Spring College on Water in Physics, Chemistry and Biology

10 - 21 April 2007

Introduction to molecular dynamics simulations methods

Mauro BOERO
Center for Computational Sciences,
University of Tsukuba,
Ibaraki 305-8577, Japan
Lecture 1: Introduction to molecular dynamics
- from classical force fields to first principles

Mauro Boero

Center for Computational Sciences, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 Japan
Outline of the lectures

• Molecular Dynamics (MD):
  Newtonian and Lagrangean dynamics
  Basic algorithms
  Time averages and ensemble averages

• Density Functional Theory (DFT) based MD:
  Brief review of DFT and Born-Oppenheimer approximation
  Car-Parrinello Molecular Dynamics (CPMD)
  Practical implementation

• Reactive CPMD:
  General problem in chemical reactions
  Blue Moon ensemble approach
  Metadynamics

• Water as a catalyst:
  Catalytic properties of supercritical water
  Synthetic reactions in supercritical H₂O – Beckmann rearrangement
Molecular Dynamics (MD)

The aim of Molecular Dynamics (MD) is to study a system of interacting particles (solid, liquid, gas) by recreating it on a computer in a way as close as possible to nature and by simulating its dynamics over a physical length of time relevant to the properties of interest.

Any MD method is an iterative numerical scheme for solving some equations of motion (EOM), coded in a computer program, that represent the physical evolution of the system under study.
\[ F = m \, a \]

\[ e^{-\frac{\Delta E}{kT}} \]

**Time and Length scales in simulations**

- **Time Scale (s)**
  - \(10^{-12}\)
  - \(10^{-8}\)
  - \(10^{-6}\)

- **Length Scale (m)**
  - \(10^{-10}\)
  - \(10^{-8}\)
  - \(10^{-6}\)
  - \(10^{-4}\)

- **Monte Carlo**
- **mesoscale continuum**
- **ab initio**
- Molecular dynamics
- \(H \psi = E \psi\)
Classical Molecular Dynamics (MD)

The objects described by Molecular Dynamics (MD) are particles (atoms, molecules, polymers) represented as deterministic variables having positions and velocities.

The Cartesian positions \((x_1,y_1,z_1),\ldots,(x_N,y_N,z_N)\) of a system of \(N\) particles can be denoted by \(\mathbf{R}_I = (x_I,y_I,z_I)\) where \(I = 1,\ldots,N\)

These particles interact via a given function of the positions \(\mathbf{R}_I\), the potential \(V(\mathbf{R}_I)\). The forces \(\mathbf{f}_I\) on each particle are simply the gradients (derivative) of this potential,

\[
\mathbf{f}_I = -\frac{\partial V(\mathbf{R}_I)}{\partial \mathbf{R}_I}
\]

and the analytical 3xN dimensional function \(\mathbf{f}_I\) is called force field
MD Simulations: how to construct a Force Field?

We have to consider all the interactions that we want to study:

- **bond stretch**
  \[ v(r_{ij}, k_{ij}) = \frac{1}{2} k_{ij} (r_{ij} - r_{ij}^0)^2 \]

- **bending**
  \[ v(\theta_{ijk}) = \frac{1}{2} k_{\theta} (\theta_{ijk} - \theta_{ijk}^0)^2 \]

- **torsion**
  \[ v(\phi_{ijkl}) \]

- **intermolecular interactions**

- **intramolecular nonbonded interactions**
MD Simulations: How do we get the parameters $k_{IJ}$, $k_{\theta}$, etc… for the force field?

From experiments
- molecular and bulk properties
- X-ray / neutron scattering structure factors
- isotopic substitutions, etc…

From *ab initio* calculations
- molecular and cluster properties
- static geometry optimization (minima, saddle points)
- first principles methods
“Eppur si muove…” (Galileo Galilei, 1564-1642)

Atoms are generally not fixed at a given position $R_I$, but they move due to e.g. finite temperature $T > 0$ K, collisions, external fields, etc…so they are rather described by dynamical variables $R_I(t)$. The evolution of $R_I(t)$ for each atom $I=1,\ldots,N$ is described in analytical mechanics by a Lagrangean(*)

\[
L(R_I, \dot{R}_I) = T - V = \sum_{I=1}^{K} \frac{1}{2} M_I \dot{R}_I^2 - V(R_I)
\]

(*): after Joseph Louis Lagrange, Torino (Italy), 1736-1813. See *Mechanique analytique* and *Miscellanea Taurinensia*, 1766-1773 (5 volumes).
The Euler-Lagrange equations give us the equations of motion (EOM) of Classical MD

The atoms move from a position \( \mathbf{R}_I(t) \) to a new position \( \mathbf{R}_I(t+dt) \) via standard dynamics:

\[
\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{\mathbf{R}}_I} \right) - \frac{\partial L}{\partial \mathbf{R}_I} = 0
\]

or, more explicitly, the *old good* Newton equation

\[
M_I \ddot{\mathbf{R}}_I = -\nabla V (\mathbf{R}_I) = f_I
\]

where \( M_I \) is the mass of the \( I^{th} \) particle of our system moving in the force field \( f_I \)
Numerical integration of the EOM: Verlet algorithm

To solve numerically the EOM, the first step is a *discretization* of the time $t$ in terms of small increments called *time steps* of (arbitrary) length $\delta t$. The system then passes across a set of time ordered configurations

$$\ldots \mathbf{R}_I(t_{m-1}) \rightarrow \mathbf{R}_I(t_m) \rightarrow \mathbf{R}_I(t_{m+1}) \ldots$$

separated in time by $\delta t = t_m - t_{m-1} = t_{m+1} - t_m = \ldots$

An easy but successful integrator in MD is the Verlet algorithm; if we make a Taylor expansion of $\mathbf{R}_I(t+\delta t)$ and $\mathbf{R}_I(t-\delta t)$ we get

$$\mathbf{R}_I(t + \delta t) = \mathbf{R}_I(t) + \delta t \cdot \mathbf{v}_I(t) + \frac{\delta t^2}{2M_I} \mathbf{f}_I(t) + \frac{\delta t^3}{6} \mathbf{b}_I(t) + O(\delta t^4)$$

$$\mathbf{R}_I(t - \delta t) = \mathbf{R}_I(t) - \delta t \cdot \mathbf{v}_I(t) + \frac{\delta t^2}{2M_I} \mathbf{f}_I(t) - \frac{\delta t^3}{6} \mathbf{b}_I(t) + O(\delta t^4)$$
We use the standard notation \( \mathbf{v}_I = \mathbf{\dot{R}}_I, \mathbf{a}_I = \mathbf{\ddot{R}}_I, \mathbf{b}_I = \mathbf{\dddot{R}}_I \)

and the EOM \( \mathbf{f}_I = \mathbf{M}_I \mathbf{a}_I \)

By simply summing up the two Taylor expansions, we get

\[
\mathbf{R}_I(t + \delta t) = 2\mathbf{R}_I(t) - \mathbf{R}_I(t - \delta t) + \frac{\delta t^2}{\mathbf{M}_I} \mathbf{f}_I(t) + O(\delta t^4)
\]

and this has an accuracy which is third order in time.

The velocity results, in a similar way, as

\[
\mathbf{v}_I(t + \delta t) = \mathbf{v}_I(t) + \frac{\delta t}{2\mathbf{M}_I} [\mathbf{f}_I(t) + \mathbf{f}_I(t + \delta t)] + O(\delta t^4)
\]

and this is what we call the velocity Verlet algorithm, a standard in any MD computer code.
Thermodynamic control methods: constant temperature (I)

From a statistical mechanics point of view, the temperature $T$ of a system of particles having masses $M_i$ and moving with velocities $v_i$ is simply given by the average

$$\langle \sum_{i=1}^{N} \frac{M_i v_i^2}{2} \rangle = \frac{3}{2} N \cdot k_B T$$

It seems than that if we can control in some way the velocities $v_i$ then we can keep under control the related parameter $T$. From statistical mechanics textbooks, we know that this control is operated by a thermostat, ideally an auxiliary system with infinite degrees of freedom in contact with our system of particles and acting as a heat reservoir.
Thermodynamic control methods: constant temperature (II)

An original idea of S. Nosé, extended by W.G. Hoover, was that we do not really need infinite degrees of freedom, but only one. This additional degree of freedom takes care of the proper scaling of the velocities, reproduces the canonical ensemble \((N,V,T)\) and reduces the problem of the thermostat to deterministic dynamical equations.

This single degree of freedom, \(s\), does a rescaling of the velocities \(v_I\) during the simulation, 

\[
\dot{R}_I = v_I \rightarrow s \cdot v_I
\]

and is similar to a friction force \(f^c_I = -s p_I\) if \(s > 0\) or to a heating process if \(s < 0\).
Thermodynamic control methods: 
constant temperature (III)

The new single degree of freedom $s$ can be added to the Lagrangean of the molecular dynamics as a new variational variable. The extended Lagrangean then reads

$$L = \frac{1}{2} \sum_{I=1}^{N} M_I s^2 \dot{R}_I^2 - V(R_I) + \frac{1}{2} Q \dot{s}^2 - (3N + 1)k_B T \ln s$$

The new variable $s(t)$ has a kinetic term in which $Q$ represents the fictitious mass, basically the time scaling of the motion of $s(t)$ with respect to the motion of $R_I(t)$. The velocities are rescaled as $sv_I$ and the virtual potential for $s(t)$ is given by the Boltzmann-like canonical term $(3N + 1)k_B T \ln s$, where $3N + 1$ are all the degrees of freedom: $3N$ for the atomic coordinates plus one for the new variable $s$. 
Thermodynamic control methods: constant temperature (IV)

The *(coupled)* Euler-Lagrange equations of motion, under the rescaling of the Nosé-Hoover thermostat become then

\[
M_I \dddot{R}_I = \frac{f_I}{s^2} - 2 \frac{s}{s} M_I \dot{R}_I
\]

\[
Q \ddot{s} = \frac{1}{s} \cdot \left[ s^2 \sum_{I=1}^{N} M_I \dot{R}_I^2 - (3N + 1)k_B T \right]
\]

and these can be solved numerically with the Verlet algorithm. Note that now we have two more input parameters: the target temperature \( T \) and the Nosé-Hoover thermostat mass \( Q \).
What can we get out of a MD simulation?

- Dynamical averages = statistical averages of several physical quantities: total and free energy, molecular velocity distributions, etc.
- Radial distribution function / pair correlation functions $g_{ij}(r)$ and angular distribution functions
- Temperature, pressure (stress tensor), crystal and non-crystal phases, etc…
- Diffusion coefficients
- Vibrational spectra and normal modes
- etc…

Let’s see each point one by one and how we can extract these data from a MD trajectory.
Connection to **Statistical Mechanics:**

time averages and ensemble averages

A successful MD run gives you, in output, a time-ordered sequence of positions $\mathbf{R}_I(t_m)$ and velocities $\mathbf{v}_I(t_m)$ at the discrete time points $t_m = m\delta t, \ m = 1, \ldots, M$ for a total simulation time $\Delta t = M \delta t$.

We call this sequence *trajectory* and the trajectory represents the set of configurations “visited” by our system during the dynamics, i.e. during its motion under the action of the force field that you selected.

So, what do we do with such a sequences of numbers ?

We can use this discrete trajectory to visualize the motion of the particles on most of the graphical PCs, workstations…
...and what do we see on our computer?

Example: the case of water... a droplet of H$_2$O molecules at $T = 300$ K and density of 1 g/cm$^3$
Connection to Statistical Mechanics: time averages and ensemble averages

…but this is just a visual inspection (= no or very little science). In the end we want always to compute ensemble averages or, within the ergodic principle, time averages of some function $A(R_f(t), v_f(t))$ of the positions and velocities, or, equivalently $A(R_f(t), p_f(t))$ of the positions and momenta. These averages are quantities that can be measured experimentally.

**Ergodic principle:** if the dynamics is long enough so that our system can explore its whole phase space $\{R_f, p_f\}$, then time averages are identical to ensemble averages, which means, if $\Delta t \to \infty$

$$\langle A \rangle_{\text{Ensemble}} \equiv \langle A \rangle_{\Delta t} = \frac{1}{\Delta t} \int_0^{\Delta t} A(R(t), p(t)) dt \equiv \frac{1}{M} \sum_{m=1}^{M} A(R(t_m), p(t_m))$$
Connection to **Statistical Mechanics**: a trivial example – **Temperature and Total Energy**

\[
\left\langle E^{Tot} \right\rangle_{\Delta t} = \left\langle \sum_{i=1}^{N} \frac{1}{2} M_i \left[ \mathbf{v}_i(t) \right]^2 + V(\mathbf{R}_i(t)) \right\rangle = \left\langle \sum_{i=1}^{N} \frac{1}{2} M_i \left[ \mathbf{v}_i(t) \right]^2 \right\rangle + \langle V \rangle = \frac{3}{2} N k_B T + \langle V \rangle
\]
Connection to Statistical Mechanics: static properties

If our system has an average density \( \rho (\rho = 1 \text{ g/cm}^3 \text{ for H}_2\text{O}) \), this density fluctuates as a function of the distance from a given point…

\[
\rho (r) = \rho g(r)
\]

The density at a given radial distance \( r \) depends on whether or not another particle is present at \( r \) and is a measure of the structure of the system. The adimensional function \( g(r) \) is called radial distribution function: It is zero at \( r = 0 \) (two particles cannot occupy the same place) and \( \lim_{r \to \infty} g(r) = 1 \) because in the whole volume \( \rho (r) = \rho \).

From a MD trajectory: 
\[
g(r) = \frac{1}{\rho N} \left\langle \sum_{I \neq J}^{N} \delta(r - r_{ij}) \right\rangle
\]
Connection to Statistical Mechanics: extracting properties from simulations

Summarizing:
static properties such as structure, pressure etc. are obtained from $g(r)$ pair (radial) distribution functions

The dynamics of the system can be measured from the displacement of its particles in time or, equivalently, from the velocities. An easy quantity to compute is the mean square displacement (MSD)

\[ \Delta R^2(t) = \frac{1}{N} \sum_{i=1}^{N} [\mathbf{R}_i(t) - \mathbf{R}_i(0)]^2 \]

In general, for a solid \( \Delta R^2(t) \) is small and almost constant in time (non-diffusive regime). Instead, in a liquid or a gas it grows roughly linearly (diffusive regime)

\[ \Delta R^2(t) = 6D t + \Delta R^2(0) \]

\[ \frac{1}{6} \frac{d}{dt} \Delta R^2(t) = D \]

\( D \) is the (self)\textit{diffusion coefficient} to be compared with experiments.
Connection to **Statistical Mechanics:**

dynamic properties

The velocities provided by the MD can be used in a mathematically equivalent way since

\[ \mathbf{R}_I(t) - \mathbf{R}_I(0) = \int_0^t \mathbf{v}_I(t') dt' \]

and if we evaluate the time derivative of the MSD,

\[ \frac{d}{dt} \Delta R^2(t) = \int dt' \int dt'' \langle \mathbf{v}(t') \cdot \mathbf{v}(t'') \rangle = 2 \int dt' \langle \mathbf{v}(t) \cdot \mathbf{v}(t') \rangle \]

\[ = 2 \int dt' \langle \mathbf{v}(t-t') \cdot \mathbf{v}(0) \rangle = 2 \int dt' \langle \mathbf{v}(t') \cdot \mathbf{v}(0) \rangle \]

using the invariance of the origin of time. So, finally,

\[ D = \frac{1}{3} \int dt' \langle \mathbf{v}(t') \cdot \mathbf{v}(0) \rangle \]
Connection to **Statistical Mechanics:**

dynamic properties

The quantity

\[ \langle v(t') \cdot v(0) \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle v_i(t') \cdot v_i(0) \rangle \]

is called velocity autocorrelation function and is useful also to compute the vibrational spectrum of the system

\[ I(\omega) = \int_{0}^{\infty} dt \langle v(t) \cdot v(0) \rangle e^{i\omega t} \]

as a simple Fourier transform from the time domain to the frequency domain.
Connection to **Statistical Mechanics:**

dynamic properties - Vibrational spectrum of water
Connection to **Statistical Mechanics**: extracting properties from simulations

**Summarizing:**
dynamic and transport properties are obtained from time correlation functions

Example: self diffusion coefficient of water

\[ D \text{ (x10}^{-5} \text{ cm}^2/\text{s}) \]

<table>
<thead>
<tr>
<th>MD (good simulation)</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8 ± 0.5</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Suggestions for further readings:

First Principles MD: instead of looking for a potential $V(R_I)$, we try to include quantum electrons and classical nuclei and to compute forces from fundamental quantum mechanics.

Beside the atom-atom (ion-ion) interaction, new ingredients must be included.
Density Functional Theory: brief review

• Define the electronic density $\rho(x)$ as a superposition of single particle Kohn-Sham (KS) orbitals

$$\rho(x) = \sum_{i}^{\text{occ}} f_{i} |\psi_{i}(x)|^{2}$$

• Write the total energy functional as

$$E[\psi_{i}, R_{I}] = E_{k} + E_{H} + E_{xc} + E_{ps} + E_{M}$$

i.e. sum of electron-electron + electron-ion + ion-ion interaction
Density Functional Theory: brief review

• Electron-electron interaction:

\[ E_k = -\frac{1}{2} \sum_i f_i \int d^3 x \psi_i^*(x) \nabla^2 \psi_i(x) \]

\[ E_H = \frac{1}{2} \int d^3 x \, d^3 x' \, \frac{\rho(x) \rho(x')}{|x-x'|} \]

\[ E_{xc} = \int d^3 x \, \epsilon_{xc}(\rho, \nabla \rho) \rho(x) \]

• \( E_k \) = kinetic energy, \( E_H \) = Coulomb interaction, \( E_{xc} \) = exchange-correlation interaction (our way to write what we do not know: the many-body interaction)
Density Functional Theory: brief review

- Electron-ion interaction:

\[ E_{ps} = \int d^3x \, \mathcal{V}^{ps}(x) \rho(x) \]

the core-valence interaction is described by pseudopotentials

- Ion-ion interaction:

\[ E_M = \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \]
H₂O: electronic (KS) states

HOMO-3
-25.50 eV

HOMO-2
-13.21 eV

HOMO-1
-9.38 eV

HOMO
-7.26 eV

LUMO
-1.00 eV

LUMO+1
+0.50 eV

LUMO+2
+1.10 eV

O-H = 0.966 Å
HOH = 104.3°
Electronic structure from QM: the case of water

Lone pairs

σ bond
Car-Parrinello Molecular Dynamics

- In 1985, Roberto Car *(Princeton University)* and Michele Parrinello *(Swiss Federal Institute of Technology)* proposed to use the fundamental quantum mechanics to replace the classical $V(R_I)$ with $E[\psi_i, R_I]$
- And extend the classical MD Lagrangean by adding the electronic degrees of freedom $\psi_i$ plus a constraint for keeping the wavefunctions orthogonality
- And any external additional variables $\alpha_q$ (e.g. thermostats, stress, etc.)

\[
\mathcal{L}^{CP} = \sum_i \mu \int d^3x \, |\dot{\psi}_i|^2 + \frac{1}{2} \sum_I M_I \dot{R}_I^2
\]

\[
+ \frac{1}{2} \sum_q \mu_q \dot{\alpha}_q^2 - E^{DFT}[\psi_i, R_I, \alpha_q]
\]

\[
+ \sum_{ij} \Lambda_{ij} \left( \int d^3x \, \psi_i^* \psi_j - \delta_{ij} \right)
\]
Car-Parrinello Molecular Dynamics

• Solve the related Euler-Lagrange EOM

\[ \mu \ddot{\psi}_i = -\frac{\delta E^{DFT}}{\delta \psi^*_i} + \sum_j \Lambda_{ij} \psi_j \]

\[ M_I \ddot{R}_I = -\nabla_{R_I} E^{DFT} \]

\[ \mu_q \ddot{\alpha}_q = -\frac{\partial E^{DFT}}{\partial \alpha_q} \]

The difference between the CP trajectories $\mathbf{R}_I^{\text{CP}}(t)$ and the Born-Oppenheimer (BO) ones $\mathbf{R}_I^{\text{BO}}(t)$ is bound by

$$| \mathbf{R}_I^{\text{CP}}(t) - \mathbf{R}_I^{\text{BO}}(t) | < C \mu^{1/2}$$

($C > 0$) if

$$\omega_0 = \sqrt{2 \cdot (\varepsilon^{\text{LUMO}} - \varepsilon^{\text{HOMO}})} / \mu > 0$$

Practical implementation

• To implement the CP-EOM numerically, the KS orbitals are generally expanded in plane waves

\[ \psi_i(x) = \sum_G c_i(G) e^{iGx} \]

• \( G \) are the reciprocal space vectors. The Hilbert space spanned by PWs is truncated to a suitable cut-off \( E^{\text{cut}} \) such that

\[ G^2/2 < E^{\text{cut}} \]

• PWs have so far been the most successful basis set, in particular for extended systems requiring periodic boundary conditions
• The basis set accuracy can be systematically improved in a fully variational way
• PWs do not depend on atomic positions (no Pulay forces)
• However they are \textit{not} localized and can be inefficient, e.g., for small molecules or clusters in a large simulation cell
Plane wave expansion: $\psi_i(x) = \sum_G c_i(G) e^{iGx}$

For each electron $i=1,\ldots,N$, $G=1,\ldots,M$ are the reciprocal space vectors. The Hilbert space spanned by PWs is truncated to a cut-off $G_{\text{cut}}^2/2 < E_{\text{cut}}$
Practical implementation

• Verlet’s algorithm on EOM gives
  \[
  \left( \frac{\mu}{\Delta t^2} \right) \cdot \left[ |\psi_i(t+\Delta t)\rangle + |\psi_i(t-\Delta t)\rangle - 2 |\psi_i(t)\rangle \right] = - \left( H_{\text{CP}} + \Lambda \right) |\psi_i(t)\rangle
  \]
  or, in \(G\)-space,
  \[
  \frac{\mu}{\Delta t^2} \left[ c_i(\mathbf{G}, t + \Delta t) + c_i(\mathbf{G}, t - \Delta t) - 2 c_i(\mathbf{G}, t) \right] = - \sum_{\mathbf{G}'} \langle \mathbf{G} | H_{\text{CP}} | \mathbf{G}' \rangle c_i(\mathbf{G}') + \sum_j \Lambda_{ij} c_j(\mathbf{G})
  \]

• The ionic degrees of freedom \( \mathbf{R}_i(t) \) are updated at a rate (speed) \( \Delta t \) while the electronic degrees of freedom \( |\psi_i(t)\rangle \) are updated at a rate \( \Delta t/\mu^{1/2} \) (\( \Delta t \sim 5 \) a.u., \( \mu \sim 500 \) a.u.), hence they are much slower (adiabatic) with respect to the ions.

• The parameter that controls the adiabaticity (BO) is \( \mu \)

• However this \( \Delta t \) is driven by QM and is much shorter than \( \delta t \) in classical MD
What can we get out of a CPMD simulation?

• Dynamical averages of several physical quantities: total and free energy, molecular velocity distributions, etc.
• Radial distribution function / pair correlation functions $g_{ij}(r)$
• Diffusion coefficients and vibrational spectra
  …as in classical MD.

So what’s new?

• **Electronic structure** evolution during the dynamics
• **Infrared** (from dipole autocorrelation function) and **Raman** spectra
• **NMR** chemical shifts and **hyperfine ESR parameters**
• Breaking and formation of **chemical bonds** if the barrier is zero or very small ($\sim k_B T$)
Suggestions for further readings:

Lecture 2: Reactive first principles molecular dynamics

Mauro Boero

Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 Japan
An easy “reaction”: breaking and formation of H-bonds

When water aggregates to form a liquid state (or solid ice), molecules stick together by forming hydrogen bonds (H-bonds):

The energy required to break one H-bond is 2.6 kcal/mol in normal liquid water*. These small barriers can easily be overcome in any MD simulation and well reproduced also by classical force fields.

...and possible H-bond configurations in H$_2$O

**Linear H-bond (LHB)**
Dipole Moment = 2.90 D

**Cyclic H-bond (CHB)**
Dipole Moment = 0.23

**Bifurcated H-bond (BHB)**
Dipole Moment = 4.38 D

**2-fold H-bond (2fHB)**
Dipole Moment = 1.80 D
When do we see these H-bonds?

Average number of the different H-bond configurations per 32 molecules of H$_2$O at different thermodynamic conditions

<table>
<thead>
<tr>
<th>Configuration</th>
<th>$\rho=0.32$ g/cm$^3$</th>
<th>$\rho=0.73$ g/cm$^3$</th>
<th>$\rho=1.00$ g/cm$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T = 647$ K</td>
<td>$T = 653$ K</td>
<td>$T = 300$ K</td>
</tr>
<tr>
<td>$sc$H$_2$O @ low density</td>
<td></td>
<td></td>
<td>Normal liquid water</td>
</tr>
<tr>
<td>CHB</td>
<td>0.615 ± 0.013</td>
<td>1.863 ± 0.021</td>
<td>0.015 ± 0.001</td>
</tr>
<tr>
<td>BHB</td>
<td>0.452 ± 0.008</td>
<td>0.320 ± 0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>2fHB</td>
<td>0.260 ± 0.012</td>
<td>1.601 ± 0.016</td>
<td>0.525 ± 0.011</td>
</tr>
</tbody>
</table>

$\Delta E(\text{CHB-LHB}) = 2.2$ kcal/mol (local minimum)

$\Delta E(\text{BHB-LHB}) = 2.8$ kcal/mol (local minimum)
An example of dynamical chemical bond breaking and formation: first principles MD of the diffusion of protons in water (electronic structure becomes important…)

The presence of a hydrogen bond network linking water molecules is crucial to propagation

A proton in normal liquid water:
\[ \rho = 1.0 \text{ g/cm}^3 \]
\[ T = 300 \text{ K} \]

simulation time displayed: 0.6 ps

total simulation time: 19.1 ps
The excess proton makes a continuous switch between Eigen $\text{H}_9\text{O}_4^+$ complex and Zundel $\text{H}_5\text{O}_2^+$ shared proton structure, forming and breaking bonds
Proton diffusion: a couple of observations

• A simple first principles MD seems to be able to simulate the basic chemistry process for a proton diffusion: the Grotthuss switching from a $\sigma$ bond (O-H) to a hydrogen bond (H-bond)
• It seems than that if we can control in some way the temperature of the water, we can speed up (heating) or slow down (cooling) such a process [Note: this is the idea behind the use of supercritical water that we shall discuss later]
• Indeed, for simple chemical reactions that occur with little (few $k_B T$) or no energy barrier, thermal fluctuations are enough to activate the process
• Yet, $k_B T = 26$ meV $= 0.60$ kcal/mol at $T = 300$ K

…so let’s try to heat up our system from 300 K to 673 K*

* Typical supercritical temperature
High temperature:
\[ \rho = 1.0 \text{ g/cm}^3 \]
\[ T = 673 \text{ K} \]

simulation time displayed: 0.6 ps
(total simulation time: 20.0 ps)

the H\(^+\) diffusion becomes faster
Theoretical diffusion coefficients computed from the velocity-velocity autocorrelation function. $D$ and from the mean square displacement, $D^{MSD}$. The experimental proton diffusion coefficient is $D_{\text{exp}}$, diffusion and $D^{\text{self}}$ is the water self-diffusion coefficient.

<table>
<thead>
<tr>
<th>$\rho$ (g cm$^{-3}$)</th>
<th>$T$ (K)</th>
<th>$D$ (cm$^2$/s)</th>
<th>$D^{MSD}$ (cm$^2$/s)</th>
<th>$D_{\text{exp}}$ (cm$^2$/s)</th>
<th>$D^{\text{self}}$ (cm$^2$/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>300</td>
<td>$13.0 \times 10^{-5}$</td>
<td>$15.0 \times 10^{-5}$</td>
<td>$9.3 \times 10^{-5}$ [a]</td>
<td>$2.5 \times 10^{-5}$</td>
</tr>
<tr>
<td>1.0</td>
<td>673</td>
<td>$70.0 \times 10^{-5}$</td>
<td>$62.0 \times 10^{-5}$</td>
<td></td>
<td>$33.0 \times 10^{-5}$</td>
</tr>
</tbody>
</table>


*However this is just a lucky case... in general a chemical reaction requires the overcoming of a much higher energy barrier.*
From a general reactant A to a product B, we have to climb above a (more or less) high mountain

- A general chemical reaction starts from reactants A and goes into products B
- The system spends most of the time either in A and in B
- …but in between, for a short time, a barrier is overcome and atomic and electronic modifications occur
- Time scale: \( \tau \sim \tau_{mol} e^{k_B T} \)

\[ F^* \sim \Delta \]
How to simulate a chemical reaction?

The reaction $A + B \rightarrow AB$ passes across a rare event (transition state) that is difficult to see.

CP molecular dynamics time scale is of the order of some tens of ps and classical simulations are of the order of few ns, but still insufficient… we cannot wait forever…
Select the reaction coordinate $\xi$ to be sampled, e.g. the interatomic distance $\xi = |\mathbf{R}_A - \mathbf{R}_B|$

Add to the CP lagrangean $L^{CP}$ the holonomic constraint

$$L^{CP} \rightarrow L^{CP} + \lambda_\xi (\xi - \xi_0) \quad (\lambda_\xi = \text{Lagrange multiplier})$$

Compute the ensemble (time) average $\langle \lambda_\xi \rangle$ why?
Blue Moon ensemble theory (II)

From the standard definition of free energy

\[ F = -k_B T \ln \langle \exp \{-[H_{CP} + \lambda_\xi (\xi - \xi_0)]/k_B T \} \rangle \]

using \( \exp(x) \sim 1 + x + O(x^2) \) and \( \ln(1+x) \sim x - O(x^2) \) we get

\[ \delta F \approx -k_B T \delta \ln \left\langle 1 - \frac{H_{CP}}{k_B T} - \frac{\lambda_\xi (\xi - \xi_0)}{k_B T} \right\rangle \]

\[ = -k_B T \delta \left\langle - \frac{H_{CP} - \lambda_\xi (\xi - \xi_0)}{k_B T} \right\rangle / k_B T = \delta \left\langle H_{CP} + \lambda_\xi (\xi - \xi_0) \right\rangle \]

\[ = \delta \left\langle H_{CP} \right\rangle + \delta \lambda_\xi \left\langle \xi - \xi_0 \right\rangle + \left\langle \lambda_\xi \right\rangle \delta (\xi - \xi_0) \]

but \( \delta \left\langle H_{CP} \right\rangle = 0 \) because \( H_{CP} \) is a constant of motion

and \( \left\langle \xi - \xi_0 \right\rangle = \left\langle \xi \right\rangle - \xi_0 = \xi_0 - \xi_0 = 0 \), \( \delta (\xi - \xi_0) = \delta \xi \)

because \( \xi_0 \) is the average value of \( \xi \) (and a constant !)
Blue Moon ensemble theory (III)

- ...so the variation of the free energy become simply

$$\delta F = \frac{\partial F}{\partial \xi} \delta \xi = \langle \lambda_\xi \rangle \delta \xi$$

- If we now integrate the average constraint force along the sampled reaction path between $\xi_{\text{init}}$ and $\xi_{\text{final}}$, we finally arrive at the following expression for the free energy difference between initial (reactant) and final (product) states

$$\Delta F = \int_{\xi_{\text{init}}}^{\xi_{\text{final}}} \langle \lambda_\xi \rangle d\xi$$
Metadynamics (I): reactive CPMD with several collective variables

1) The atomic and electronic configuration of our initial system is given by a set of variables

\[ \{ R_1, \ldots, R_N \}_{t=0} \oplus \{ \psi_1, \ldots, \psi_M \} \]

2) And we assume that some known functions of few of them (collective variables) are necessary and sufficient to describe the process we are interested in

\[ s_\alpha (R_I ; \psi_i) \quad \alpha = 1, \ldots, n < N, M \]

3) …so that the FES is a function of these changing variables

\[ F(s) \quad s \equiv s(t) = \{ s_\alpha (t) \}_{\alpha=1,\ldots,n} \]
Metadynamics (II)

- Artificial dynamics in the space of a few collective variables [1]
- The CPMD dynamics is biased by a history-dependent potential constructed as a sum of Gaussians [2].
- The history dependent potential compensates, step after step, the underlying free energy surface [3,4].


What is it used for?

- For reconstructing $F(s)$ in many dimensions as a function of one or more collective variables.
- Used as a tool for escaping local (free energy) minima, i.e. accelerating rare events, reconstructing the free energy in the selected interval of reaction coordinates (not everywhere!).
Metadynamics (III): how does it work?

• Put a “small” Gaussian
• The dynamics brings you to the closest local minimum of $F(s)$ plus the sum of all the Gaussians
Metadynamics (IV): summarizing…

In the one dimensional example shown, the system freely moves in a potential well (driven by MD).

• Adding a penalty potential in the region that has been already explored forces the system to move out of that region,

• …but always choosing the minimum energy path, i.e. the most natural path that brings it out of the well.

• Providing a properly shaped penalty potential, the dynamics is guaranteed to be smooth

• Therefore the systems explores the whole well, until it finds the lowest barrier to escape.
Metadynamics (V):
plugging the method in CPMD – yet another extended Lagrangean

\[ \mathcal{L} = \mathcal{L}_{CP} + \sum_{\alpha} \frac{1}{2} M_\alpha \dot{s}_\alpha^2 - \sum_{\alpha} \frac{1}{2} k_\alpha (S_\alpha(R) - s_\alpha)^2 - V(s, t) \]

Fictitious kinetic energy

Restrain potential: coupling fast and slow variables
\[ \sqrt{(k_\alpha / M_\alpha)} \ll \omega_1 \]

\[ \frac{\partial F}{\partial s_\alpha} \approx \langle k_\alpha (S_\alpha(R) - s_\alpha) \rangle \]

History-dependent potential
Metadynamics (VI): Equations of motion for the collective variables

We use (again) the velocity Verlet algorithm to solve the EOM

\[ s_\alpha(t + \delta t) = 2s_\alpha(t) - s_\alpha(t - \delta t) + \frac{1}{2} \delta t^2 \frac{f_\alpha(t)}{M_\alpha} \]

and we have two new contributions to the force

\[ f_\alpha(t) = k_\alpha (S_\alpha(R, t) - s_\alpha(t)) + \frac{\partial V(t, s)}{\partial s_\alpha(t)} \]
Metadynamics (VII): note that...

1. The parameter $t$ in $s_\alpha(t)$ is not a real time, but rather an index labeling the configurations that explore the FES
2. $\Delta s_\alpha = s_\alpha(t+\delta t) - s_\alpha(t)$ can be arbitrarily large (coarse-grained metadynamics) and gives the accuracy in the sampling of the FES along the trajectory described by the $s_\alpha(t)$ variables
3. The dynamics at each fixed $s_\alpha$ is a true dynamics used to explore the local portion of FES
4. The FES is smoother than the PES, being dependent on a reduced number of variables
Metadynamics (VIII):
Discrete form of $V(s,t)$ used in the implementation

$$V(s,t) = \sum_{t_i < t} W_i \exp\left(-\frac{1}{2} \left(\frac{(s - s_i)^2}{(\Delta s_i^\perp)^2}\right)\right) \exp\left[-\frac{1}{2} \left(\frac{(s_{i+1} - s_i) \cdot (s - s_i)^2}{(\Delta s_i^\parallel)^4}\right)\right]$$

where the Dirac $\delta$ has been expressed in the approximate Gaussian form

$$\delta(x) \approx \frac{1}{\beta \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{x^2}{\beta^2}\right) \quad \beta = |\dot{s}(t)| \Delta t$$

and the discrete time step $\Delta t$ must be such that

CPMD time step $\Delta t^{CPMD} \ll \Delta t \ll \omega_s^{-1}$

Highest oscillation frequency of $s_\alpha$
Metadynamics (IX):
FES reconstruction, i.e. what $V(s,t)$ is used for

When the (meta)dynamics is over and the walker has explored all the portion of the \{s\} phase space available, we have completed our job (at large $t$) and filled all the local minima, then the shape of $V(s,t)$ is similar to the FES apart from a sign and an arbitrary additive constant

$$\lim_{t \to \infty} V(s,t) = -F(s) + const.$$ 

In practice: the number of gaussians required to fill a minimum is proportional to $(1/\delta \sigma)^n$ ($n =$ dimensionality of the problem) and

$$W / \delta \sigma e^{-1/2} = \gamma \langle f_{\alpha}^2 \rangle^{1/2} \quad \gamma \approx 0.5.$$
Metadynamics (X):
Example of FES exploration by two collective variables
$F(s_1, s_2)$ (kcal/mol)

(Example from a true simulation on ATP hydrolysis)

$\Delta F^\# = F^{\text{ATP}} - F^{\text{TS}} = 3.0 \text{ kcal/mol}$

$\Delta F = F^{\text{ATP}} - F^{\text{ADP}} = 6.9 \text{ kcal/mol}$

(exp. 7.1 kcal/mol)
Metadynamics (XI):

caveat

• Parameters are generally system dependent
• The collective variables $s_\alpha$ must be identified by the user
• …and must be able to discriminate between initial and final state
• …and must keep into account all the slow degrees of freedom
• Although the trajectory generated describes the most probable reaction pathway, it is not the true dynamics
Note: connection between metadynamics and Blue Moon

In the case of a single collective variable (reaction coordinate), $\alpha = 1$

\[
\mathcal{L} = \mathcal{L}_{CP} + \frac{1}{2} M \dot{s}^2 - \frac{1}{2} k (s(R_I) - s)^2 - V(s, t)
\]

• Do not update the Gaussians: $V(s, t) \rightarrow 0$

• Do not move dynamically $s(t)$: $\dot{s} = 0$

\[
\mathcal{L} = \mathcal{L}_{CP} - \lambda (s(R_I) - s)^2
\]

where $\lambda = k/2$.

…Then we are back to a Blue Moon formulation.
Further readings:

• M. B. et al., *J. Chem. Theory Comput.* **1**, 925 (2005); CPMD metadynamics
Lecture 3: Water as a catalyst

Mauro Boero

Center for Computational Sciences, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 Japan
Main interests in scH$_2$O

- Water above its critical point $T$=654 K, $P$=22.1 MPa, $\rho$=0.32 g/cm$^3$ is found in nature inside the earth mantel and in the “black jets” of deep oceanic trenches (like the Japan trench).

- scH$_2$O has important technological applications:
  
  i) it is an advanced tool for the treatment of hazardous wastes such as dioxin and PCB [1]

  ii) it promotes non-catalytic chemical reactions and can replace non-polar solvent and acid catalysts which are environmentally dangerous [2]

Hydrogen-bond network in water

Normal water
\[ T = 300 \, \text{K} \]
\[ \rho = 1.00 \, \text{g/cm}^3 \]

Supercritical water:
Disrupted H-bond NW
\[ T = 653 \, \text{K} \]
\[ \rho = 0.73 \, \text{g/cm}^3 \]

\[ scH_2O \]
\[ T > 647 \, \text{K} \]
\[ P > 22.1 \, \text{MPa} \]
\( sc\text{H}_2\text{O} :\)
\[ \rho = 0.66 \text{ g/cm}^3 \quad T = 673 \text{ K} \]

simulation time displayed: 0.6 ps

total simulation time: 18.4 ps
Supercritical water: $\rho = 0.66 \text{ g/cm}^3$, $T = 673 \text{ K}$

A more complex and varied scenario arises in SCW and the proton displaces faster than in the normal liquid state

Eigen $(H_7O_3)^+$  unstable $(H_5O_2)^+$  Zundel $(H_5O_2)^+$
Eigen complex pair correlation functions
Computing the diffusion of protons water:

In general a proton $H^+$ is transferred to a $H_2O$ molecule but it is a different one that subsequently leaves the $H_3O^+$

$$H^+ + H-O-H \rightarrow H-OH_2 \rightarrow H-O-H + H^+$$

We then follow the center of mass of the Eigen/Zundel complex:

$$R_{EZ}(t_i) = \frac{1}{M_{EZ}} \sum_{i=1}^{N_{EZ}} M_i R_i(t_i)$$

where $N_{EZ}$ is the number of atoms forming the Eigen or Zundel complex, respectively. $M_{EZ}$ is of course the related total mass

In practice, $R_{EZ}(t_i)$ corresponds to the position of the central $O$ atom in the case of Eigen complex and to the position of the shared proton in the case of Zundel complex.
Computing the diffusion of protons water:

- We then constructed the velocity of the \textit{migration of the structural defect} as

\[ v_{EZ}(t_i) = \frac{R_{EZ}(t_{i+1}) - R_{EZ}(t_i)}{t_{i+1} - t_i} \]

- Then computed the \textbf{diffusion coefficient}, by using the Einstein relation expressed in terms of the Green-Kubo formula, assuming this velocity definition in the autocorrelation function

\[ D = \frac{1}{3} \int_0^{\Delta t} \langle v_{EZ}(t) \cdot v_{EZ}(0) \rangle \]

being $\Delta t$ half of the simulation time ($\sim 9$ ps) and as

\[ D^{MSD} = \frac{1}{6} \lim_{t \to \infty} \frac{d}{dt} \left\langle \left| R_{EZ}(t) - R_{EZ}(0) \right|^2 \right\rangle \]
Let’s complete our table of theoretical diffusion coefficients computed from the velocity-velocity autocorrelation function. \( D \) and from the mean square displacement, \( D^{MSD} \). The experimental proton diffusion coefficient is \( D_{exp} \), diffusion and \( D_{self} \) is the water self-diffusion coefficient.

<table>
<thead>
<tr>
<th>( \rho ) [g cm(^{-3})]</th>
<th>( T ) [K]</th>
<th>( D ) [cm(^2) s(^{-1})]</th>
<th>( D^{MSD} ) [cm(^2) s(^{-1})]</th>
<th>( D_{exp} ) [cm(^2) s(^{-1})]</th>
<th>( D_{self} ) [cm(^2) s(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>300</td>
<td>13.0 \times 10^{-5}</td>
<td>15.0 \times 10^{-5}</td>
<td>9.3 \times 10^{-5} [a]</td>
<td>2.5 \times 10^{-5}</td>
</tr>
<tr>
<td>1.00</td>
<td>673</td>
<td>70.0 \times 10^{-5}</td>
<td>62.0 \times 10^{-5}</td>
<td>-</td>
<td>33.0 \times 10^{-5}</td>
</tr>
<tr>
<td>0.66</td>
<td>673</td>
<td>54.0 \times 10^{-5}</td>
<td>55.0 \times 10^{-5}</td>
<td>49.0 \times 10^{-5} [b]</td>
<td>54.0 \times 10^{-5}</td>
</tr>
<tr>
<td>0.32</td>
<td>673</td>
<td>37.0 \times 10^{-5}</td>
<td>41.0 \times 10^{-5}</td>
<td>17.0 \times 10^{-5} [b]</td>
<td>177.0 \times 10^{-5}</td>
</tr>
</tbody>
</table>


The experimental value closest to 0.66 g cm\(^{-3}\) is \( \rho = 0.63 \text{ g cm}\(^{-3}\) and \( T = 623.15 \text{ K} \).
Beckmann rearrangement for $\varepsilon$-Caprolactam

1. **Commercially important** for production of synthetic fibers (nylon 6). About 4 million tons per year.

2. Known to be catalyzed only by **strong acids** in conventional **non-aqueous** systems.

3. Formation of **byproducts** (ammonium sulfate, $(\text{NH}_4)_2\text{SO}_4$) of low commercial value in acid catalyst: byproducts = 1.7 × products (in weight).

4. Environmentally **harmful**: acid wastes are produced.

**Points 2, 3 and 4 and related problems can be eliminated in scH$_2$O: no acid required & no byproducts.**

Beckmann rearrangement reaction

proton attack to O

in strong acid and supercritical H₂O

in supercritical H₂O

hydrolysis in H₂O and superheated H₂O

cyclohexanone oxime

proton attack to N

cyclohexanone (+NH₂OH+H⁺)

ε-caprolactam
Schematic outline of reaction mechanism

oxime
R–C–R’
||
N–OH
H^+

amide
H
R–N=C–R’
||
R–N=C+=R’

H_2O
R–N=CR’
+ H_2O

H_2O
R–N=CR’

H
R–N=CR’

H
R–N=CR’
Which are the important ingredients that make water special at supercritical conditions?

- Proton attack is the trigger (experimental outcome!)

**High efficiency:** fast proton diffusion

**High selectivity:** difference in hydration between O and N
Computational details

• 1 cyclohexanone oxime immersed in a system of 60 H$_2$O molecules
• DFT-based first principles molecular dynamics within BLYP gradient corrected approach
• Systems: $T= 673$ K, $\rho =0.66$ g/cm$^3$, $L=14.262$ Å : \textcolor{red}{supercritical}
  \hspace{1cm} $T= 673$ K, $\rho =1.00$ g/cm$^3$, $L=12.430$ Å : \textcolor{red}{super heated}
  \hspace{1cm} $T= 300$ K, $\rho =1.00$ g/cm$^3$, $L=12.430$ Å : \textcolor{red}{normal}
• Temperature control via Nosé-Hoover thermostat chain
• Integration step of 4.0 a.u. (0.0967 fs) and fictitious electronic mass of 400 a.u.
• Reaction path sampling within both \textcolor{red}{metadynamics} and \textcolor{red}{Blue Moon}
Cyclohexanone-oxyme in $sc\text{H}_2\text{O}$:

- The energy barrier seems rather high and the reaction pathway not unique
- The reaction is generally acid catalyzed, hence protons are expected to be essential in triggering the process
- At supercritical conditions, however, the $K_w$ of water increase, hence $H^+$ and $OH^-$ can be around in the solvent in non-negligible concentration
- And small amounts of weak acids greatly enhance reaction rates
Cyclohexanone-oxime in scH₂O in presence of an excess proton H⁺:

\[ T = 673 \text{ K} \]
Cyclohexanone-oxyme in $s\text{c}H_2O$ (+ $H^+$)

A very small activation barrier (about 1 kcal/mol) is required for the $N$ insertion process.
Contrary to normal liquid water, \( sc\text{H}_2\text{O} \) accelerates selectively the formation of the first intermediate.

\[ \sim 0.9 \text{ ps in } sc\text{H}_2\text{O} \]

\[ \sim 5.2 \text{ ps in } n\text{-H}_2\text{O} \]
Superheated water: 
($\rho = 1.00 \text{ g/cm}^3$) 

$\sim 3.4 \text{ ps at } 673 \text{ K}$ 

Proton attacks N 

Cycrohexanon (byproduct) formation
Efficient reaction in $scH_2O$ due to fast proton diffusion and acid properties of the (broken) Eigen-Zundel complexes

Acid catalyst?
Proton diffusion in ordinary liquid and supercritical water

- Hydrogen bond network is disrupted in SCW.

Is the proton diffusion slowed down in \( \text{scH}_2\text{O} \)? Not really…
Proton diffusion: normal water and supercritical water

Proton (structural defect) diffusion coefficient estimation in the 3 systems:

<table>
<thead>
<tr>
<th>System</th>
<th>Diffusion constant $D$ (cm$^2$/s)</th>
<th>Hydrogen bond network</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-H$_2$O (normal water)</td>
<td>$15.0 \times 10^{-5}$</td>
<td>continuous</td>
</tr>
<tr>
<td>Superheated H$_2$O</td>
<td>$62.0 \times 10^{-5}$</td>
<td>continuous (fast switch)</td>
</tr>
<tr>
<td>scH$_2$O (supercritical water)</td>
<td>$55.0 \times 10^{-5}$</td>
<td>disrupted</td>
</tr>
</tbody>
</table>

In scH$_2$O the network is disrupted and the motion occurs in sub-networks that join and break apart rapidly due to density fluctuations; two diffusion regimes are cooperating: hydrodynamics (vehicular) and Grotthuss
complex formation in the process of proton diffusion in normal water

The excess proton makes a continuous switch between Eigen $\text{H}_9\text{O}_4^+$ complex and Zundel $\text{H}_5\text{O}_2^+$ shared proton structure and turns out to be rather stabilized. Life time $\sim 1.0$ ps.
Unstable complex in \( sc\text{H}_2\text{O} \) (\( \rho = 0.66 \text{ g/cm}^3 \))

The Eigen \( \text{H}_9\text{O}_4^+ \) complex is incomplete (\( \text{H}_7\text{O}_3^+ \)) most of the time due to the disrupted H-bond network and the stabilization of the excess proton in the H-bond sub-network is reduced.

Pseudo-Eigen \[ \rightarrow \]
Zundel

life time: \( \sim 0.1 \text{ ps} \)

The Eigen \( \text{H}_9\text{O}_4^+ \) complex is incomplete (\( \text{H}_7\text{O}_3^+ \)) most of the time due to the disrupted H-bond network and the stabilization of the excess proton in the H-bond sub-network is reduced.
Reaction selectivity?

Selective reaction in scH₂O due to different solvation of O and N
Cyclohexanone-oxyime in $scH_2O (+ H^+)$: 

*the selectivity*

$T = 673 \, \text{K}$
Comparison of the pair correlation functions

(a) = n-H₂O, (b) = superhetaed H₂O, (c) = scH₂O

Depletion (hole around N)
...and now the second step: C-O bond formation

oxime

\[ R-C-R' \]

\[ N-OH \]

\[ H^+ \]

\[ R \rightarrow C \rightarrow R' \]

\[ N \rightarrow OH \]

\[ H_2O \]

\[ R-N=\text{C}^+\rightarrow R' \]

\[ R-N=C+\rightarrow R' \]

\[ H \]

\[ \text{OH}^- \]

\[ R-N=C-\rightarrow R' \]

\[ R-N=C^+-\rightarrow R' \]
...and now the second step: Blue Moon approach

$\Delta E = 5.9 \text{ kcal/mol}$
$\Delta F = 5.1 \text{ kcal/mol}$

Approach of an $\text{H}_2\text{O}$ molecule, $\xi = |O_{\text{wat}}-C|$
The last step: eventually the ε-caprolactam
The final step: Solvent-solute proton exchange via metadynamics

- Collective variables used:
  \( s_1(N-H_w) = \) coordination number of N with H of water molecules of \( scH_2O \)
  \( s_2(O_w-H) = \) coordination number of O of \( scH_2O \) with H of the solute

- \( M_\alpha = 40.0 \) a.u. for both \( s_1(N-H_w) \) and \( s_2(O_w-H) \)
  \( k_\alpha = 0.6 \) a.u.

- \( V(s_\alpha,t) \) updated every 100 CPMD steps
  \( \Delta s_\perp = 0.1 \)
Proton exchange with \textit{water}
Total energy: a very rugged landscape
Free energy surface: a less rugged landscape
Water solvation and proton diffusion: what did we learn?

- Proton diffuses with proton wire mechanism mediated by H bond. For an atom to be attacked by a proton, the atom has to be solvated.
- In normal water and superheated water, both O and N are solvated by water molecules. byproducts
- In supercritical water, only O is solvated and N is dry. desired products only
- Moreover, solvation water around N prevents N insertion into ring.

\[ \text{R—C—R'} \quad \text{scH}_2\text{O} \]

\[ \text{R—C—R'} \quad \text{N}^+ \quad \text{H—O} \quad \text{H} \quad \text{normal H}_2\text{O} \]

\[ \text{R—C—R'} \quad \text{N}^+ \quad \text{H} \quad \text{super heated H}_2\text{O} \]
Further readings:

Acknowledgements

• Michele Parrinello, ETHZ-USI and Pisa University
• Roberto Car, Princeton University
• Kiyoyuki Terakura, JAIST, AIST and Hokkaido University
• Michiel Sprik, Cambridge University
• Pier Luigi Silvestrelli, Padova University
• Alessandro Laio, SISSA, Trieste
• Marcella Iannuzzi, Zurich University
• Michael L. Klein, University of Pennsylvania
• Giacinto Scoles, ICTP, Trieste
• Erio Tosatti, SISSA, Trieste
• Roger Rousseau, SISSA, Trieste
• Sandro Scandolo, SISSA, Trieste

…and all of you for listening despite the rough, incomplete and approximate level of these lectures.