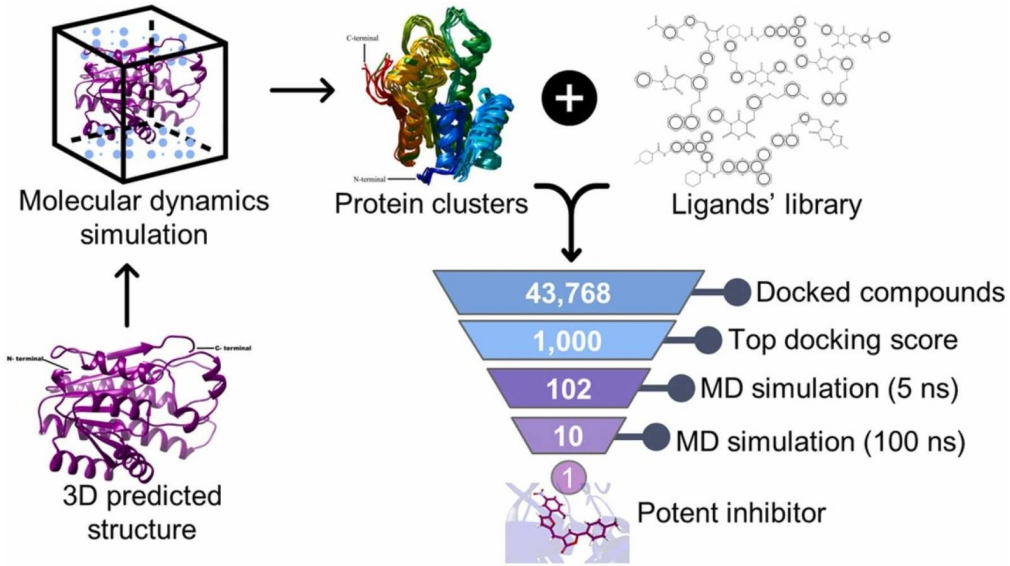
https://www.biorxiv.org/content/10.1101/2022.04.04.486986v1.full-text



https://elifesciences.org/articles/43542

Generating ensembles of conformations for virtual screening

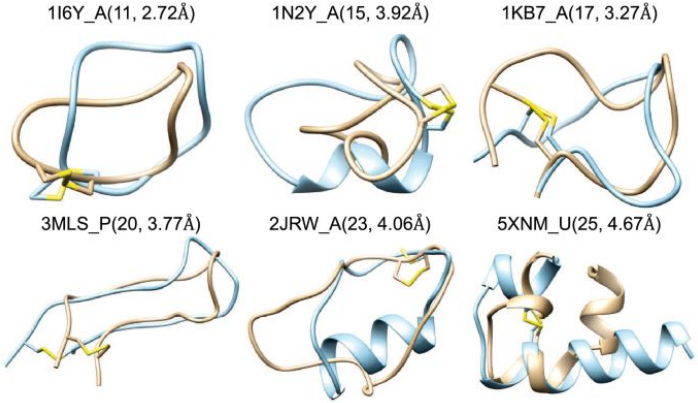
- X-ray or CryoEM structure is almost never what actually works in the cell (due to artificial environment, harsh experimental conditions)
- MD allows obtaining set of representative dynamic structures - an ensemble of “real” conformations.
- VS to the ensemble is always better than VS to crystal but more resource intensive.
- After docking another round of MD can be performed *with* docked ligands to assess their stability and pose correctness.



https://www.sciencedirect.com/science/article/pii/S1476927123000026

Exploring conformations of peptide binders

- Peptide drugs is on the steep rise in recent years.
- Major problem with peptides - they are very flexible.
- Cyclopeptides are used to reduce this flexibility, but still to many possible rotamers.
- MD is often used to explore the peptide conformational state and to select the most relevant clusters for VS and peptide-protein docking.

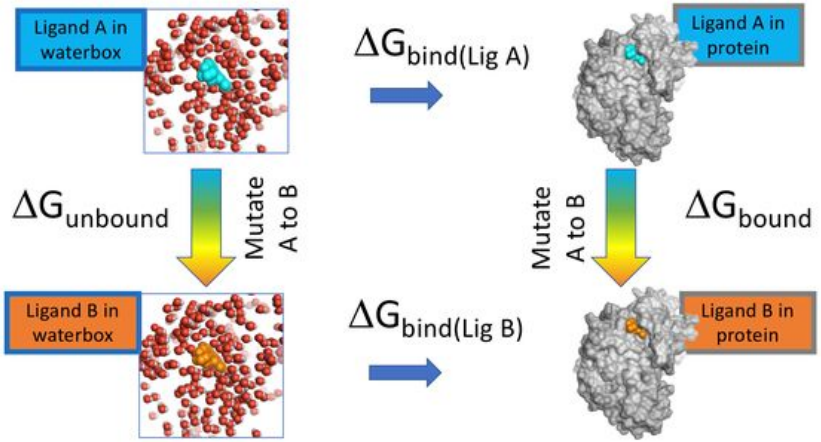


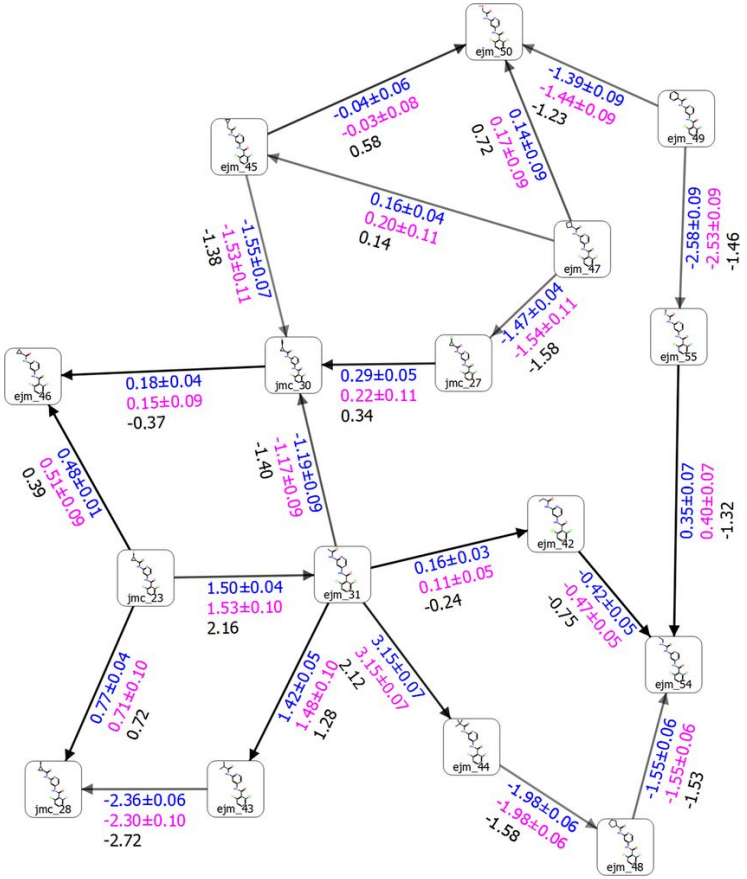
https://jcheminf.biomedcentral.com/articles/10.1186/s13321-022-00605-8

https://jcheminf.biomedcentral.com/articles/10.1186/s13321-022-00605-8

Free Energy Perturbation (FEP) for lead optimization

- FEP allows computing:
  - Binding free energy of ligand binding to target protein.
  - *Differences* in the free energies of binding between similar ligands.
- Works by “alchemically” mutating one molecular topology to the other in place in the course of MD using special simulation protocols and algorithms.
- Works best with series of similar compounds → usage on the lead optimization steps.
- Very resource intensive, fragile and hard to set up manually.





Conclusions

- MD is large and mature field in life sciences, which become a separate discipline by its own.
- It has multiple applications in drug discovery and related fields.
- Understanding its capabilities and limitations is critical for successful career in computational drug discovery in both academia and industry.
- Reading tutorial doesn’t make you an MD expert :) Practical hands on experience is required to “feel” it.