

Statistical Sampling via Molecular Dynamics Simulations

Vibhu Vardhan Singh

PI: Prof. Francesco Paesani

University of California San Diego



UC San Diego
Chemistry & Biochemistry



LABORATORY FOR THEORETICAL AND COMPUTATIONAL CHEMISTRY

Statistical Mechanics

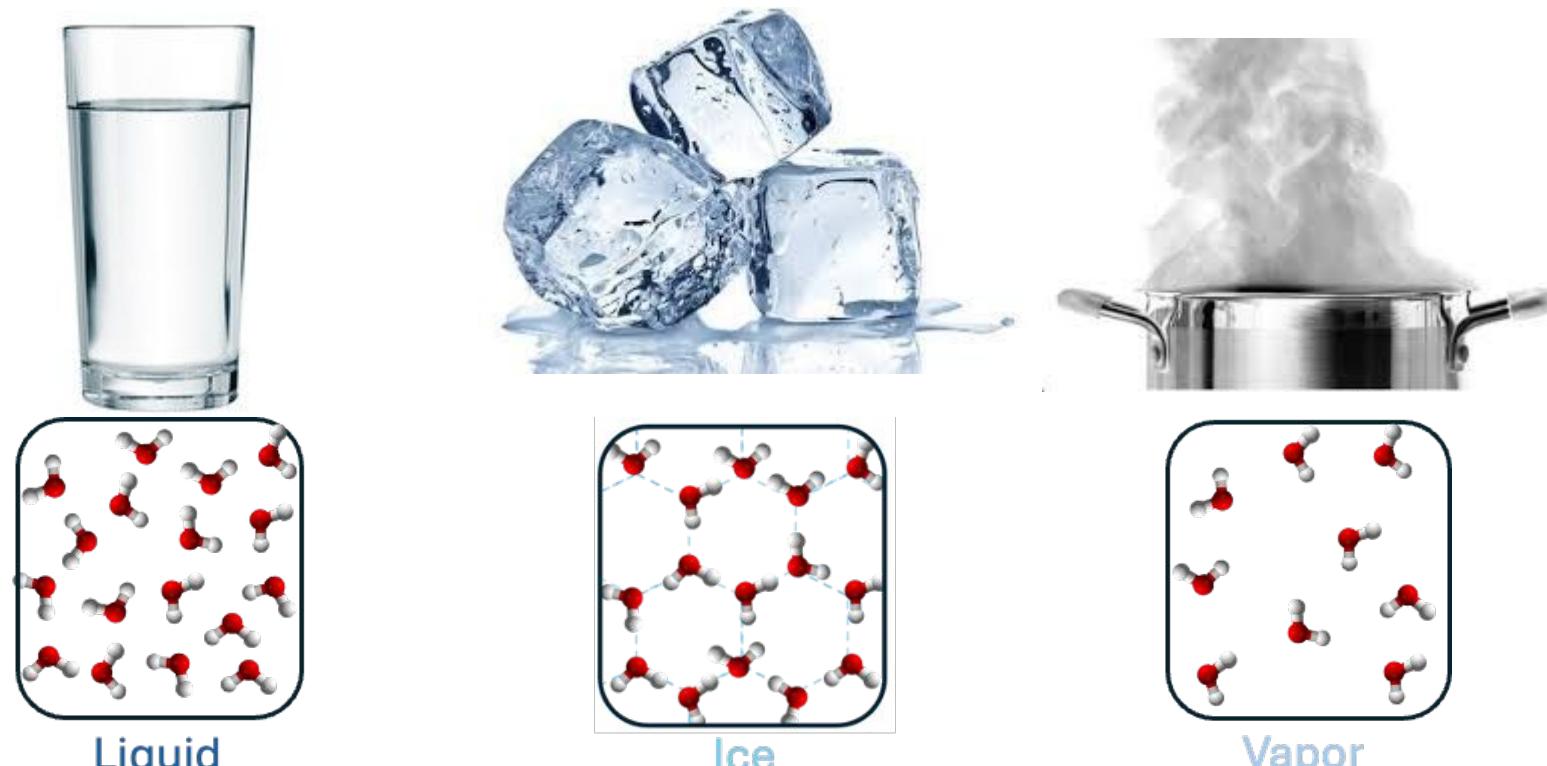
Thermodynamics: Study of relationship between Macroscopic properties (Density, pressure, viscosity etc)



How is the structure at a molecular level?



Statistical Mechanics



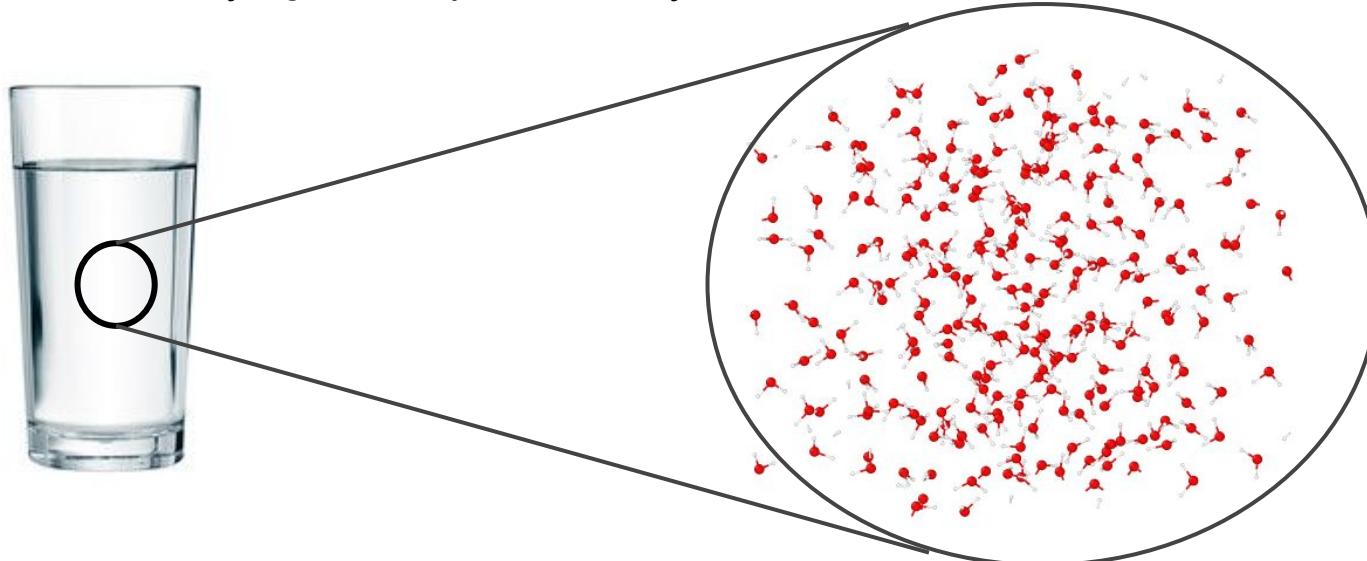
How is the structure at a molecular level?

From these structures, can we calculate the bulk properties?



Statistical Mechanics

How to connect such macroscopic properties to the systems to the microscopic properties that we were studying for the past few days?



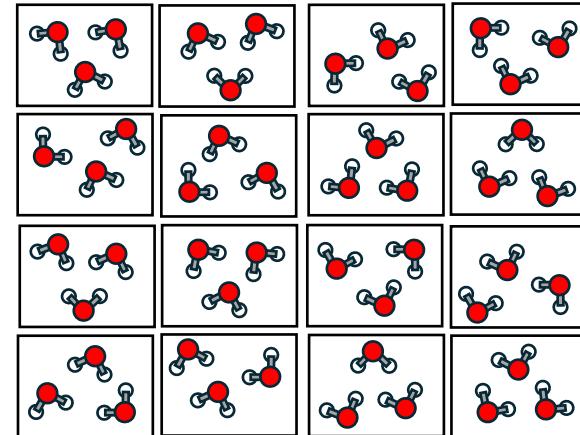
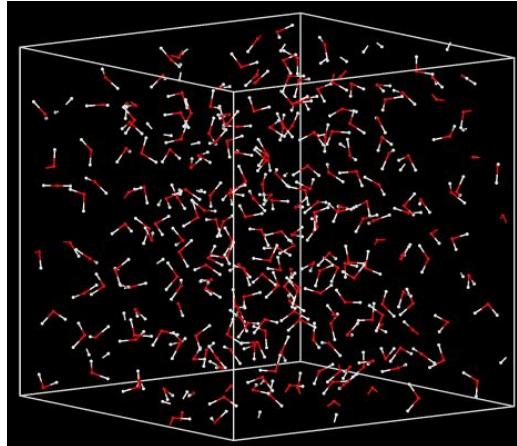
Statistical Mechanics/Thermodynamics makes the connection between the bulk properties and microscopic properties



Statistical Ensemble

What is an ensemble?

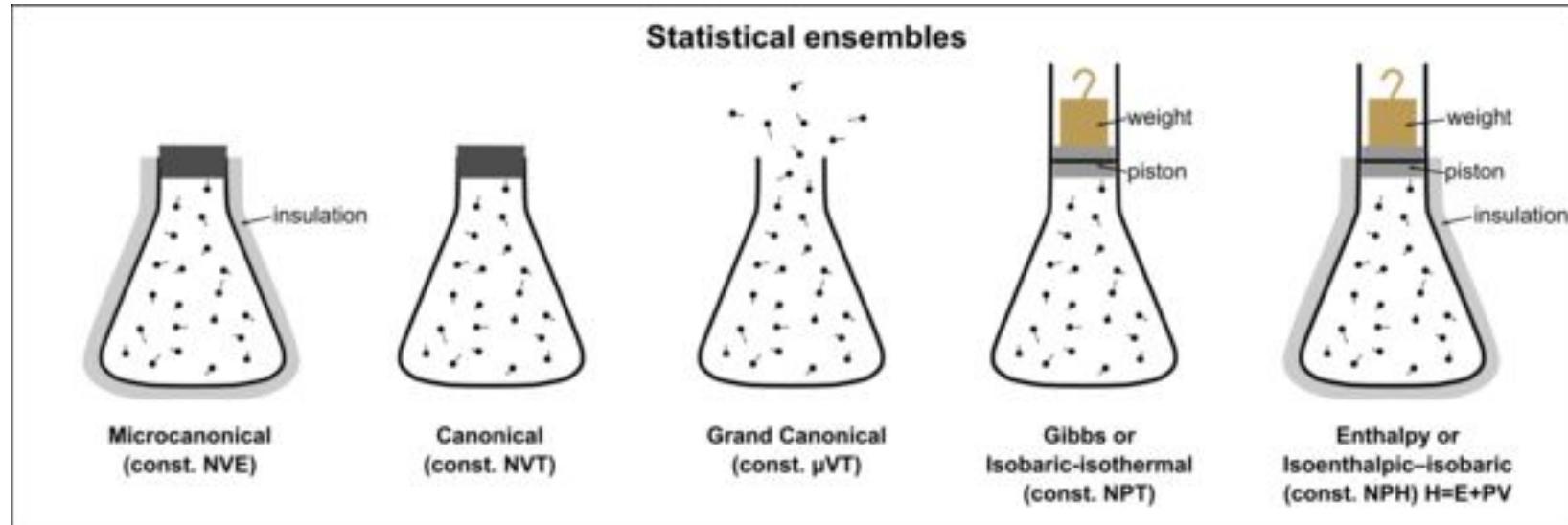
Consider it as a virtual copy of the system, each of which represent a particular state in which the system is



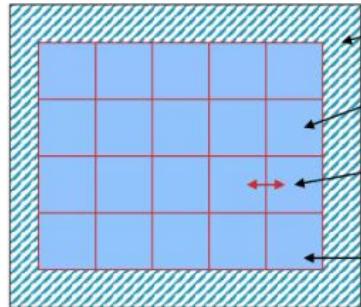
Statistical Ensemble

What is an ensemble?

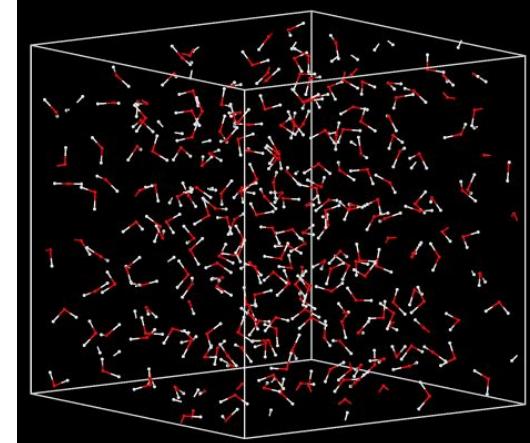
It is a collection of particles inside a mechanical system that has some unique properties



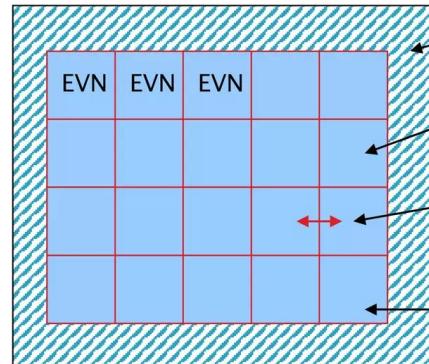
Canonical Ensemble



- Large reservoir (constant T)*
- All the ensemble members have the same (n, V, T)*
- Energy can be exchanged but particles cannot*
- Number of Microstates $N \rightarrow \infty$*



Micro Canonical Ensemble

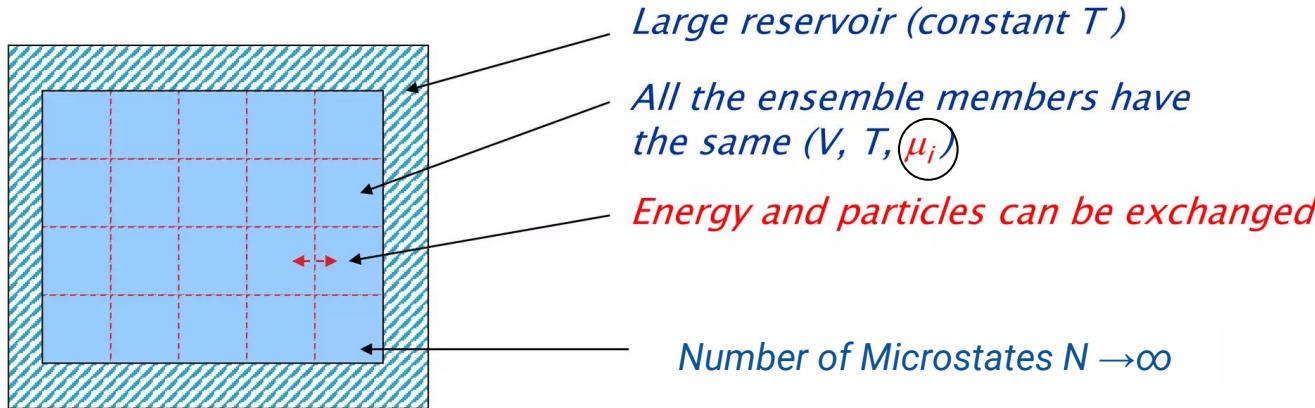


- impermeable*
- All the ensemble members have the same (E, V, N)*
- there is no transfer of Energy and mass*
- isolated*

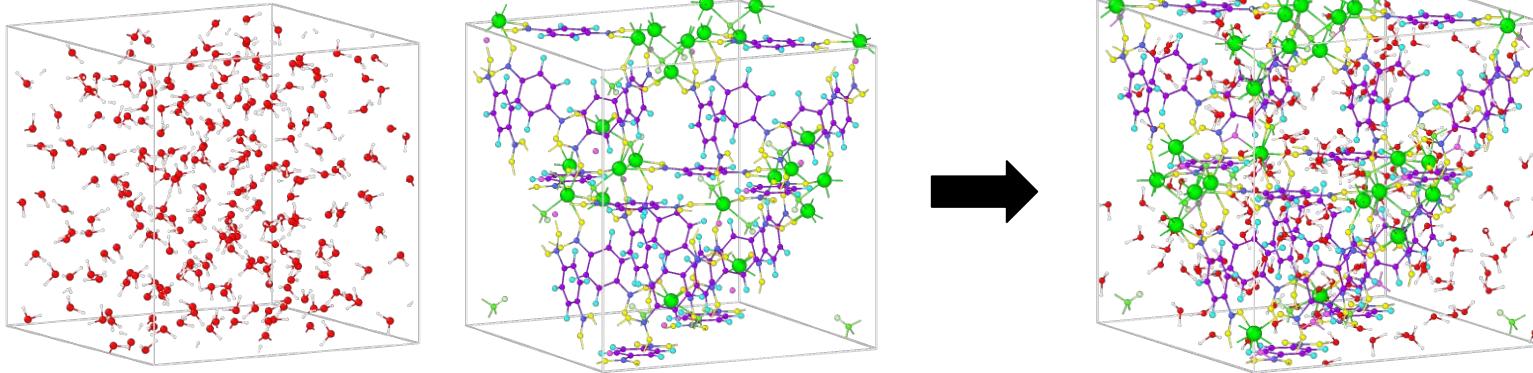
Different Frames correspond to different ensemble



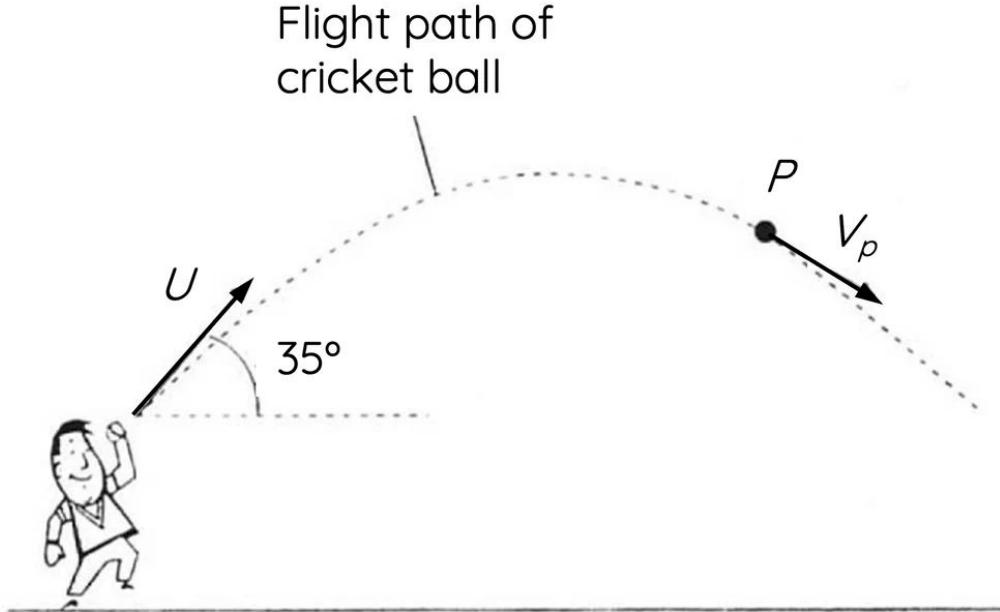
Grand Canonical Ensemble



Chemical potential is the reason why diffusion happens!!



Classical Physics and Newton's laws of motion



Once we throw a cricket ball, do we know what's the path it's going to follow?

Using the knowledge of projectile motion, Yes! we can.



Classical Physics and Newton's equations of motion

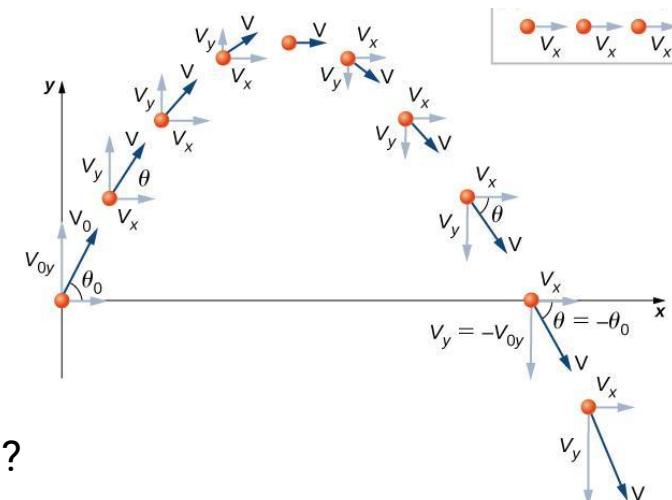
Similar concept can be applied when for atoms. If we know the initial momentum (velocity and mass) and position, we can see how the trajectory evolves with time if we know the force

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

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2m} [\mathbf{F}(t) + \mathbf{F}(t + \delta t)] \delta t$$

Velocity Verlet Algorithm (Numerical Integration of Newton's motion)

Question is how to get the respective forces or force field?



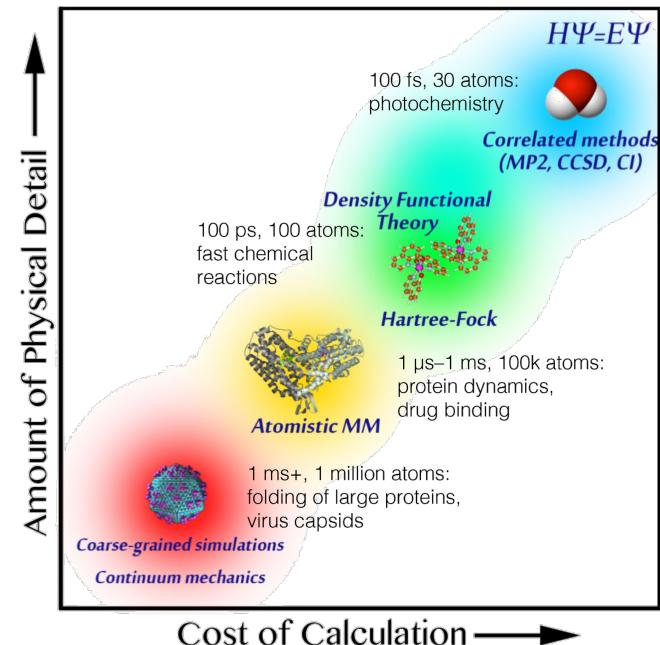
Quantum vs Classical Molecular Dynamics?

Limitations of Quantum MD

Computational cost increases as the size of the system increases

Higher level of theory means higher computational cost

Classical MD: Follows Classical laws of mechanics
(Electronic degree of freedom is not involved) to evolve the state as a function of time.



Lee Ping Wang
bios-203-lecture-3-classical-molecular-dynamics



Force Fields in Classical MD

$$\mathbf{F} = -\nabla U$$

How to find U?

$$U(R) = \sum_{\text{bonds}} k_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} k_\phi(1 + \cos[n\phi - \gamma]) + \sum_{\text{impropers}} k_\omega(\omega - \omega_{\text{eq}})^2 + \sum_{i < j}^{\text{atoms}} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] + \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

bond



angle



dihedral



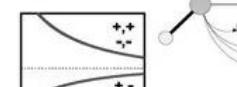
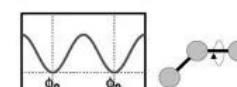
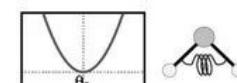
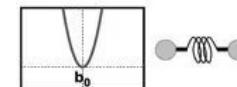
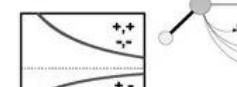
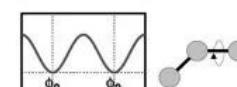
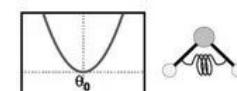
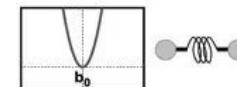
improper



van der Waals



electrostatic



[Bond function simulator](#)

Bonded
Interactions

Non-Bonded
Interactions



Ergodic Hypothesis and Time evolution of a system

Ensemble Average:

$$\langle f \rangle_e = \frac{1}{N} \sum_i^N f(i)$$

Time Average:

$$\langle f \rangle_t = \frac{1}{T} \sum_i^T f(i)$$

f is any macroscopic property of the system

At infinite time, the particle will visit the whole phase space



Statistical Mechanics and relation with Classical MD

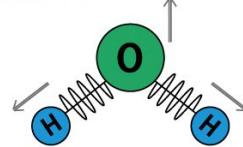
Statistical Mechanics requires sampling the system in a way that all possible states/configurations of a system are visited

Classical MD can sample the space but only for a finite time

How can we calculate the macroscopic properties of a system when it's only sampled for a finite time or finite number of steps? (Generally 1 timestep is in the order of femtoseconds)

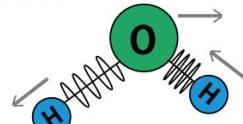
Why the timestep is so small ???

- Symmetric stretching vibration at 3652 cm^{-1}



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

- Anti-symmetric stretching vibration at 2349 cm^{-1}

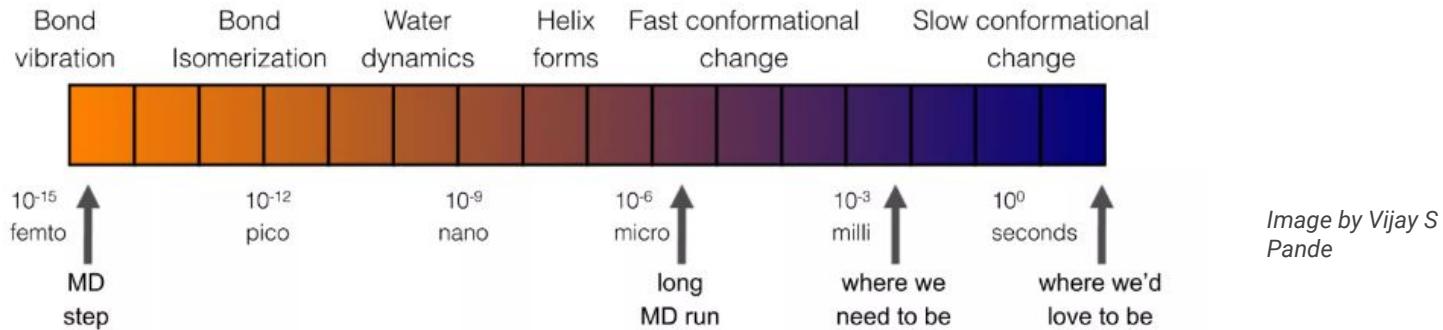


$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2m} [\mathbf{F}(t) + \mathbf{F}(t + \delta t)] \delta t$$

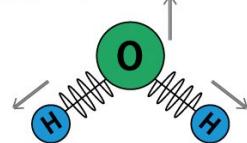
Calculate the frequency in seconds?
Hint: $1\text{cm}^{-1} = 3*10^{10}\text{ Hz(sec}^{-1})$



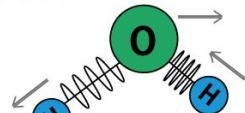
Statistical Mechanics and relation with Classical MD



• Symmetric stretching vibration at 3652 cm^{-1}



• Anti-symmetric stretching vibration at 2349 cm^{-1}



$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2m}\mathbf{F}(t)\delta t^2$$
$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2m} [\mathbf{F}(t) + \mathbf{F}(t + \delta t)] \delta t$$

It takes around 9 and 14 femtoseconds respectively to complete one vibration



Thermodynamic Equilibrium in MD

When there is minimal change in variables of the system, say total energy or volume or temperature of the system, the system is said to be in equilibrium

In MD, we have to equilibrate the system first before calculating any properties

Thermodynamic equilibrium means that the system is in its most stable states, or their probabilities are the most, hence just by finite sampling of a system, we can get most of the macroscopic properties. **Remember the Ergodic hypothesis!!!**

In a Thermodynamic equilibrium:

- Properties are history independent
- Properties should be invariant with time



Basics of Molecular Dynamics

- Thermostat
- Barostat
- Short range interactions, Long range interactions
- Periodic boundary conditions



Thermostat



- Controls the temperature of the system by providing optimal velocities to the particle
- Nose-Hoover Thermostat is one of the common used thermostat in LAMMPS
- LAMMPS provides flexibility to apply thermostat to a particular region in system, specific particles or only to any of the x,y,z components of the particles.

Temperature of the system is calculated using:

$$T = \frac{2E_{\text{kin}}}{N_{\text{DOF}}k_B} \quad \text{with} \quad E_{\text{kin}} = \sum_{i=1}^{N_{\text{atoms}}} \frac{1}{2}m_i v_i^2 \quad \text{and} \quad N_{\text{DOF}} = n_{\text{dim}}N_{\text{atoms}} - n_{\text{dim}} - N_{\text{fixDOFs}}$$



Barostat

- Controls the pressure of the system by adjust the velocity and volume of the system
- Nose-Hoover Barostat is one of a common used Barostat in LAMMPS
- LAMMPS allows us to set the pressure along any of the x,y,z components of the simulation box

The pressure of the system is calculated using:

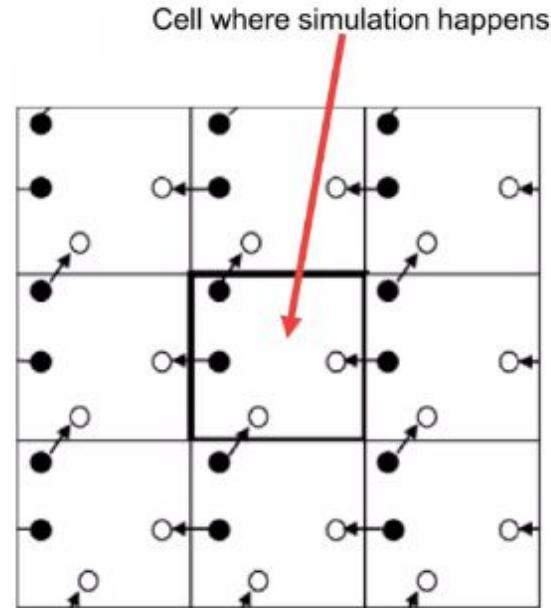
$$P = \frac{Nk_B T}{V} + \frac{1}{Vd} \sum_{i=1}^{N'} \vec{r}_i \cdot \vec{f}_i$$

d is the dimensionality



Periodic Boundary Conditions

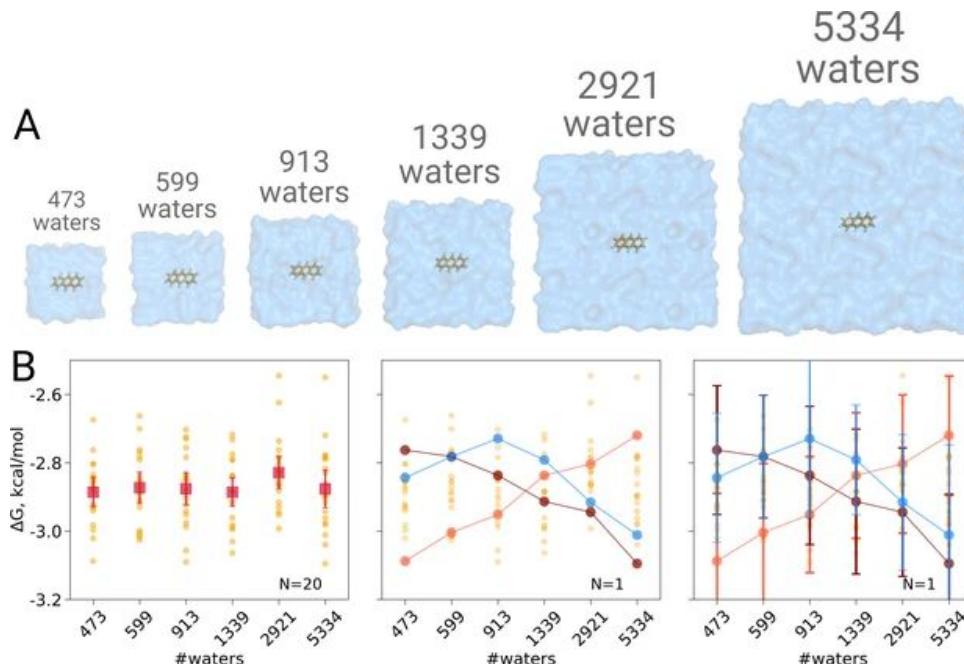
- Number of atoms in a simulation cell is finite. But in practical, they are surrounded by other atoms all the time
- To address the issue, a periodic image of simulation box is created
- If particle leaves the box from one side, it enters from the other



This increases the number of Interactions!!!
Which increase computational cost!!!!



Periodic boundary condition in MD



PBC eliminates the need for a bigger simulation box.

Statistics (different simulations, different initial conditions, providing sufficient time to equilibrate) are more important than system size

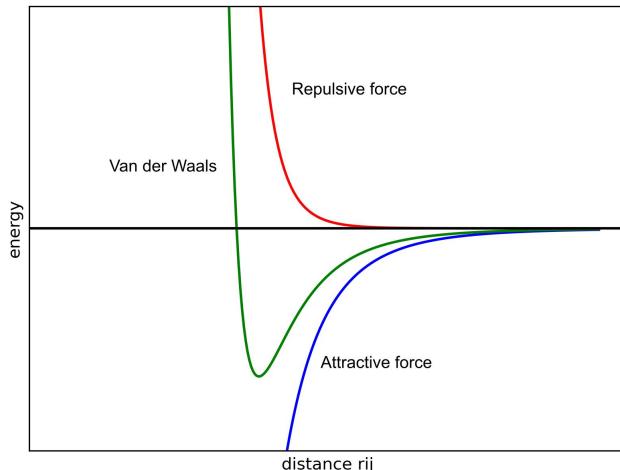
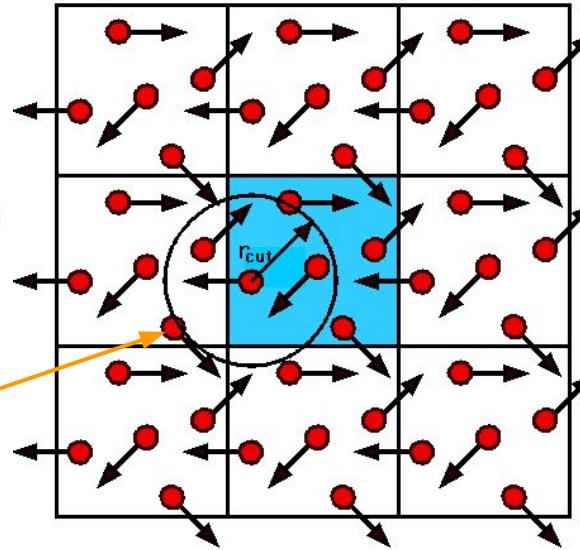
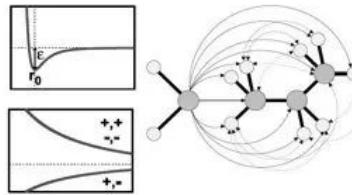
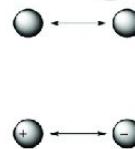
Vytautas Gapsys, Bert L de Groot (2020) On the importance of statistics in molecular simulations for thermodynamics, kinetics and simulation box size



Short and Long range interactions

$$\sum_{i < j} \text{atoms} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] \text{van der Waals}$$

$$\sum_{i < j} \text{atoms} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \text{electrostatic}$$



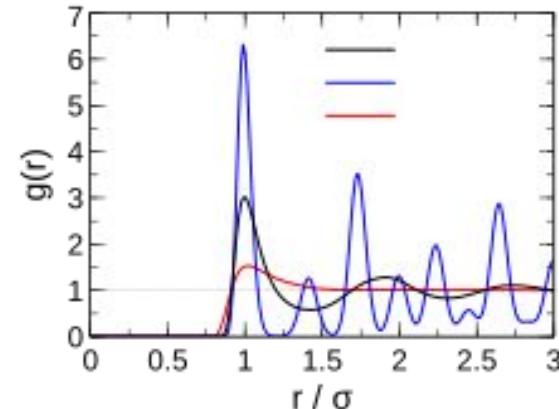
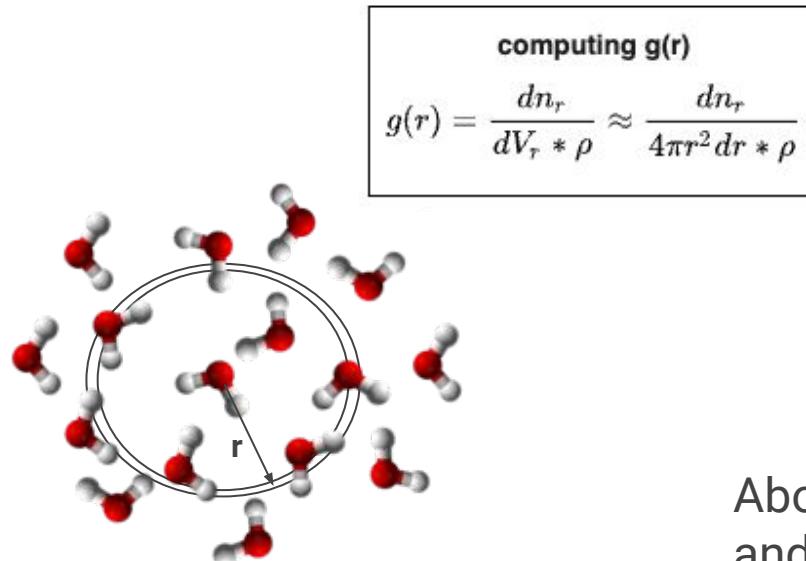
[function simulator](#)

- Cutoff is set for both vdw and electrostatic potentials
- Long range interactions for electrostatics use different algorithms (Handled in reciprocal space)



Radial Distribution Function

- Variation of Density as a function of distance w.r.t. a reference particle

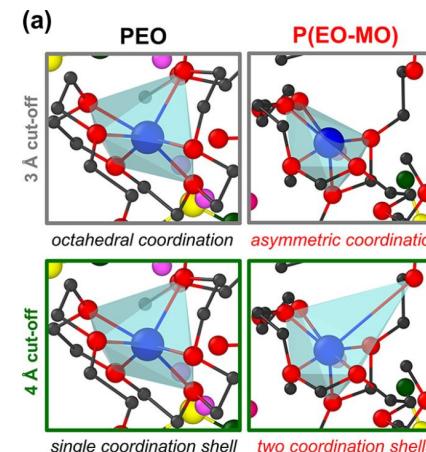
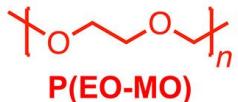
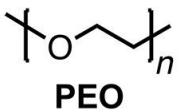
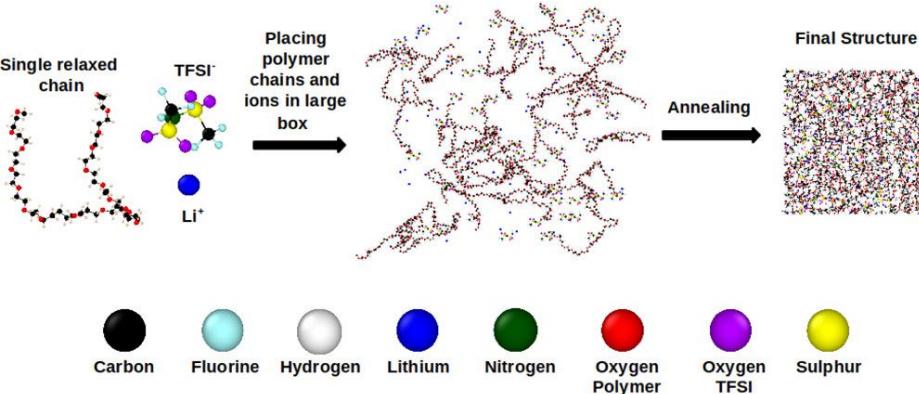


Above plot is for Argon in solid, liquid and gas. Guess which one is for whom?

https://en.wikibooks.org/wiki/Molecular_Simulation/Radial_Distribution_Functions



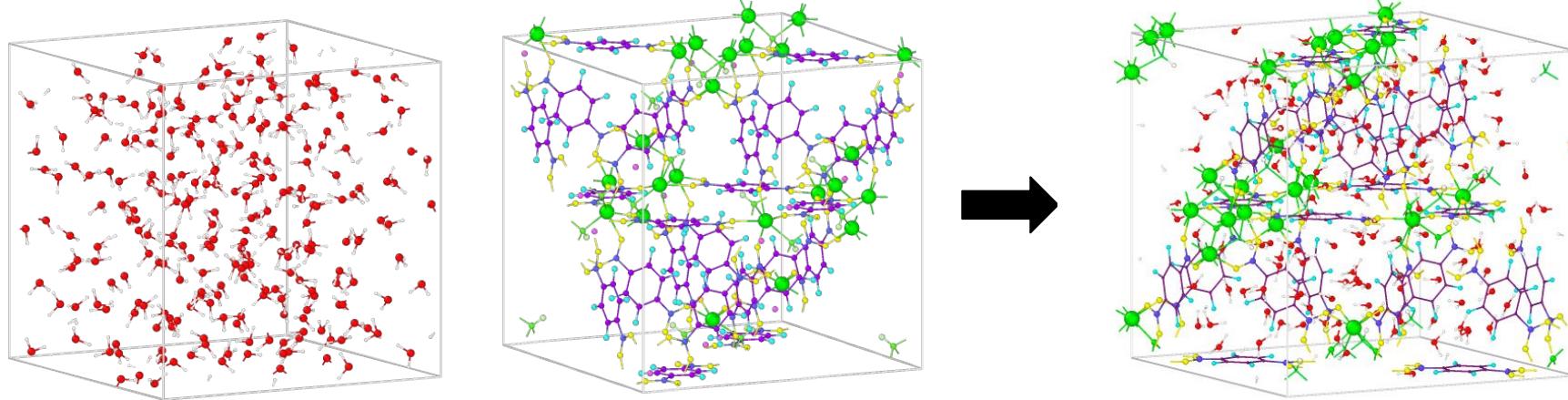
Applications of RDF



Macromolecules 2021, 54, 18, 8590–8600



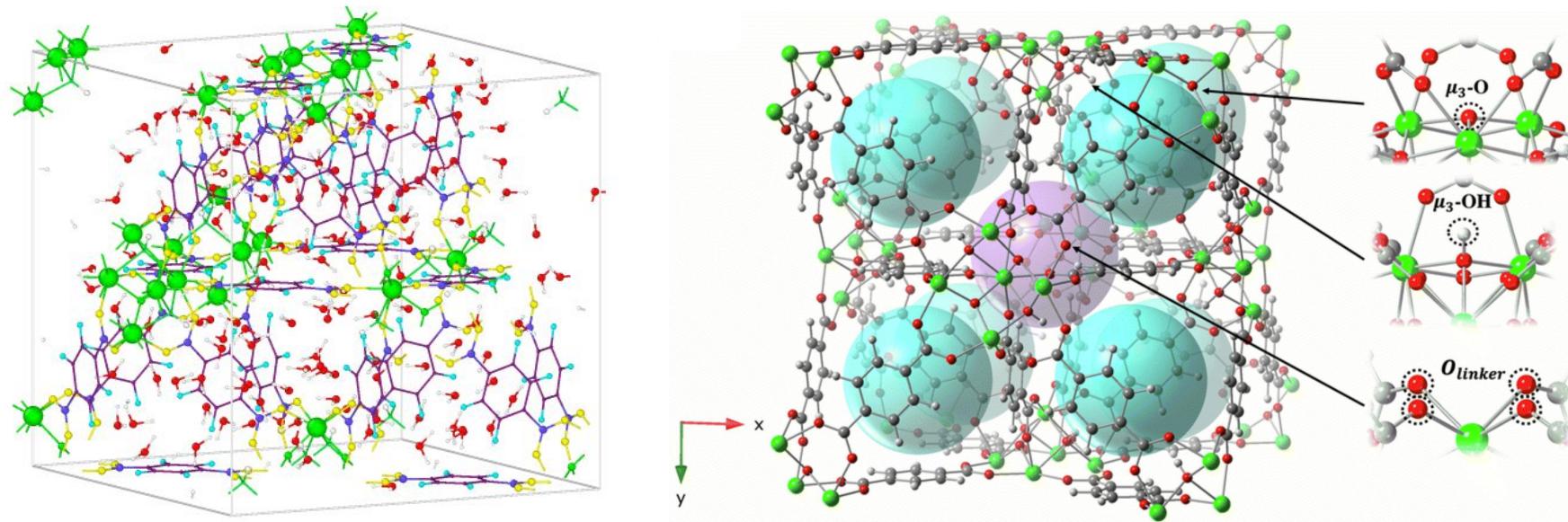
RDF Applications - Water in MOF



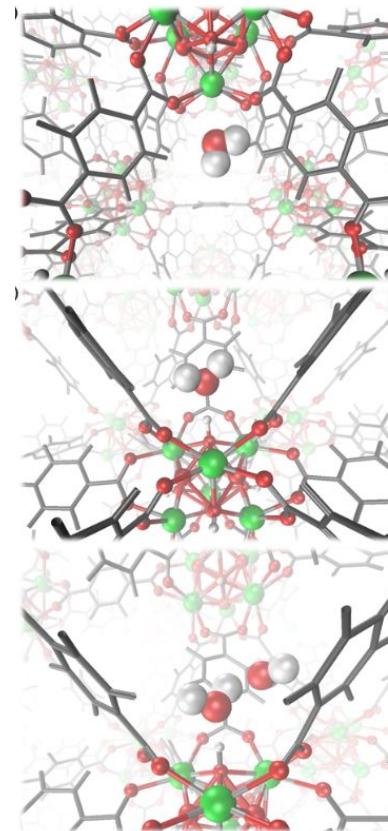
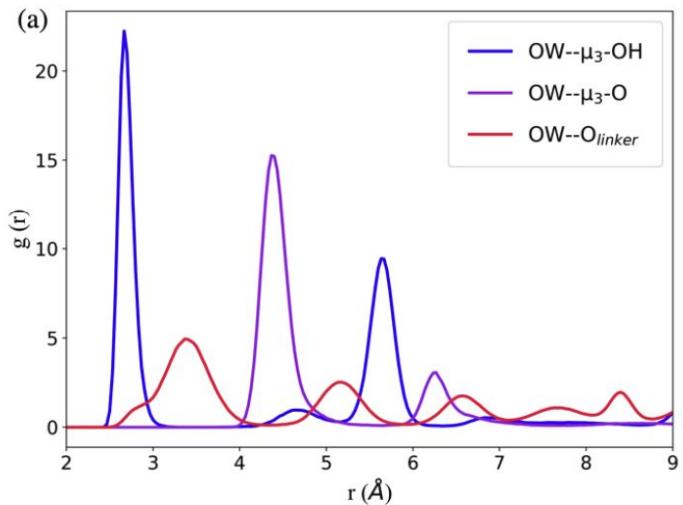
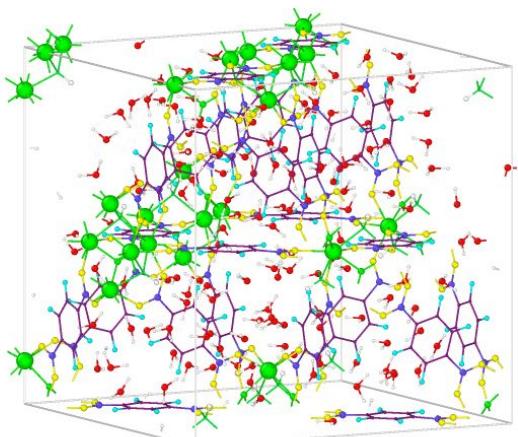
Watching a movie is cool, but we need statistical insights to understand system



RDF Applications - Water intake mechanism in MOF



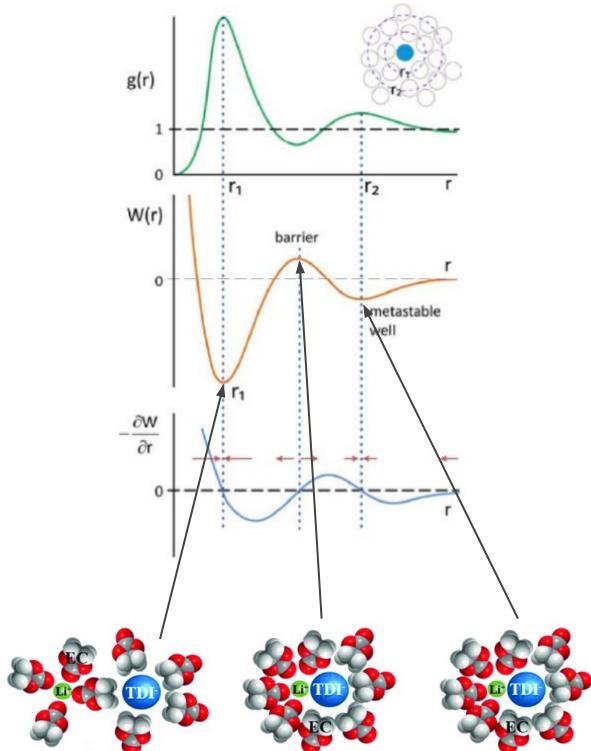
Simulation of water in MOF Zr UiO-66 using MB-pol



Computational Insights into the Interaction of Water with the UiO-66 Metal-Organic Framework and Its Functionalized Derivatives



Using $g(r)$ to calculate the activation barrier



$g(r)$ is related to the potential of mean force $W(r)$

$$g(r) = \exp \left[-\frac{w^{(2)}(r)}{kT} \right] \text{ function calculator}$$

$$w^{(2)}(r) = -kT \ln(g(r))$$

Potential of mean force can be understood as the free energy surface along the chosen coordinate

Caution! This can be calculated only when the possible configurations were sampled (Showed system is only for representation)

https://chem.libretexts.org/Bookshelves/Biological_Chemistry/Concepts_in_Biophysical_Chemistry



Thank You



LABORATORY FOR THEORETICAL AND COMPUTATIONAL CHEMISTRY