Chimica Computazionale

Molecular Mechanics and Dynamics

Emanuele Coccia

UNIVERSITÀ DEGLI STUDI DI TRIESTE

DSCF Dipartimento di Scienze Chimiche e Farmaceutiche

PhotoInduced Quantum Dynamics (PIQD) Group

PIQD

Molecular mechanics

Molecular mechanics (MM)

QUANTUM CHEMISTRY

PROS

- (Possible) high accuracy
- Explicit electronic structure
- First-principle approach

CONS

- Large computational effort
- Not applicable to (very) large systems

MOLECULAR MECHANICS

PROS

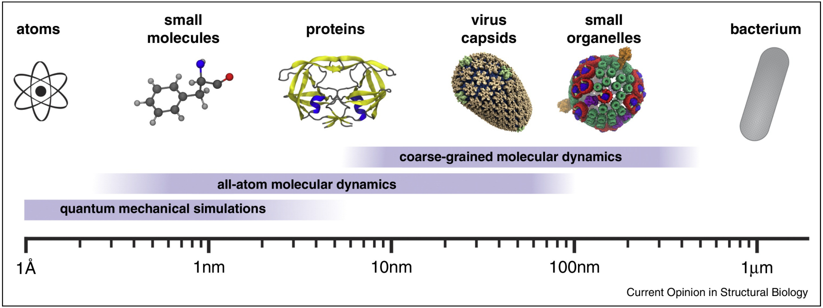
- Less time-consuming
- Chemical bond pattern assumed
- Applied to liquids, solids and complex biological systems

CONS

- Chemical reactivity missing
- No electrons described
- Chemical bond pattern assumed

Simulations vs system size

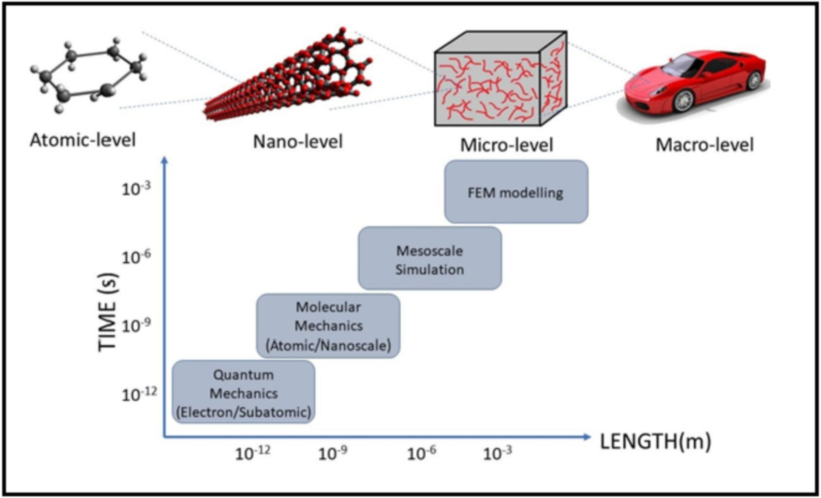
Towards biological systems



Current Opinion in Structural Biology

Simulations vs system size

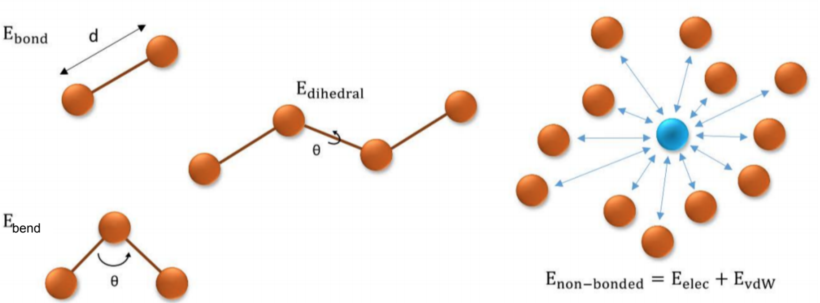
Towards macroscopic materials



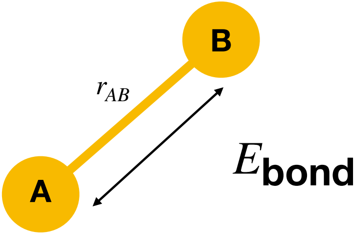
Force field in MM

- Set of equations to compute energy and force using MM
- Functions of nuclear positions (classical!)
- Parametrized set of equations

$$\begin{aligned} E_{\mathrm{FF}} &= E_{\mathrm{bonded}} + E_{\mathrm{nonbonded}} \\ E_{\mathrm{bonded}} &= E_{\mathrm{bond}} + E_{\mathrm{bend}} + E_{\mathrm{dihedral}} \\ E_{\mathrm{nonbonded}} &= E_{\mathrm{elec}} + E_{\mathrm{vdW}} \end{aligned}$$



Bond stretching

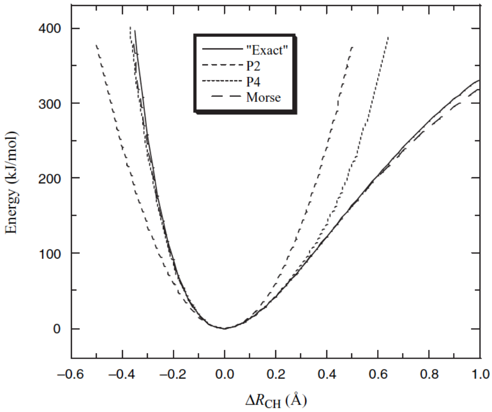


- Taylor expansion around the "equilibrium" bond length r_0

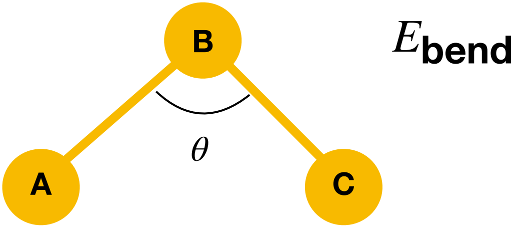
$$\begin{aligned} E_{\text{bond}} &= k_{AB}(r_{AB} - r_0)^2 = k_{AB}\Delta r^2 \quad \text{(P2)} \\ E_{\text{bond}} &= k_{AB,2}\Delta r^2 + k_{AB,3}\Delta r^3 + k_{AB,4}\Delta r^4 \quad \text{(P4)} \end{aligned}$$

Bond stretching

- Stretching of the C-H bond in CH_4
- "Exact": quantum-chemistry calculations
- Morse: $E_{\text{Morse}} = D(1 - e^{-\alpha\Delta r})^2$, $\alpha = \sqrt{k/(2D)}$
- P2: simple harmonic approximation
- P4: P2 + quartic term (and the cubic one)



Molecular bending

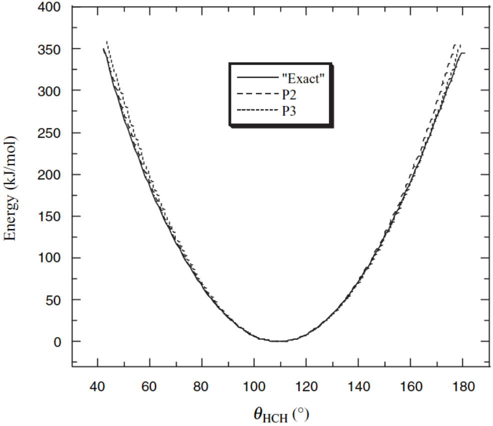


- E_{bend}: energy required for bending an angle formed by three atoms A-B-C
- Taylor expansion around the "equilibrium" angle value θ_0

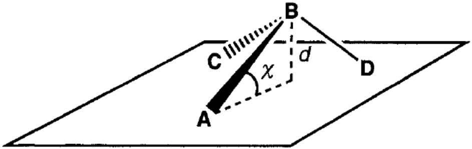
$$
\begin{aligned}
E_{\text{bend}} &= k_{ABC}(\theta_{ABC} - \theta_0)^2 \quad (\text{P2}) \\
E_{\text{bend}} &= k_{ABC,2}(\theta_{ABC} - \theta_0)^2 + k_{ABC,3}(\theta_{ABC} - \theta_0)^3 \quad (\text{P3})
\end{aligned}
$$

Molecular bending

- Bending energy for CH_4
- "Exact": quantum-chemistry calculations
- P2: simple harmonic approximation
- P3: P2 + cubic term



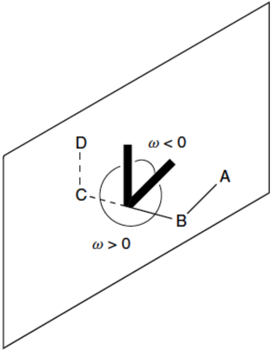
Out-of-plane bending



- Central atom B connected to the 3 peripheral atoms A, C and D
- E_{oop} used to enforce planarity
- Also dubbed as improper torsion
- Example of formaldehyde CH_2O

$$\begin{aligned} E_{\text{oop}} &= k_B \chi^2 \\ E_{\text{oop}} &= k'_B d^2 \end{aligned}$$

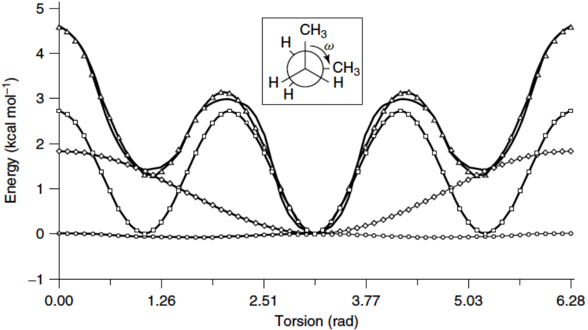
Molecular torsion



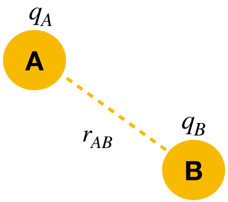
- Four atoms connected in sequence (ABCD)
- The torsional angle ω defined as the angle between AB and CD bonds, when projected into the plane bisecting the BC bond
- ω is periodic and defines the stiffness

Molecular torsion

$$
\begin{aligned}
E_{\text{dihedral}} &= \frac{1}{2}\sum_n A_n\left[1+(-1)^{n+1}\cos(n\omega)\right] \quad (\triangle)\\
&= \frac{1}{2}A_1\left[1+\cos(\omega)\right] \quad (\diamond)\\
&+ \frac{1}{2}A_2\left[1-\cos(2\omega)\right] \quad (\circ)\\
&+ \frac{1}{2}A_3\left[1+\cos(3\omega)\right]+\cdots(\square)
\end{aligned}
$$



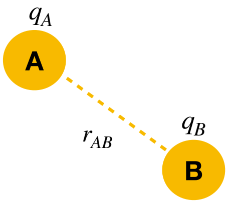
Electrostatic interaction



$$E_{\textbf{elec}} = \frac{1}{\epsilon}\frac{q_A q_B}{r_{AB}}$$

- Partial charges assigned to each atom
- Assigned by fitting to the electrostatic potential calculated by quantum chemistry

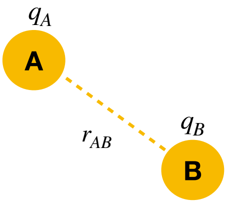
Electrostatic interaction



$$E_{\textbf{elec}} = \frac{1}{\epsilon}\frac{q_A q_B}{r_{AB}}$$

- Partial charges assigned to each atom
- Assigned by fitting to the electrostatic potential calculated by quantum chemistry
- Effective dielectric constant ϵ: $\epsilon = 1$ for vacuum

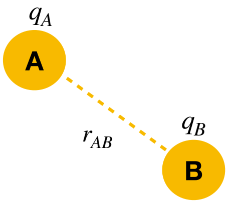
Electrostatic interaction



$$E_{\textbf{elec}} = \frac{1}{\epsilon}\frac{q_A q_B}{r_{AB}}$$

- Partial charges assigned to each atom
- Assigned by fitting to the electrostatic potential calculated by quantum chemistry
- Effective dielectric constant ϵ: $\epsilon = 1$ for vacuum
- E_{elec} for atom pairs separated by three bonds or more

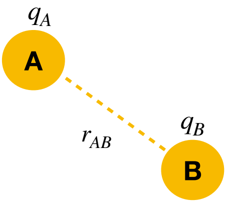
Electrostatic interaction



$$E_{\textbf{elec}} = \frac{1}{\epsilon}\frac{q_A q_B}{r_{AB}}$$

- Partial charges assigned to each atom
- Assigned by fitting to the electrostatic potential calculated by quantum chemistry
- Effective dielectric constant ϵ: $\epsilon = 1$ for vacuum
- E_{elec} for atom pairs separated by three bonds or more
- Pairwise interaction

Electrostatic interaction



$$E_{\textbf{elec}} = \frac{1}{\epsilon} \frac{q_A q_B}{r_{AB}}$$

- Partial charges assigned to each atom
- Assigned by fitting to the electrostatic potential calculated by quantum chemistry
- Effective dielectric constant ϵ: $\epsilon = 1$ for vacuum
- E_{elec} for atom pairs separated by three bonds or more
- Pairwise interaction
- Improvements:
  - Including dipoles, quadrupoles etc.
  - Beyond fixed charges: polarizable force fields

Fitting charges

- Electrostatic potential $\phi_{\text{esp}}(\mathbf{r})$

$$\phi_{\text{esp}}(\mathbf{r}) = \sum_{k}^{K} \frac{Z_k}{|\mathbf{R}_k - \mathbf{r}|} - \int \frac{\rho(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}'$$

- $\phi_{\text{esp}}(\mathbf{r})$ from quantum chemistry
- Fitting to get the charges $\mathbf{q} = \{q_1, q_2, ..., q_k, ..., q_K\}$
- Minimize $\text{ErrF}(\mathbf{q})$

$$\text{ErrF}(\mathbf{q}) = \sum_{i}^{N_{\text{points}}} \left(\phi_{\text{esp}}(\mathbf{r}_i) - \sum_{k}^{K} \frac{q_k(\mathbf{R}_k)}{|\mathbf{R}_k - \mathbf{r}_i|} \right)^2$$

van der Waals interaction

- E_{vdW}: van der Waals energy describing the repulsion or attraction between atoms not directly bonded
  - Repulsion: overlap of the electron clouds of the two atoms
  - Attraction: instantaneous dipole-dipole interaction ($\propto r_{AB}^{-6}$)

van der Waals interaction

- E_{vdW}: van der Waals energy describing the repulsion or attraction between atoms not directly bonded
  - Repulsion: overlap of the electron clouds of the two atoms
  - Attraction: instantaneous dipole-dipole interaction ($\propto r_{AB}^{-6}$)
- General functional form

$$E_{\mathrm{vdW}} = E_{\mathrm{repulsion}} - \frac{C_{AB}}{r_{AB}^{6}}$$

van der Waals interaction

- E_{vdW}: van der Waals energy describing the repulsion or attraction between atoms not directly bonded
  - Repulsion: overlap of the electron clouds of the two atoms
  - Attraction: instantaneous dipole-dipole interaction ($\propto r_{AB}^{-6}$)
- General functional form

$$E_{\text{vdW}} = E_{\text{repulsion}} - \frac{C_{AB}}{r_{AB}^{6}}$$

- Lennard-Jones potential

$$E_{\text{vdW}} = 4\epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^{6} \right]$$

van der Waals interaction

- E_{vdW}: van der Waals energy describing the repulsion or attraction between atoms not directly bonded
  - Repulsion: overlap of the electron clouds of the two atoms
  - Attraction: instantaneous dipole-dipole interaction ($\propto r_{AB}^{-6}$)
- General functional form

$$E_{\text{vdW}} = E_{\text{repulsion}} - \frac{C_{AB}}{r_{AB}^6}$$

- Lennard-Jones potential

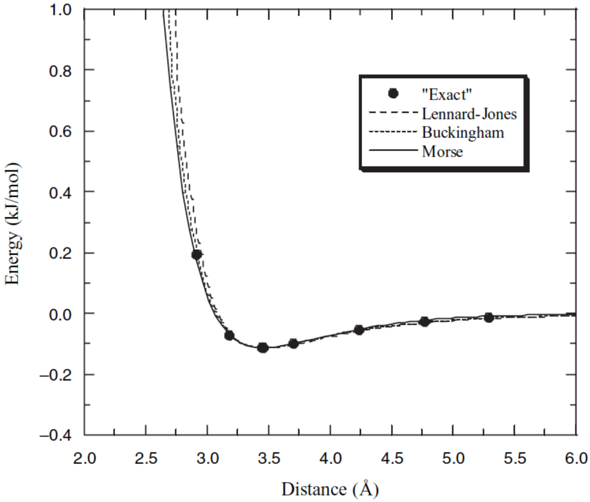
$$E_{\text{vdW}} = 4\epsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right]$$

- Buckingham potential

$$E_{\text{vdW}} = A_{AB}\exp[-B_{AB}r_{AB}] - \frac{C_{AB}}{r_{AB}^6}$$

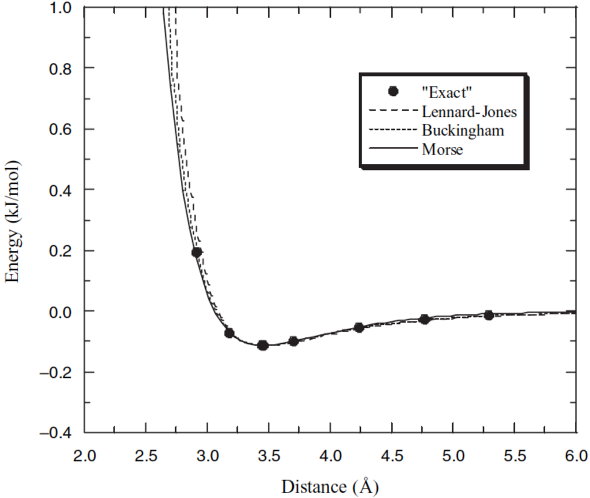
van der Waals interaction

H_2-He interaction



van der Waals interaction

H_2-He interaction

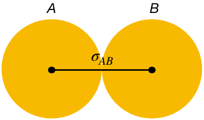


He_2 exists! (average bond length of 55 Å, 0.09 kJ mol^{-1})

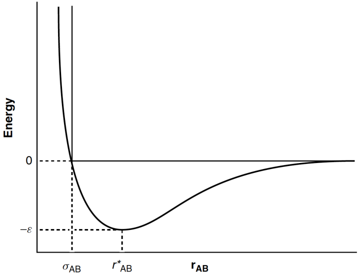
Lennard-Jones potential

- ϵ_{AB}: well depth for E_{vdW}
- σ_{AB}: distance for $E_{\mathrm{vdW}} = 0$

$$\begin{aligned} \epsilon_{AB} &= \sqrt{\epsilon_A \epsilon_B} \\ \sigma_{AB} &= \sigma_A + \sigma_B \end{aligned}$$



- $r^*_{AB} = 2^{1/6}\sigma_{AB}$: distance of the minimum E_{vdW}



Force-field parametrization

- Parameter values from:
    - Experimental data
    - Quantum-chemistry calculations

Force-field parametrization

- Parameter values from:
  - Experimental data
  - Quantum-chemistry calculations
- Experimental data:
  - X-ray crystallography (equilibrium bond lengths and angles)
  - Infrared spectroscopy (force constants)
  - Bulk properties such as density (non-bonded terms)

Force-field parametrization

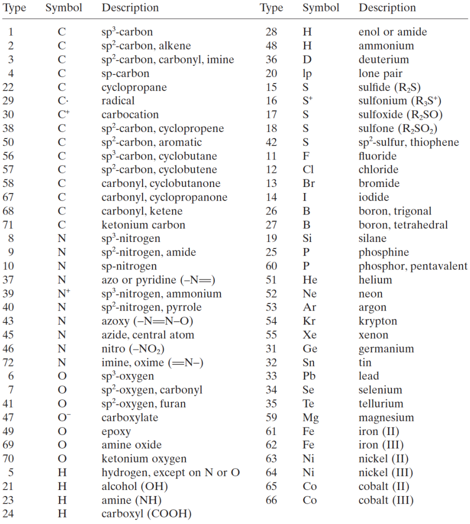
- Parameter values from:
  - Experimental data
  - Quantum-chemistry calculations
- Experimental data:
  - X-ray crystallography (equilibrium bond lengths and angles)
  - Infrared spectroscopy (force constants)
  - Bulk properties such as density (non-bonded terms)
  - Minimizing a penalty function Z

$$Z = \left[\sum_{i}^{\text{Observables}} \sum_{j}^{\text{Occurrences}} \frac{(\text{calc}_{i,j} - \text{expt}_{i,j})^2}{\omega_i^2} \right]^{1/2}$$

  - Parameters selected minimizing Z
- Generality versus accuracy

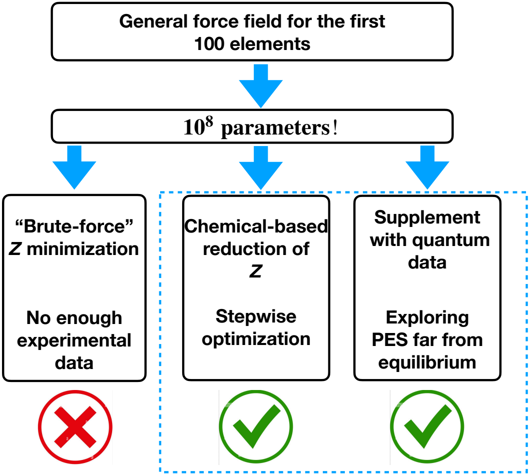
Force-field parametrization

- Atom types: atoms sharing the same expected bonding and interaction properties



Type	Symbol	Description	Type	Symbol	Description
1	C	sp^3-carbon	28	H	enol or amide
2	C	sp^2-carbon, alkene	48	H	ammonium
3	C	sp^2-carbon, carbonyl, imine	36	D	deuterium
4	C	sp-carbon	20	lp	lone pair
22	C	cyclopropane	15	S	sulfide (R_2S)
29	C·	radical	16	S^+	sulfonium (R_3S^+)
30	C^+	carbocation	17	S	sulfoxide (R_2SO)
38	C	sp^2-carbon, cyclopropene	18	S	sulfone (R_2SO_2)
50	C	sp^2-carbon, aromatic	42	S	sp^2-sulfur, thiophene
56	C	sp^3-carbon, cyclobutane	11	F	fluoride
57	C	sp^2-carbon, cyclobutene	12	Cl	chloride
58	C	carbonyl, cyclobutanone	13	Br	bromide
67	C	carbonyl, cyclopropanone	14	I	iodide
68	C	carbonyl, ketene	26	B	boron, trigonal
71	C	ketonium carbon	27	B	boron, tetrahedral
8	N	sp^3-nitrogen	19	Si	silane
9	N	sp^2-nitrogen, amide	25	P	phosphine
10	N	sp-nitrogen	60	P	phosphor, pentavalent
37	N	azo or pyridine (–N=)	51	He	helium
39	N^+	sp^3-nitrogen, ammonium	52	Ne	neon
40	N	sp^2-nitrogen, pyrrole	53	Ar	argon
43	N	azoxy (–N=N–O)	54	Kr	krypton
45	N	azide, central atom	55	Xe	xenon
46	N	nitro ($-NO_2$)	31	Ge	germanium
72	N	imine, oxime (=N–)	32	Sn	tin
6	O	sp^3-oxygen	33	Pb	lead
7	O	sp^2-oxygen, carbonyl	34	Se	selenium
41	O	sp^2-oxygen, furan	35	Te	tellurium
47	O^-	carboxylate	59	Mg	magnesium
49	O	epoxy	61	Fe	iron (II)
69	O	amine oxide	62	Fe	iron (III)
70	O	ketonium oxygen	63	Ni	nickel (II)
5	H	hydrogen, except on N or O	64	Ni	nickel (III)
21	H	alcohol (OH)	65	Co	cobalt (II)
23	H	amine (NH)	66	Co	cobalt (III)
24	H	carboxyl (COOH)			

Force-field parametrization



General force field for the first 100 elements

10^8 parameters!

"Brute-force" *Z* minimization

No enough experimental data

✗

Chemical-based reduction of *Z*

Stepwise optimization

✓

Supplement with quantum data

Exploring PES far from equilibrium

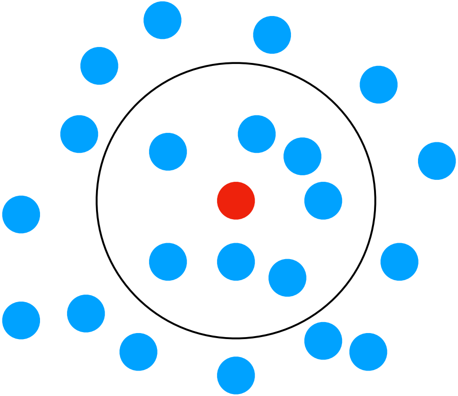
✓

- Specific force fields for liquid water (TIP3P)
- General biological force fields (AMBER, CHARMM)

Periodic systems and cut-offs

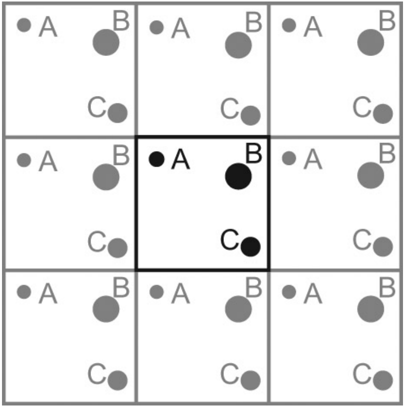
- Non-bonded (vdW) interactions: N atoms and $\approx N^2/2$ terms
- Many of pairwise interaction are negligible

$$V_{\text{cutoff}}(r) = \begin{cases} V(r) - V(r_{\text{cutoff}}) & \text{if } r \leq r_{\text{cutoff}} \\ 0 & \text{otherwise} \end{cases}$$



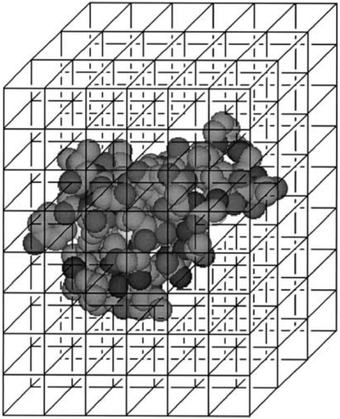
Periodic systems and cut-offs

- Electrostatic interactions: larger than VdW ones for a given r_{cutoff}
- Alternative technique: periodic boundary conditions
    - Modelling an infinitely sized system
    - $(x, y, z) \rightarrow (x + n_x L, y + n_y L, z + n_z L)$



Docking

- Prediction of strength and specificity with which molecule may bind to a biological macromolecule
- Rigid structures
- Only nonbonded interactions



Molecular dynamics

Phase space

- Classical system: its state completely given by coordinates and momenta of the N particles
- $6N$-dimensional space: phase space
- One point in the phase space at given time

$$\mathbf{X} = (x_1, y_1, z_1, p_{x1}, p_{y1}, p_{z1}, x_2, y_2, z_2, p_{x2}, p_{y2}, p_{z2}, \ldots)$$

Phase space

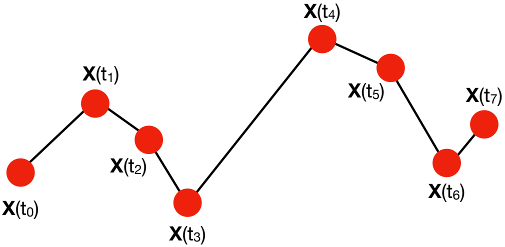
- Classical system: its state completely given by coordinates and momenta of the N particles
- $6N$-dimensional space: phase space
- One point in the phase space at given time

$$\begin{aligned}
\mathbf{X} &= (x_1, y_1, z_1, p_{x1}, p_{y1}, p_{z1}, x_2, y_2, z_2, p_{x2}, p_{y2}, p_{z2}, \ldots) \\
\mathbf{q} &= (x_1, y_1, z_1, x_2, y_2, z_2, \ldots) \\
\mathbf{p} &= (p_{x1}, p_{y1}, p_{z1}, p_{x2}, p_{y2}, p_{z2}, \ldots)
\end{aligned}$$

Phase space

- Classical system: its state completely given by coordinates and momenta of the N particles
- $6N$-dimensional space: phase space
- One point in the phase space at given time

$$\begin{aligned}
\mathbf{X} &= (x_1, y_1, z_1, p_{x1}, p_{y1}, p_{z1}, x_2, y_2, z_2, p_{x2}, p_{y2}, p_{z2}, \ldots) \\
\mathbf{q} &= (x_1, y_1, z_1, x_2, y_2, z_2, \ldots) \\
\mathbf{p} &= (p_{x1}, p_{y1}, p_{z1}, p_{x2}, p_{y2}, p_{z2}, \ldots) \\
\mathbf{X} &= (\mathbf{q}, \mathbf{p})
\end{aligned}$$



System at equilibrium

- System with constant N, V and T
- Computing the property A of the system at equilibrium

$$\langle A \rangle = \int \int A(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p}$$

System at equilibrium

- System with constant N, V and T
- Computing the property A of the system at equilibrium

$$\langle A\rangle = \int\int A(\mathbf{q},\mathbf{p})P(\mathbf{q},\mathbf{p})d\mathbf{q}d\mathbf{p}$$

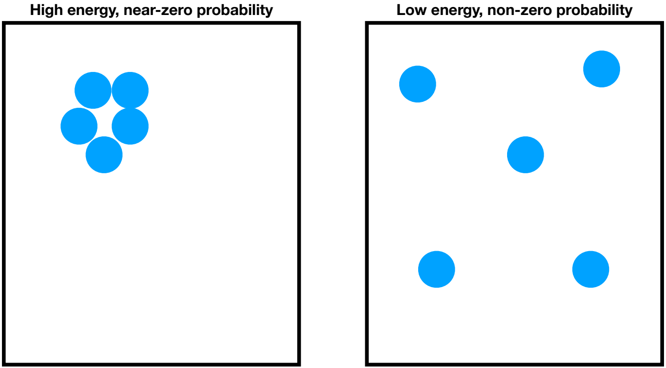
- P is the probability of the system to be in a point of the phase space

$$\begin{aligned} P &= Q^{-1}e^{-E(\mathbf{q},\mathbf{p})/k_BT} \\ Q &= \int\int e^{-E(\mathbf{q},\mathbf{p})/k_BT}d\mathbf{q}d\mathbf{p} \\ E(\mathbf{q},\mathbf{p}) &= V(\mathbf{q}) + \sum_i \frac{\mathbf{p}_i^2}{2m_i} \end{aligned}$$

- $e^{-E(\mathbf{q},\mathbf{p})/k_BT}$: Boltzmann factor
- E: energy of the system
- $k_B = 1.38 \times 10^{-23}$ JK^{-1}: Boltzmann constant
- Q: canonical partition function

System at equilibrium

- Phase space is huge (e.g., $N \approx 10^4 - 10^6$ in a protein)
- How to compute $\langle A \rangle$?
- Sampling important points in the phase space
- High-energy phase points have near-zero probability



How to compute $\langle A \rangle$

$$\langle A \rangle = ?$$

Monte Carlo methods:

- Sampling the phase space with random numbers
- Equilibrium properties
- No access to dynamic properties

Molecular dynamics:

- Time trajectories in the phase space
- Equilibrium properties
- Access to dynamic properties

WITH

Molecular mechanics

Monte Carlo method

- No equation of motion to integrate

Monte Carlo method

- No equation of motion to integrate
- A is separable in terms of **q** and **p**

$$\langle A \rangle = \int A(\mathbf{q})P(\mathbf{q})d\mathbf{q} + \int A(\mathbf{p})P(\mathbf{p})d\mathbf{p}$$

- $3N$- instead of $6N$-dimensional integrals

$$\langle A \rangle = \frac{1}{R}\sum_{i}^{R} A(\mathbf{q}_i)$$

- R random points
- Choosing phase points randomly is inefficient

Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm

Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm
    1. Choose $\mathbf{q}_i$, compute $V(\mathbf{q}_i)$ and $A(\mathbf{q}_i)$

Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm
    1. Choose $\mathbf{q}_i$, compute $V(\mathbf{q}_i)$ and $A(\mathbf{q}_i)$
    2. $\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta\mathbf{q}_{\text{trial}}$ and compute $V(\mathbf{q}_{i+1})$

Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm
  1. Choose $\mathbf{q}_i$, compute $V(\mathbf{q}_i)$ and $A(\mathbf{q}_i)$
  2. $\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta\mathbf{q}_{\text{trial}}$ and compute $V(\mathbf{q}_{i+1})$
  3. In *NVT* ensemble, acceptance probability p

$$p = \min\left[1, \frac{e^{-\frac{V_{i+1}}{k_B T}}}{e^{-\frac{V_i}{k_B T}}}\right]$$

Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm
  1. Choose $\mathbf{q}_i$, compute $V(\mathbf{q}_i)$ and $A(\mathbf{q}_i)$
  2. $\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta\mathbf{q}_{\text{trial}}$ and compute $V(\mathbf{q}_{i+1})$
  3. In *NVT* ensemble, acceptance probability p

$$p = \min\left[1, \frac{e^{-\frac{V_{i+1}}{k_B T}}}{e^{-\frac{V_i}{k_B T}}}\right]$$

  4. If $V(\mathbf{q}_{i+1}) \leq V(\mathbf{q}_i)$, the move is accepted (compute $A(\mathbf{q}_{i+1})$)
  5. Otherwise, p is compared with an uniform random number z
     - Move accepted if $p \geq z$ (compute $A(\mathbf{q}_{i+1})$)
     - Otherwise, the move is rejected

Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm
  1. Choose $\mathbf{q}_i$, compute $V(\mathbf{q}_i)$ and $A(\mathbf{q}_i)$
  2. $\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta\mathbf{q}_{\text{trial}}$ and compute $V(\mathbf{q}_{i+1})$
  3. In *NVT* ensemble, acceptance probability p

$$p = \min\left[1, \frac{e^{-\frac{V_{i+1}}{k_B T}}}{e^{-\frac{V_i}{k_B T}}}\right]$$

  4. If $V(\mathbf{q}_{i+1}) \leq V(\mathbf{q}_i)$, the move is accepted (compute $A(\mathbf{q}_{i+1})$)
  5. Otherwise, p is compared with an uniform random number z
     - Move accepted if $p \geq z$ (compute $A(\mathbf{q}_{i+1})$)
     - Otherwise, the move is rejected
  6. Repeat 2) to 5)

Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm
  1. Choose $\mathbf{q}_i$, compute $V(\mathbf{q}_i)$ and $A(\mathbf{q}_i)$
  2. $\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta\mathbf{q}_{\text{trial}}$ and compute $V(\mathbf{q}_{i+1})$
  3. In *NVT* ensemble, acceptance probability p

$$p = \min\left[1, \frac{e^{-\frac{V_{i+1}}{k_B T}}}{e^{-\frac{V_i}{k_B T}}}\right]$$

  4. If $V(\mathbf{q}_{i+1}) \leq V(\mathbf{q}_i)$, the move is accepted (compute $A(\mathbf{q}_{i+1})$)
  5. Otherwise, p is compared with an uniform random number z
     - Move accepted if $p \geq z$ (compute $A(\mathbf{q}_{i+1})$)
     - Otherwise, the move is rejected
  6. Repeat 2) to 5)
- Markov chain

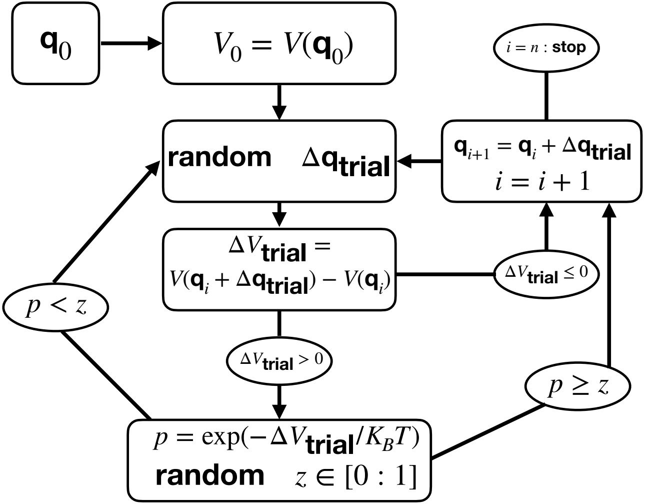
Monte Carlo method

- Points $\mathbf{q}_i$ selected through the Metropolis algorithm
  1. Choose $\mathbf{q}_i$, compute $V(\mathbf{q}_i)$ and $A(\mathbf{q}_i)$
  2. $\mathbf{q}_{i+1} = \mathbf{q}_i + \Delta\mathbf{q}_{\text{trial}}$ and compute $V(\mathbf{q}_{i+1})$
  3. In *NVT* ensemble, acceptance probability p

$$p = \min\left[1, \frac{e^{-\frac{V_{i+1}}{k_B T}}}{e^{-\frac{V_i}{k_B T}}}\right]$$

  4. If $V(\mathbf{q}_{i+1}) \leq V(\mathbf{q}_i)$, the move is accepted (compute $A(\mathbf{q}_{i+1})$)
  5. Otherwise, p is compared with an uniform random number z
     - Move accepted if $p \geq z$ (compute $A(\mathbf{q}_{i+1})$)
     - Otherwise, the move is rejected
  6. Repeat 2) to 5)
- Markov chain
- Step too small or too large

Monte Carlo method: flowchart



Average over trajectory

- Reasonable phase point as initial condition (t_0)
- Evolution over time
- Samples relevant regions of the phase space

$$\langle A \rangle = \frac{1}{M} \sum_i^M A(t_i)$$

- M: number of times A is computed

Average over trajectory

- Reasonable phase point as initial condition (t_0)
- Evolution over time
- Samples relevant regions of the phase space

$$\langle A \rangle = \frac{1}{M} \sum_{i}^{M} A(t_i)$$

- M: number of times A is computed
- Continuous formulation

$$\langle A \rangle = \lim_{t \to \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(\tau) d\tau$$

- $\langle A \rangle$ independent of t_0: 'ergodic hypothesis'

Molecular dynamics (MD)

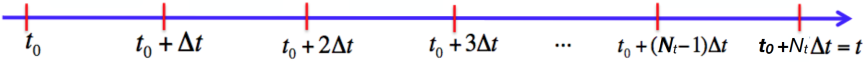
- How to follow a phase-space trajectory?

Molecular dynamics (MD)

- How to follow a phase-space trajectory?
- Classical representation of nuclei, second Newton's law

$$\mathbf{F} = m\mathbf{a} \quad \Rightarrow \quad -\frac{dV}{d\mathbf{q}} = m\frac{d^2\mathbf{q}}{dt^2}$$

- V: potential energy at position $\mathbf{q}$
- $\mathbf{q}_i = \mathbf{q}(t_i)$
- MD trajectory is deterministic



Time discretization

- Taylor expansion: movement forward with a Δt later

$$\begin{aligned}
\mathbf{q}_{i+1} &= \mathbf{q}_i + \frac{\partial \mathbf{q}}{\partial t}(\Delta t) + \frac{1}{2}\frac{\partial^2 \mathbf{q}}{\partial t^2}(\Delta t)^2 + \frac{1}{6}\frac{\partial^3 \mathbf{q}}{\partial t^3}(\Delta t)^3 + \ldots \\
\mathbf{q}_{i+1} &= \mathbf{q}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 + \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \ldots
\end{aligned}$$

Time discretization

- Taylor expansion: movement forward with a Δt later

$$\begin{aligned}\mathbf{q}_{i+1} &= \mathbf{q}_i + \frac{\partial \mathbf{q}}{\partial t}(\Delta t) + \frac{1}{2}\frac{\partial^2 \mathbf{q}}{\partial t^2}(\Delta t)^2 + \frac{1}{6}\frac{\partial^3 \mathbf{q}}{\partial t^3}(\Delta t)^3 + \dots \\ \mathbf{q}_{i+1} &= \mathbf{q}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 + \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \dots\end{aligned}$$

- Movement backward with a Δt earlier $(-\Delta t)$

$$\mathbf{q}_{i-1} = \mathbf{q}_i - \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 - \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \dots$$

Time discretization

- Verlet algorithm

$$\begin{aligned}\mathbf{q}_{i+1} &= \mathbf{q}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 + \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \dots \\ &+ \\ \mathbf{q}_{i-1} &= \mathbf{q}_i - \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 - \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \dots \\ \hline \mathbf{q}_{i+1} &= (2\mathbf{q}_i - \mathbf{q}_{i-1}) + \mathbf{a}_i(\Delta t)^2 + \dots\end{aligned}$$

Time discretization

- Verlet algorithm

$$\begin{array}{rcl}
\mathbf{q}_{i+1} & = & \mathbf{q}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 + \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \ldots \\
 & + & \\
\mathbf{q}_{i-1} & = & \mathbf{q}_i - \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 - \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + \ldots \\
\hline
\mathbf{q}_{i+1} & = & (2\mathbf{q}_i - \mathbf{q}_{i-1}) + \mathbf{a}_i(\Delta t)^2 + \ldots \\
\mathbf{q}_{i+1} & = & (2\mathbf{q}_i - \mathbf{q}_{i-1}) - \frac{1}{m}\frac{dV}{d\mathbf{q}}(\Delta t)^2 + \ldots.
\end{array}$$

- About the initial point

$$\mathbf{q}_{-1} = \mathbf{q}_0 - \mathbf{v}_0 \Delta t$$

Time discretization

- Verlet algorithm

$$\begin{array}{rcl} \mathbf{q}_{i+1} & = & \mathbf{q}_i + \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 + \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + ... \\ & + & \\ \mathbf{q}_{i-1} & = & \mathbf{q}_i - \mathbf{v}_i(\Delta t) + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 - \frac{1}{6}\mathbf{b}_i(\Delta t)^3 + ... \\ \hline \mathbf{q}_{i+1} & = & (2\mathbf{q}_i - \mathbf{q}_{i-1}) + \mathbf{a}_i(\Delta t)^2 + ... \\ \mathbf{q}_{i+1} & = & (2\mathbf{q}_i - \mathbf{q}_{i-1}) - \frac{1}{m}\frac{dV}{d\mathbf{q}}(\Delta t)^2 + ... \end{array}$$

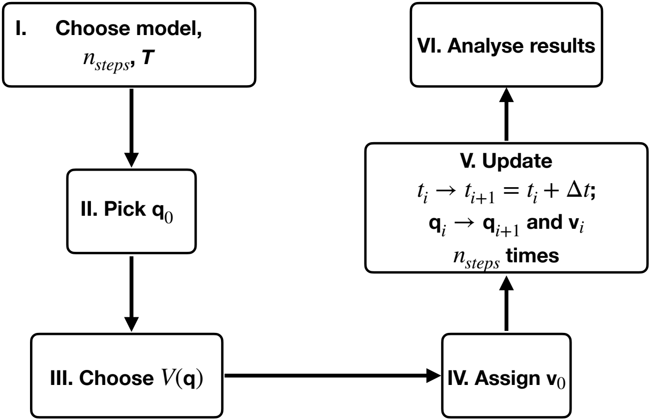
- About the initial point

$$\mathbf{q}_{-1} = \mathbf{q}_0 - \mathbf{v}_0 \Delta t$$

- Velocity Verlet algorithm

$$\begin{array}{rcl} \mathbf{q}_{i+1} & = & \mathbf{q}_i + \mathbf{v}_i \Delta t + \frac{1}{2}\mathbf{a}_i(\Delta t)^2 \\ \mathbf{v}_{i+1} & = & \mathbf{v}_i + \frac{1}{2}(\mathbf{a}_i + \mathbf{a}_{i+1})\Delta t \end{array}$$

Molecular dynamics: flowchart (*NVT*)



I. Choose model, n_{steps}, *T*

II. Pick $\mathbf{q}_0$

III. Choose $V(\mathbf{q})$

IV. Assign $\mathbf{v}_0$

V. Update
$t_i \rightarrow t_{i+1} = t_i + \Delta t$;
$\mathbf{q}_i \rightarrow \mathbf{q}_{i+1}$ **and** $\mathbf{v}_i$
n_{steps} **times**

VI. Analyse results

MD and ensembles

- *NVE* ensemble
  - N, V and E are constant
  - *P* and *T* fluctuate
  - Natural ensemble in MD

MD and ensembles

- *NVE* ensemble
  - N, V and E are constant
  - *P* and *T* fluctuate
  - Natural ensemble in MD
- Total energy *E*
  - Conservation of energy
  - Finite precision and time steps
  - *E* preserved within a numerical threshold

MD and ensembles

- *NVE* ensemble
  - N, V and E are constant
  - *P* and *T* fluctuate
  - Natural ensemble in MD
- Total energy *E*
  - Conservation of energy
  - Finite precision and time steps
  - *E* preserved within a numerical threshold
- Temperature *T* from average kinetic energy

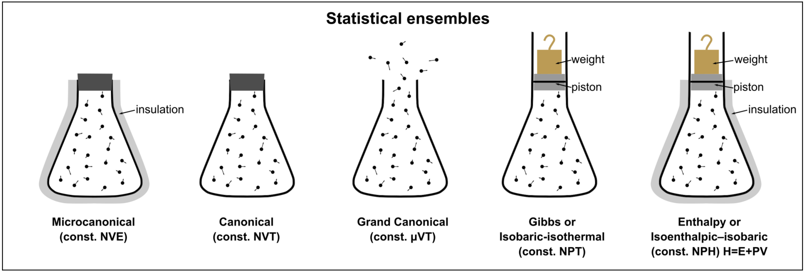
$$E_K = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i} = \sum_i^N \frac{1}{2} m_i \mathbf{v}_i^2 = \frac{3}{2} N k_B T$$

MD and ensembles

- NVT ensemble
    - Scaling velocities $\mathbf{v}_i$
    - System coupled to a heat bath
    - Nosé-Hoover method: bath with dynamic variables

MD and ensembles

- NVT ensemble
  - Scaling velocities $\mathbf{v}_i$
  - System coupled to a heat bath
  - Nosé-Hoover method: bath with dynamic variables
- NPT ensemble
  - Scaling coordinates $\mathbf{q}_i$
  - Pressure bath and Nosé-Hoover method



Example: liquid water

- Box length of 20 Å, 270 molecules, density of around 0.997 g cm^{-3}
- Initial structure: r(O-H) = 1 Å and H-O-H angle of 100°

Example: liquid water

- Box length of 20 Å, 270 molecules, density of around 0.997 g cm^{-3}
- Initial structure: r(O-H) = 1 Å and H-O-H angle of 100°
- TIP3P force field (molecular mechanics)
  - Stretching and bending terms
    - $r_0 = 0.9572$ Å, $k_{\mathrm{stretc}} = 600$ kcal mol^{-1} Å^{-2}
    - $\theta_0 = 104.52°$, $k_{\mathrm{bend}} = 75$ kcal mol^{-1} rad^{-2}
  - -0.834 (0.417) au charges on O (H)
  - Lennard-Jones term for O
    - $\epsilon = 0.1521$ kcal mol^{-1}, $\sigma_{\mathrm{OO}} = 3.15061$ Å

Example: liquid water

- Box length of 20 Å, 270 molecules, density of around 0.997 g cm^{-3}
- Initial structure: r(O-H) = 1 Å and H-O-H angle of 100°
- TIP3P force field (molecular mechanics)
    - Stretching and bending terms
        - $r_0 = 0.9572$ Å, $k_{\text{stretc}} = 600$ kcal mol^{-1} Å^{-2}
        - $\theta_0 = 104.52°$, $k_{\text{bend}} = 75$ kcal mol^{-1} rad^{-2}
    - -0.834 (0.417) au charges on O (H)
    - Lennard-Jones term for O
        - $\epsilon = 0.1521$ kcal mol^{-1}, $\sigma_{\text{OO}} = 3.15061$ Å
- *NVT* conditions, thermostat at 298 *K*

Example: liquid water

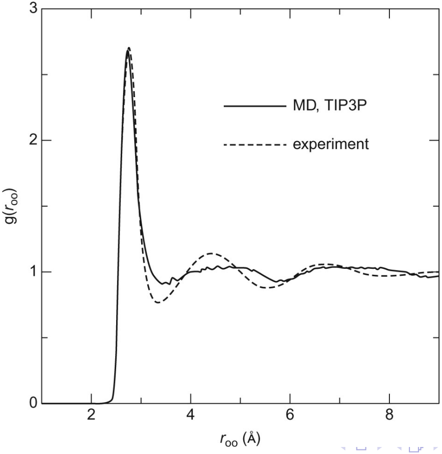
- Box length of 20 Å, 270 molecules, density of around 0.997 g cm^{-3}
- Initial structure: r(O-H) = 1 Å and H-O-H angle of 100°
- TIP3P force field (molecular mechanics)
  - Stretching and bending terms
    - $r_0 = 0.9572$ Å, $k_{\text{stretc}} = 600$ kcal mol^{-1} Å^{-2}
    - $\theta_0 = 104.52°$, $k_{\text{bend}} = 75$ kcal mol^{-1} rad^{-2}
  - -0.834 (0.417) au charges on O (H)
  - Lennard-Jones term for O
    - $\epsilon = 0.1521$ kcal mol^{-1}, $\sigma_{\text{OO}} = 3.15061$ Å
- *NVT* conditions, thermostat at 298 *K*
- Equilibration phase

Example: liquid water

- Box length of 20 Å, 270 molecules, density of around 0.997 g cm^{-3}
- Initial structure: r(O-H) = 1 Å and H-O-H angle of 100°
- TIP3P force field (molecular mechanics)
  - Stretching and bending terms
    - $r_0 = 0.9572$ Å, $k_{\text{stretc}} = 600$ kcal mol^{-1} Å^{-2}
    - $\theta_0 = 104.52°$, $k_{\text{bend}} = 75$ kcal mol^{-1} rad^{-2}
  - -0.834 (0.417) au charges on O (H)
  - Lennard-Jones term for O
    - $\epsilon = 0.1521$ kcal mol^{-1}, $\sigma_{\text{OO}} = 3.15061$ Å
- *NVT* conditions, thermostat at 298 *K*
- Equilibration phase
- 50 ps simulated

Example: liquid water

- Radial distribution function $g(r_{OO})$
- Peak at 3 Å due to hydrogen bonding (coordination number of water, 4)
- Secondary peaks as next-nearest-neighbour effects



How to compute $g(r_{AB})$

- g from X-ray or neutron diffraction measurements

How to compute $g(r_{AB})$

- g from X-ray or neutron diffraction measurements
- g from MD

$$\frac{1}{V} g(r_{AB}) = \frac{1}{N_A N_B} \left\langle \sum_i^{N_A} \sum_j^{N_B} \delta[r_{AB} - r_{A_i B_j}] \right\rangle$$

$$P\{\mathrm{A}, \mathrm{B}, r_{AB}, \Delta r\} = \frac{4\pi r_{AB}^2}{V} g(r_{AB}) \Delta r$$

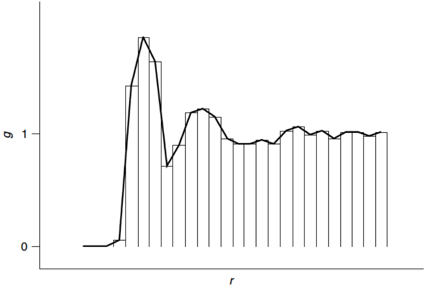
How to compute $g(r_{AB})$

- g from X-ray or neutron diffraction measurements
- g from MD

$$\frac{1}{V}g(r_{AB}) = \frac{1}{N_A N_B}\left\langle \sum_i^{N_A}\sum_j^{N_B}\delta[r_{AB} - r_{A_iB_j}]\right\rangle$$

$$P\{\mathrm{A},\mathrm{B},r_{AB},\Delta r\} = \frac{4\pi r_{AB}^2}{V}g(r_{AB})\Delta r$$

- In practice:
  - All A-B distances computed at each time snapshot
  - Each occurrence added to a bin of the histogram



Dynamical properties

- Autocorrelation function $C(t)$ or $\tilde{C}(t)$ for property A

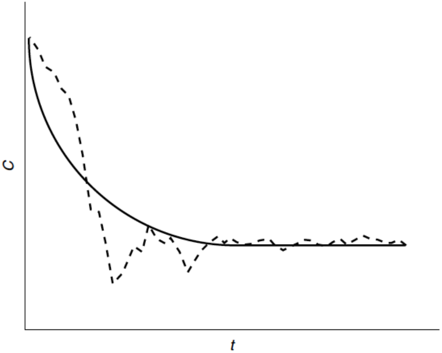
$$C(t) = \langle A(t_0)A(t_0+t)\rangle_{t_0}$$

Dynamical properties

- Autocorrelation function $C(t)$ or $\tilde{C}(t)$ for property A

$$\begin{aligned} C(t) &= \langle A(t_0)A(t_0+t)\rangle_{t_0} \\ \tilde{C}(t) &= \frac{\langle [A(t_0)-\langle A\rangle][\langle [A(t_0+t)-\langle A\rangle]\rangle_{t_0}}{\langle A^2\rangle - \langle A\rangle^2} \end{aligned}$$

- Used for:
  - Determining diffusion coefficients
  - Computing vibrational infrared spectra



Reducing simulation complexity

GOAL:

- Reduce the number of degrees of freedom
- Computational saving

STOCHASTIC DYNAMICS:

- Useful for solute dynamics
- Implicit environment
- Stochastic modelling of system/ environment interaction

COARSE-GRAINED MODELS:

- Reduce atomic resolution
- Applied to very large systems
- Simulating complex membranes, fibrils, cells etc.

Stochastic dynamics

- Molecular deterministic dynamics: accurate estimation of thermodynamic information
- Stochastic dynamics:
    - Dynamical evolution of a system embedded in a larger one (solute in a solvent)
    - Larger system modelled stochastically
    - Computational saving
- Langevin dynamics

$$\mathbf{a}(t) = \zeta \mathbf{p}(t) + \frac{1}{m}[\mathbf{F}_{\text{intra}}(t) + \mathbf{F}_{\text{continuum}}(t)]$$

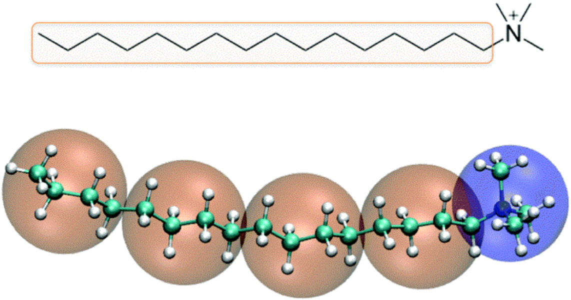
- Continuum interacting with the smaller system
    - Microscopic friction coefficient ζ
    - $\mathbf{F}_{\text{intra}}$ with usual force fields
    - $\mathbf{F}_{\text{continuum}}$ for modelling the effect of the larger system

Coarse-grained methods

- Computational expense unaffordable for atomistic MM and MD
- Building blocks larger than single atoms
- Advantages:
  - Computational saving by reducing the number of degrees of freedom
  - Larger time steps in time propagation

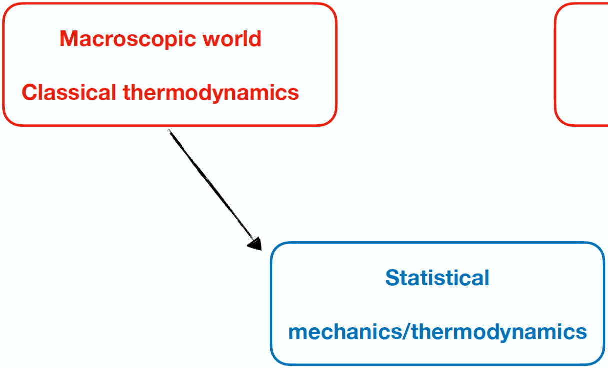
Coarse-grained methods

- Computational expense unaffordable for atomistic MM and MD
- Building blocks larger than single atoms
- Advantages:
  - Computational saving by reducing the number of degrees of freedom
  - Larger time steps in time propagation
- Disadvantage:
  - Atomic resolution lost



Statistical thermodynamics

Statistical thermodynamics



Macroscopic world

Classical thermodynamics

Microscopic world

Quantum mechanics

Statistical

mechanics/thermodynamics

Statistical thermodynamics and equilibrium

- Bridge between microscopic and macroscopic world

Statistical thermodynamics and equilibrium

- Bridge between microscopic and macroscopic world
- Partition function used to compute thermodynamic properties
- Chemical equilibrium

$$\Delta_r^0 G = -RT \ln K$$

Statistical thermodynamics and equilibrium

- Bridge between microscopic and macroscopic world
- Partition function used to compute thermodynamic properties
- Chemical equilibrium

$$\Delta_r^0 G = -RT \ln K$$

- Free energy of a mole of ideal gas at pressure P and temperature T

$$G(P,T) = G(T = 0\ K) - RT \ln \left[\frac{q(P,T)}{N_A} \right]$$

Statistical thermodynamics and equilibrium

- Bridge between microscopic and macroscopic world
- Partition function used to compute thermodynamic properties
- Chemical equilibrium

$$\Delta_r^0 G = -RT \ln K$$

- Free energy of a mole of ideal gas at pressure P and temperature T

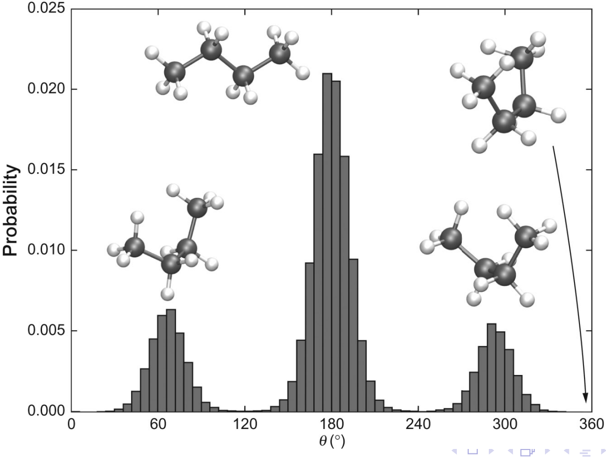
$$G(P,T) = G(T = 0\ K) - RT \ln \left[\frac{q(P,T)}{N_A} \right]$$

- With q

$$q = \int\int e^{-\epsilon(\mathbf{q},\mathbf{p})/k_BT} d\mathbf{q} d\mathbf{p}$$

Example: *n*-butane

- Distribution of the dihedral angle of *n*-butane at 273 K
- 50 *n*-butane molecules, 120 ps simulated
- Sampling every 100 fs (1200 snapshots)
- $50 \times 1200 = 60000$ structures



Example: *n*-butane

- *Anti* conformer ($\theta = 180^\circ$, most stable)
- Enantiomeric *gauche* conformers ($\theta = 60^\circ$ and 300°)

Example: *n*-butane

- *Anti* conformer ($\theta = 180^\circ$, most stable)
- Enantiomeric *gauche* conformers ($\theta = 60^\circ$ and 300°)
- *a* conformers with $\theta \in [120^\circ : 240^\circ]$ ($n(a) = 38,637$)
- g^+ conformers with $\theta \in [0^\circ : 120^\circ]$ ($n(g^+) = 11,599$)
- g^- conformers with $\theta \in [240^\circ : 360^\circ]$ ($n(g^-) = 9,764$)

Example: *n*-butane

- *Anti* conformer ($\theta = 180^\circ$, most stable)
- Enantiomeric *gauche* conformers ($\theta = 60^\circ$ and 300°)
- a conformers with $\theta \in [120^\circ : 240^\circ]$ ($n(a) = 38,637$)
- g^+ conformers with $\theta \in [0^\circ : 120^\circ]$ ($n(g^+) = 11,599$)
- g^- conformers with $\theta \in [240^\circ : 360^\circ]$ ($n(g^-) = 9,764$)
- $n(a/g^+/g^-) \propto$ partition function of the conformers
- *Anti* conformer as reference $\rightarrow \Delta G(a) = 0$

$$\begin{aligned} \Delta G(g^+) &= -RT \ln \frac{n(g^+)}{n(a)} = 2.73 \text{ kJ mol}^{-1} \\ \Delta G(g^-) &= -RT \ln \frac{n(g^-)}{n(a)} = 3.12 \text{ kJ mol}^{-1} \end{aligned}$$

Statistical thermodynamics and equilibrium

- Combining quantum mechanics and statistical thermodynamics
- ϵ as a discrete quantity
- Single-particle partition function and probability

$$
\begin{aligned}
q &= \sum_i g_i \exp\left[-\frac{\epsilon_i}{k_B T}\right] \\
p_i &= \frac{1}{q} g_i \exp\left[-\frac{\epsilon_i}{k_B T}\right]
\end{aligned}
$$

Statistical thermodynamics and equilibrium

- Combining quantum mechanics and statistical thermodynamics
- ϵ as a discrete quantity
- Single-particle partition function and probability

$$\begin{aligned}
q &= \sum_i g_i \exp\left[-\frac{\epsilon_i}{k_B T}\right] \\
p_i &= \frac{1}{q} g_i \exp\left[-\frac{\epsilon_i}{k_B T}\right] \\
\epsilon_i &= \epsilon_{\mathrm{tr},i} + \epsilon_{\mathrm{vib},i} + \epsilon_{\mathrm{rot},i} + \epsilon_{el,i} \\
q &= q_{\mathrm{tr}} q_{\mathrm{vib}} q_{\mathrm{rot}} q_{\mathrm{el}}
\end{aligned}$$

Statistical thermodynamics and equilibrium

- For a generic aA + bB $\rightleftharpoons$ cC + dD reaction:

$$\Delta_r G^0 = \Delta_r G^0(0\ K) - RT \ln \frac{(q_C^0/N_A)^c (q_D^0/N_A)^d}{(q_A^0/N_A)^a (q_B^0/N_A)^b}$$

Statistical thermodynamics and equilibrium

- For a generic $a\mathrm{A} + b\mathrm{B} \rightleftharpoons c\mathrm{C} + d\mathrm{D}$ reaction:

$$\Delta_r G^0 = \Delta_r G^0(0\ K) - RT \ln \frac{(q_C^0/N_A)^c (q_D^0/N_A)^d}{(q_A^0/N_A)^a (q_B^0/N_A)^b}$$

- In practice
  1. Compute with quantum chemistry

$$\begin{aligned} \Delta_r G^0(T=0\ K) &= cG^0(T=0\ K, \mathrm{C}) + dG^0(T=0\ K, \mathrm{D}) \\ &- \left(aG^0(T=0\ K, \mathrm{A}) + bG^0(T=0\ K, \mathrm{B})\right) \end{aligned}$$

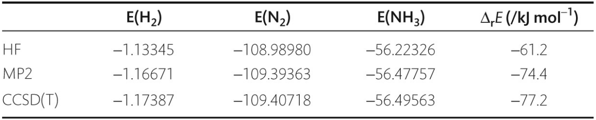
  2. Compute $\Delta_r G^0(P,T)$ with q for A, B, C and D moieties

Example: $N_2 + 3H_2 \rightleftharpoons 2\ NH_3$

- Computing the equilibrium constant

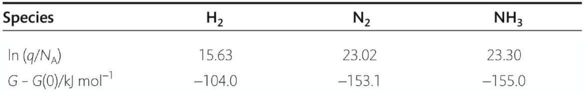
$$\Delta_r G^0 = -RT \ln \left(\frac{P^2_{NH_3} (P^0)^2}{P_{N_2} P^3_{H_2}} \right)$$

- Using the protocol of the previous slide
  1. $\Delta_r G^0 (T = 0\ K) = \Delta_r E$



	$E(H_2)$	$E(N_2)$	$E(NH_3)$	$\Delta_r E$ (/kJ mol^{-1})
HF	−1.13345	−108.98980	−56.22326	−61.2
MP2	−1.16671	−109.39363	−56.47757	−74.4
CCSD(T)	−1.17387	−109.40718	−56.49563	−77.2

  2. At 800 K



Species	H_2	N_2	NH_3
$\ln (q/N_A)$	15.63	23.02	23.30
$G - G(0)$/kJ mol^{-1}	−104.0	−153.1	−155.0

$$\begin{aligned} \Delta_r G^0(800K) &= -77.2 + (2 \times -155.0 - (-153.1 + 3 \times -104.0)) \\ &= 77.9 \text{ kJ mol}^{-1} \\ K_p &= 8.2 \times 10^{-6} \\ K_p^{\text{ref}} &= 9.0 \times 10^{-6} \quad \text{(reference)} \end{aligned}$$