Ab intio molecular dynamics

So far all simulation have been static or either minima or transition state search. No real dynamics of the molecules have been considered. The ab initio molecular dynamics is still easy since we can use the BO approximation (the electronic Schrödinger equation can be solved with fixed atomic positions) which allows the classical treatment of the atoms. Now the dynamics can be described with Newton's equation

$$F = ma$$

With given atom configuration we can compute the forces and update the atomic positions with some algorithm. One of the simplest one is the Verlet algorithm

$$R_I(t + \Delta t) = 2R_I(t) - R_I(t - \Delta t) + \frac{F_I(t)}{m_I}\Delta t^2 + O(\Delta t^4)$$

The main point is that the position updates have to be done very often. The Δt is very small, **typically 1 fs**. To simulate 1s we need 10^{15} steps! This is not possible but of course shorter simulation times are possible. In AIMD the simulation time scales are ca. 50 ps (and up to 1 ns, the desired computer time for time step is less than 10s). Naturally the size of the system affects a lot of the total time scale. In AIMD systems of few 100 water molecules can be simulated. This is hard to improve due to the N^3 scaling of the DFT methods. The linear scaling DFT methods are not here yet.

The method above is called AIMD or BO-AIMD. Its main bottle neck is the wave function optimization since at every MD step one need to solve the DFT (or equivalent) equations. It is essential that the optimizer is fast and the information of the previous MD steps are used effectively.

We can mark the optimized as $\psi_{new} =$ Opt$(\psi_n; R_I)$ which solves the new wave functions corresponding the atomic positions R and a guess wave function ψ_n. In AIMD the simplest case is

$$\psi_n(t + dt) = Opt[\psi_n(t), R_I(t + dt)]$$

But this is not very effective. It is better to extrapolate the guess wave function. With two wave functions

$$\tilde{\psi}_n = 2 * \psi_n(t) - \psi_n(t - dt) + O(dt^2); \quad \psi_n(t + dt) = Opt[\tilde{\psi}_n, R_I(t + dt)]$$

Or more general (typically: n=3,4)

$$\tilde{\psi}_n = P[\psi_n(t), \psi_n(t - dt), \psi_n(t - 2dt), .., \psi_n(t - ndt)]; \quad \psi_n(t + dt) = Opt[\tilde{\psi}_n, R_I(t + dt)]$$

If the electronic structure of the system is difficult to converge the AIMD is slow. The insulators and semiconducting materials are easy but metals are not.

There is also another way to do AIMD. In 1985 R. Car and M. Parrinello suggest so called Car-Parrinello (CP) method. In this methods all the parameters (atoms and electronic) are treated as dynamical variables (they can be updated with Newton's equations). The electronic parameters are the coefficients of the basis set $\psi_n(t) = \sum_i C_{n,i}(t)\xi_i(R)$. Now one can compute the forces for all parameters (E_{KS} is the Kohn-Sham energy)

$$R_I = -\frac{\partial E_{KS}}{\partial R_I}, \quad F(C_i) = -\frac{\partial E_{KS}}{\partial C_i} + \sum_j \Lambda_{ji}C_j; \quad F = M\ddot{R}, \quad F(C) = \mu\ddot{C}$$

The additional terms in the electronic force comes from the wave functions orthonormality constraints. If one optimize the wave functions at the beginning one can follow the CP dynamics quite a few steps. The CP method works surprisingly well for insulators and semiconductors. Even metallic systems can be done but one need to keep the "temperature" of the electronic system in control. The key parameter in the CP method is the mass μ of the electronic parameters. This need to be much smaller that the mass of atoms. Typically value of 300 m_e (mass of electron) is used. Then the adiabatic effect is valid. This means that the light particles will follow the movements of heavier particles. (A classic example it the planetary system.) Unfortunately, the light mass force the usage of small time step. The CP time step is typically 0.1-0.15 fs. In CP one electronic update is needed so if BO-AIMD method require less than ca. 10 iterations then it is faster. The algorithms used in BO-AIMD has improved and it is now better than CP. CP has also some problems with slow heating of the electrons, so CP is not used much. The CP is also problematic with atom centered basis functions.

The 100 ps time scale is very short for most of the chemical reactions. If realistic reaction attempt frequency is 1 ps, in

room temperature (RT=2.4 kJ/mol) barrier of 11 kJ/mol will lead to reaction time scale of 100 ps ($\tau_0 \exp(-\Delta E/RT)$). Seldom it is realistic to do AIMD simulations for chemical reaction and some constrained methods are needed. But there are several systems in which the molecular movement is important. Of these the most important are liquids. In liquids it is essential that the molecules will move. At the moment only liquids consist of small molecules like water, methanol, ammonia, (ethanol), are possible to simulate for reasonable time scales. Of course the key component is the number of electrons. In water there is 8 valence electrons, in methanol (14 valence e), ethanol (20 e), di-cloro-ethane (26 e). These are doable but systems like ionic lquids are not (well even they have been simulated with AIMD).

DFT Approximations

The AIMD simulations are mostly done with DFT-GGA models (BLYP, PBE). The hybrid simulations, like B3LYP, PBE0 are possible but they are time consuming. It is almost mandatory to use the empirical van der Waals corrections (Grimme generation 3, -G3). Because the DFT converge rapidly with the basis functions the key approximation is the used DFT model.

Interesting systems

AIMD can be applied to any system but the solid systems are not very interesting since very little happen on the AIMD time scale. One can study light atom, like Li, diffusion but even that is better to do with some constraint. One can push the atom to some direction and monitory the constraint force.

Small molecular cluster are rather interesting. Especially water clusters but not much AIMD simulations were done. The most common systems are the liquids. In liquid systems the periodic boundaries are used. For the electronic structure the Γ-point k=(0,0,0), is almost always used but other k-point sampling is possible. The system should contain at least 100 molecules and preferably 200+ molecules. The minimum time scale for pure liquids relaxations is ca. 10 ps water to 25 ps for ethanol. But if there are some ions, like Na+, Cl-, in the system the equilibration time is much longer. Like 50-100 ps.
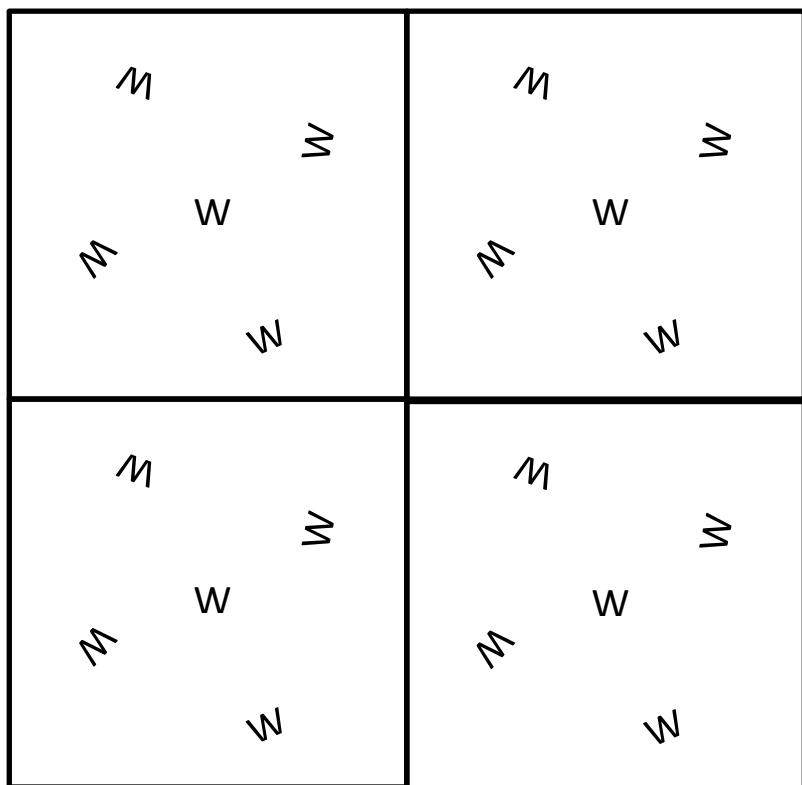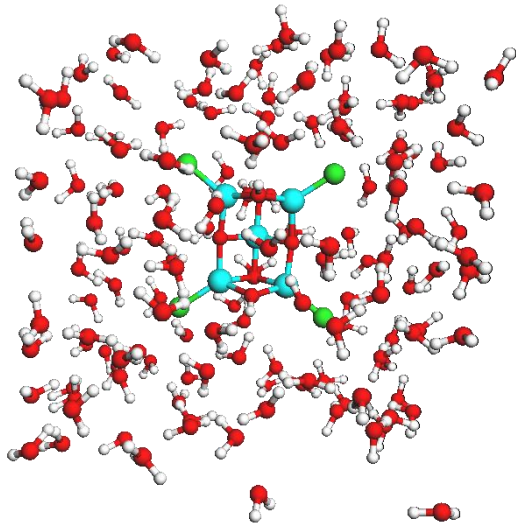
A very small system: Al2-Cl2-OH molecule + 65 water molecules.

This system is so small that the molecule will fit to the box but not the solvation layers of the Cl atoms. It is not so easy to see this from a static picture.
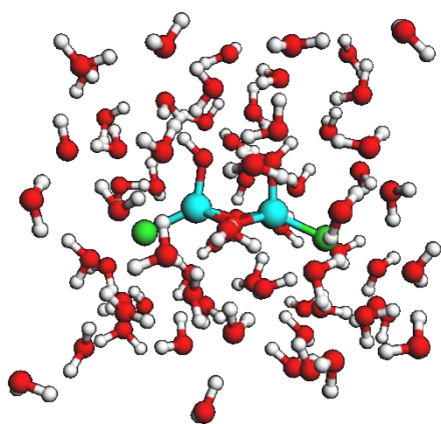
A larger system: Al4-Cl4-OH-O molecule and 144 water molecules.

What can be learned from AIMD?
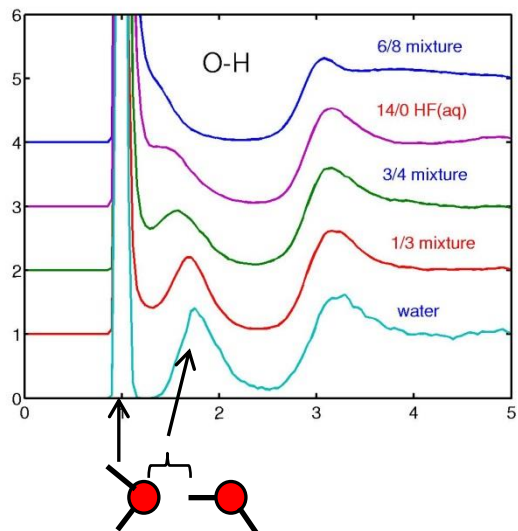
Almost anything.

The most important quantity is the pair correlation function. It describes the atoms radial probability distribution. It tells how likely that atom B is at distance d from atom A.

Below there is OO and FO pair correlation functions of water, concentrated HF (44 at%) and HF/HCl (HF 19 at%, HCl 25 at%). Here we see that that the O-O distance will reduce a lot.
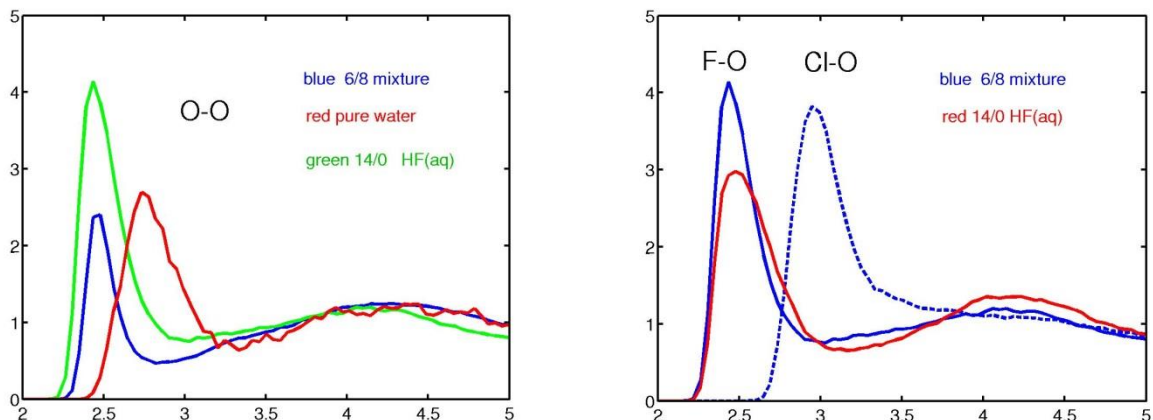

The OH pair correlation is even more interesting. The most interesting is the O..H-O peak. In pure water this peak is very clear but when the acid concentration (both HCl and HF) will increase this peak start moving towards the molecular O-H peak and highest concentration it merges to the H-O peak. This means that

the water molecules start losing their structure!! The hydrogen
bonded hydrogen is almost as close to the O than the normal H.


Reaction in liquid

With AIMD one can study almost anything but one of the most
interesting topic is the chemical reactions. As said earlier the
direct simulations is usually far too slow and because all the
atoms are moving the NEB type methods are not useful since they
assume that everything else except than the path is in minimum.
The simplest method it the constraint method where some constraint
is imposed. The simplest constraint is the bond distance; $g=d(AB)-d_0$. ($d(AB)$ is the distance between the atoms and d_0 is the value of
the constrain. Now one can do AIMD with this constraint and one
need to monitor the force acting to the constraint. In this case
it is easy $f_{cons}=(F_A-F_B).(R_A-R_B)$. Naturally the atoms will collide
each other and this force is very noisy. One need to collect data
for relatively long time (20-30 ps) to get a proper force; now we
have $<f_{cons}(d0)>$.
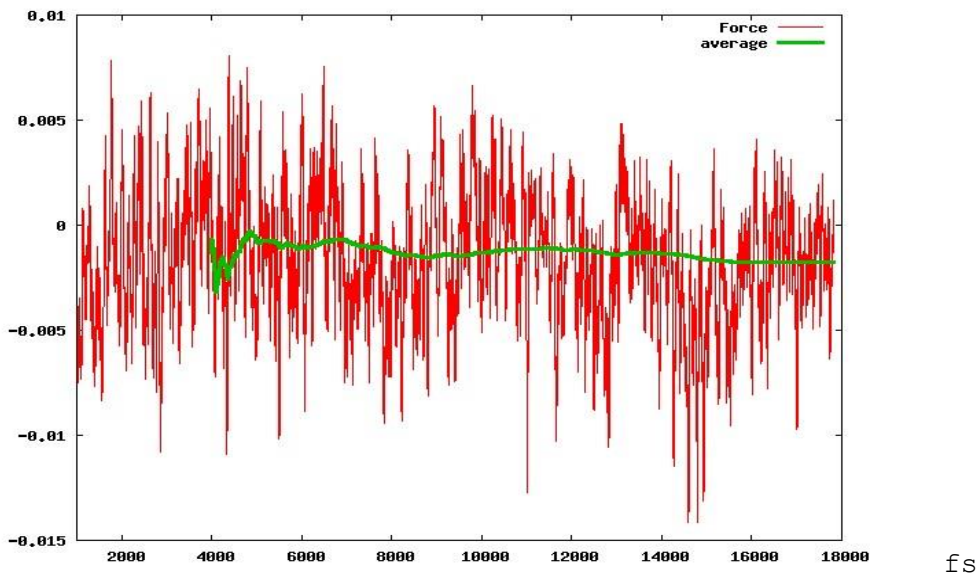
Once we are happy with the $<f_{cons}(d0)>$ we can change the d0 value. The we collect statistic with this constraint etc. The free energy difference can be computed as an integral over the forces at each distance:

$$\Delta G = \int_{d0}^{d1} \langle f_{cons}(x) \rangle \, dx$$

The constraint method is very general and the constraint can be any functions that depend on the atomic positions, like angle, torsional angle, coordination number, etc. The main problem of the method is that the real reaction coordinate is not a function of the constraint.


40 ± 5 kJ mol^{-1}

14 ± 3 kJ mol^{-1}
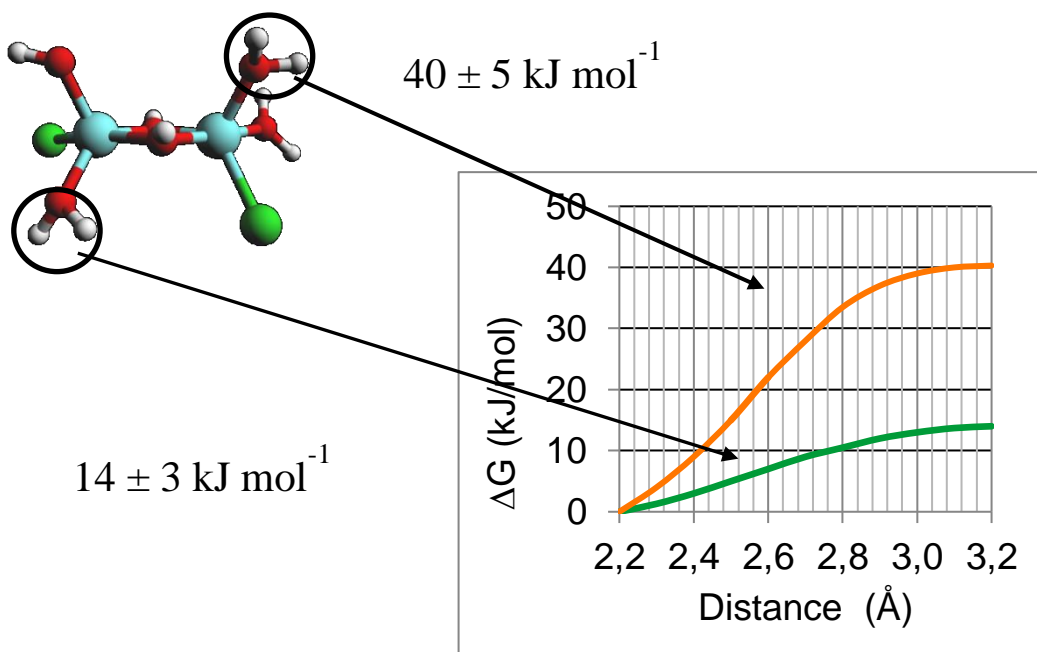
Metadynamics

There are several methods for computing the Free energy barriers but we cannot go through all of them. I present one biased dynamics method, the Metadynamics. The Metadynamics has been developed by Laio and Parrinello (Laio ja Parrinello, PNAS, 2002). A good review article of the topic is Laio ja Gervasio, Rep. Prog. Phys. 71 126601 (2008))

The idea is to choose few collective variables (CV) to parametrize the reaction. The CV usually depend on atoms. Possible CV's can atom distances, etc. The CV's have always some values and during the dynamics this value fluctuates. After some MD steps one adds an small repulsive Gaussian function to the CV's average value <CV>. We will build a bias potential Vbias(CV2,CV2,CV3,..) during the simulation and this potential will push the system away from the starting minima. The smart thing is that we do not force the system to go any particular direction. We only "fill" the current minima. After the system has moved to a new minima we start filling that and the system will move to a new minima. After several steps we have a good map of the free energy. Providing that the CV have been reasonable. With one CV the Metadynamics is not very effective and same results can be found by scanning the CV with the constraint method. The key strength is that two (or more) CV's can be used. Naturally the Metadynamics become very time consuming with more CV's and typically two CV's are used.
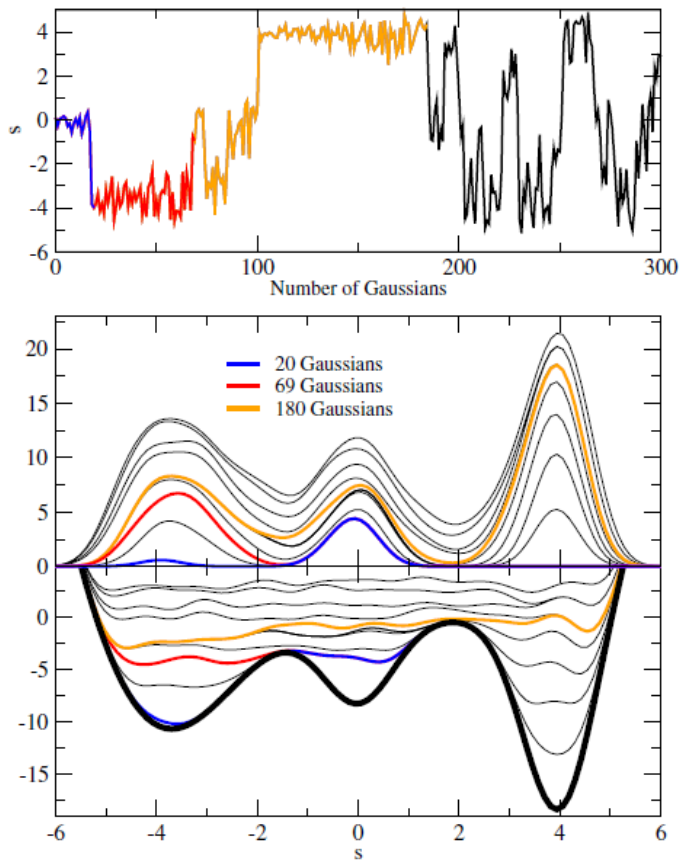
Problem: why Metadynamics is not effective in 1-dim.

In more complex reaction the Metadynamics with 2 CV's is a good choice. But choice of the CV's is non-trivial. They need to represent well the real reaction coordinate which we do not know.
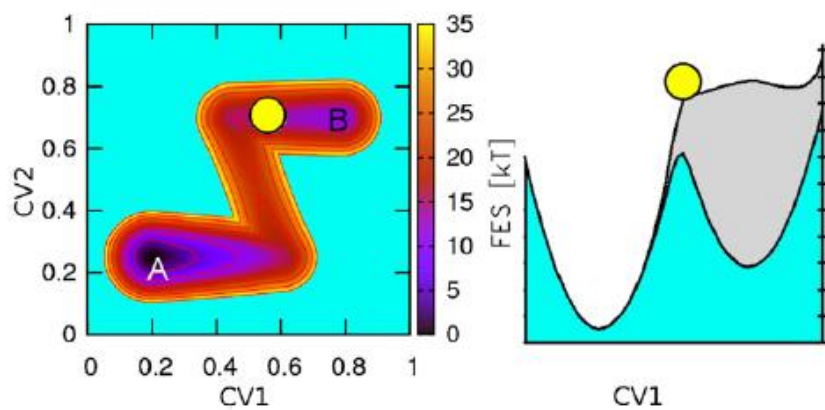
Figure 2. The effect of neglecting a relevant degree of freedom. Left side: 2D Z-shaped potential. Right side: the trajectory of a metadynamics simulation generated using only s1 as CV. Transitions from A to B are properly described by CV1, causing strong hysteresis in the reconstructed free energy.

An example Al3OnHm cluster

We have studied a lot of small $Al_nO_mH_k$ clusters. Even a given cluster in gas phase have several structures and very little is known in their structures in water. The Mass Spectra data suggest also that there are huge amount of different clusters but the MS do not give information of the structures.

$[Al_5O_8H_6Cl_4]^+$ m/z = 409   **$[Al_5O_9H_8Cl_4]^+$ m/z = 427**

ΔE [kJ/mol]

$\Delta E_{PBE/TZVP} = +161$
$\Delta E_{PBE-D/TZVPP} = +172$
$\Delta E_{COSMO} = 0$
(e2)

$\Delta E_{PBE/TZVP} = +220$
$\Delta E_{PBE-D/TZVPP} +232$
$\Delta E_{COSMO} = +150$
(d3)

$\Delta E_{PBE/TZVP} = +230$
$\Delta E_{PBE-D/TZVPP} = +244$
$\Delta E_{COSMO} = +13$
(d2)

$\Delta E_{PBE/TZVP} = +17$
$\Delta E_{PBE-D/TZVPP} +17$
$\Delta E_{COSMO} = 0$
(c3)

$\Delta E_{PBE/TZVP} = +3$
$\Delta E_{PBE-D/TZVPP} = +2$
$\Delta E_{COSMO} = +95$
(c2)

$\Delta E_{PBE/TZVP} = +3$
$\Delta E_{PBE-D/TZVPP} +2$
$\Delta E_{COSMO} = +68$
(b3)

$\Delta E_{PBE/TZVP} = +4$
$\Delta E_{PBE-D/TZVPP} = 0$
$\Delta E_{COSMO} = +287$
(b2)

$\Delta E_{PBE/TZVP} = 0$
$\Delta E_{PBE-D/TZVPP} = +0.4$
$\Delta E_{COSMO} = +245$
(a2)

$\Delta E_{PBE/TZVP} = 0$
$\Delta E_{PBE-D/TZVPP} 0$
$\Delta E_{COSMO} = +66$
(a3)

The cluster are prepared from $AlCl_3$ that is added to water. From this solution on can observe huge amount of cluster (several hundreds) and they are of various sizes. Thus the $AlCl_3$ clusters will react rapidly in water. We wanted to understand what happen to the $Al_3O_nH_m$ cluster in water and we used Metadynamics (MetaD) to it. We had a good collection of the gas phase cluster structures but the cluster interact very strongly with water. So we start with the best gas phase cluster and placed that to water. We

choose two CV's (d(Al1-Al2)-d(Al2-Al3) and RMS deviation from the starting structure) and start a MetaD simulation.

The Free energy surface from the MetaD is below. It is not very reliable but it is indicative and it show that this cluster have very many structures that are not very different in Free energy. It also show that the MetaD is a good tool to map complex surfaces and we do not need to decide how the system will behave. There are also some problems in the structure A where the cluster has broken, and I do not think that the CV's can handle this.
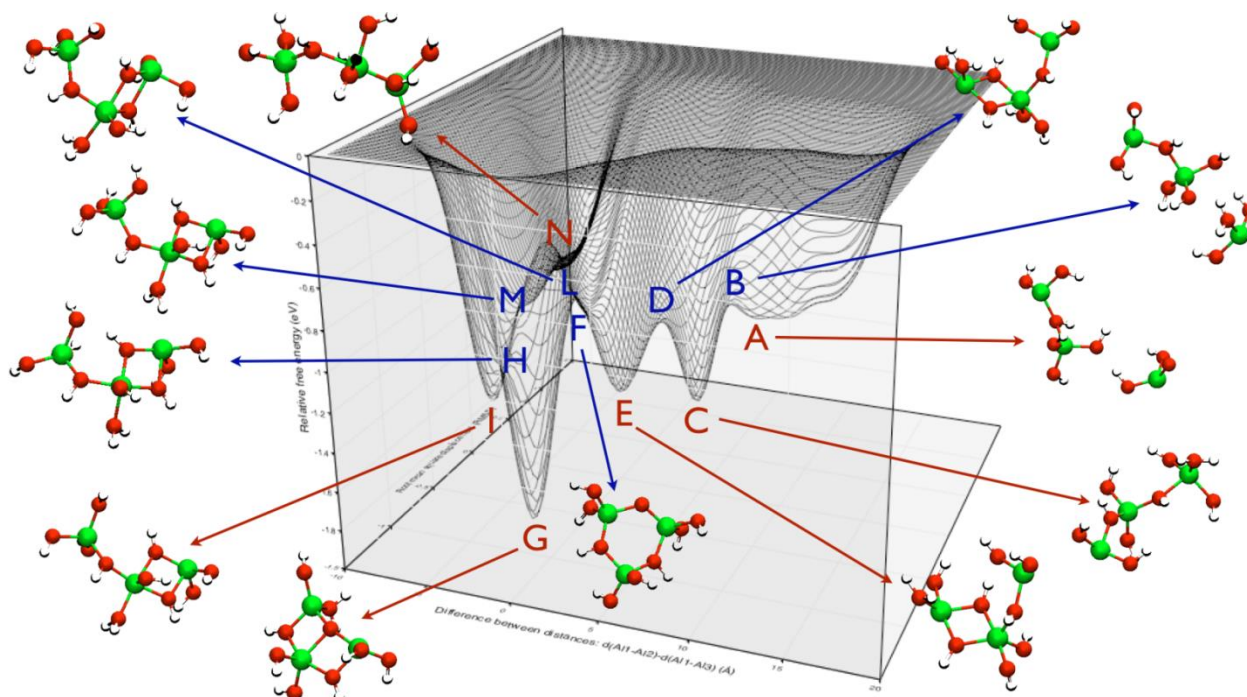

Figure: Free energy surface of Al2OnHm in water, Lanzani et al. JPCC (2017)