KJ2900 - Bachelorprosjekt i kjemi

Can redox reactions be studied with RETIS?

Abstract: Simulating chemical reactions can be time consuming to the point where some reactions can be impossible to simulate. The aim of this thesis is to compare different simulation methods, such as ab initio molecular dynamics (ab initio MD) to classical MD, and brute force ab initio MD to replica exchange transition interface sampling with ab initio MD. In comparing ab initio MD to classical MD, it was not possible to use classical MD for chemical- or redox reactions. This is because of the difficulties in getting accurate results, even when using force fields such as ReaxFF, because of the difficulties in parameterizing the force fields. RETIS with ab initio MD is also compared to brute force ab initio MD for dissossiation of water, oxidation of iron, and simulating aqueous silicate condensation, in which it proved to be faster by a factor of roughly 100 000 times for the condensation. RETIS also yielded more accurate result using decorrelated moves which reduces statistical error.

Author: B. Welfler
Supervisor: Prof. T. S. van Erp
April 30, 2020

Department of Chemistry
Faculty of Natural Sciences
Norwegian University of Science and Technology

Contents

Introduction

Computational simulation is becoming increasingly more relevant for studying the different properties of chemical reactions. This is because computational power is increasing, meaning it is possible to simulate over longer time scales and gain more insight about different properties of atoms and molecules, and how they interact.

Still, we have not reached the point where it is possible to simulate the full time scale of every chemical reaction. With brute force molecular dynamics (MD) you might have to spend centuries simulating just to get information about the crossing of the transition between two states in a chemical reaction. This is because most of the time the system want to be in the reactant state or the product state. This means it is very difficult to gain information about what goes on during a transition between two states. Luckily there are techniques that can help us out! Events that happen rarely are appropriately named *rare events*[1]. There are several available algorithms that can help us gain information about these rare events such as crossing the transition states in a reaction. In addition to molecular simulations, rare event simulation is also applicable in other fields.

Rare event simulation can be used for everything that is defined as a rare occurrence. It can be used to compute failure probability of a building under earthquake excitation[1], estimation of probabilities for future possible terrorist events[2], studying packet loss in real-time video traffic in a packet-switched telecommunication network. In the insurance field where most of the cases are rare events where you have to take into account the wealth of the company, income, and outcome due to insurance claims[3] and especially relevant now, the spreading of diseases[4]. As shown, there are myraids of applications for rare event simulation. Albeit these different fields each require different algorithms than the ones that will be discussed in this paper, the idea of rare event simulation still applies to a wide range of fields.

In this paper, the basics of computational simulations such as MD and Monte Carlo simulations will be presented along with different rare event sampling methods, such as transition path sampling (TPS), transition interface sampling (TIS) and replica exchange transition interface sampling (RETIS). Some basic theory on electron transfer reactions (redox reactions) and reaction rate theory will be introduced. Finally the different methods presented in the thesis will be compared in terms of usefulness for different types of reactions.

1   Molecular simulations

For advanced systems, it can be hard to predict different states and outcomes. For these systems, applying mechanics and simplified methods, and then just let the simulations run makes it possible to observe the changes in a system. There are two common simulation techniques: Molecular dynamics (MD) and Monte Carlo (MC). This chapter will explain how these methods work, where, and when they are applied.

1.1   Molecular dynamics

MD is a method that generates the conformation of a system by integrating Newton's laws of motion. By solving the differential equation it is possible to obtain the different trajectories.

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{F_i}{m_i} \tag{1.1}$$

Where i is the index for a given particle, x_i is position for particle i, t is time, m_i is the mass of the particle i, and F_i is the force affecting particle i.[5]. As seen in equation 1.1, this is a time dependent function. After applying Newton's laws of motion on the system, MD then simulates what would happen at a given *time step*. A time step is the time between each calculated state of the system. The importance of having a suitable time step is that a too small time step will make the simulation time unnecessarily long, and too large of a time step can cause instabilities in the integration algorithm used, which will give inaccurate information about the system in return[5].

In general, an MD step is between 1 and 10 femtoseconds (10^{-15} second), and a typical MD system consists of 100 to 100 000 molecules with a time scale of nanoseconds to microseconds[6]. The size of an MD system depends on the computational power and the type system. The D.E. Shaw group have the record for classical MD calculations in which a Joint AMBER-CHARMM benchmark system containing 23 558 atoms (dihydrofolate reductase surrounded by water) was simulated 16.4 μs (10^{-6} second) per day[7]. A system using a version of the NAMD software has managed to produce 100 ns (10^{-9} second) per day for a 30 000 atom system[8].

1.1.1   Ab initio MD

Ab initio means *from the beginning*[9] in Latin and is a special case of MD. Classical MD is calculating the intramolecular forces using premade potentials such as the Lennard-Jones potential[5] to speed up the calculations. Meanwhile ab initio

MD calculates all the intramolecular forces from scratch at every MD step. This method is more computationally demanding, but can be necessary for some calculations, especially to study a process in which chemical bonds are formed or broken.

Ab initio MD offers some advantages over classical MD. Because classical MD uses empirical potentials to describe the forces on pairs of atoms or molecules there are some cases where classical MD has difficulties describing the physical behavior of the system. For example, in liquids where the force between two atoms or molecules can be affected by the presence of one or more other atoms or molecules, meaning the complete electronic state can change. There are force fields that takes these cases into account (many body potentials), but making these force fields is difficult[9].

Because ab inito MD requires more processing power, the systems and time scales is smaller. A system with 100-300 atoms is possible to simulate with a time scale of 10-100 ps (10^{-12} second)[10].

1.2 Monte Carlo

In contrast to MD, MC is a technique which is not time reliant. In MD, the conformation and velocities of the system is calculated at discrete time intervals, while in MC you do not get the time evolution of the system, but you get the statistical distribution of the different conformations.

The method is based on attempting to make a random change in the system, then either accept or discard the change. When a change is made, the energy of the new conformation is calculated. If the energy is lower than the old conformation, the change is accepted. If the energy is higher, a random number between 0 and 1 is generated and compared with *Boltzmann factor* given below

$$e^{\frac{-\nu(r_{new}^N)-\nu(r_{old}^N)}{k_B T}} \tag{1.2}$$

where $\nu(r^N)$ is the potential energy of the conformation as a function of the reaction coordinate r, the subscript *new* and *old* refers to the new and old conformations before the attempted move, k_B is the Boltzmann constant and T is the temperature. If the random number is higher than the one generated in eq. 1.2 , the change is rejected, if it is lower, the change is accepted[5].

2   Rare event simulation techniques

Rare events is a term used for events occuring with a low frequency[1]. This does not imply that the event seldom occurs, but compared to the time scale of an MD simulation it will count as a *rare event*. A way to visualize this is a chemical reaction where you have molecule at a low energy state, S_1, that has to pass a higher energy barrier to end up at a lower energy state, S_2. The molecule does not want to be in the transition state, but rather in the stable states (S_1 or S_2), meaning it is difficult to sample the crossing of the transition state. Rare event simulation can help describe these situation, such the transition state between S_1 and S_2 in a computationally efficient way.

This thesis will focus on two different approches of simulating rare events. One is based on *free energy calculations* and the other is based on *path sampling*

2.1   Free energy based methods

Free energy based methods sample the free energy of the system as a function of a *reaction coordinate*. A reaction coordinate is ”*an abstract one-dimensional coordinate which represents progress along a reaction pathway. It is usually a geometric parameter that changes during the conversion of one or more molecular entities*”(Tuckerman, 2008)[11]. Free energy methods sample the conformation space along the full scale of the reaction coordinate including the reaction barrier. The energy profile then provides the reaction's free energy and the free energy barrier. There are two main methods for obtaining the free energy profile, *thermodynamic integration* and *umbrella sampling*[12].

2.1.1   Thermodynamic integration

Thermodynamic integration is a free energy based method where the free energy is calculated by assuming that the Hamiltonian is defined as a function of a continuously varying parameter giving

$$\Delta F(\lambda^*) = F(\lambda^*) - F(\lambda_A) = \int_{\lambda_A}^{\lambda^*} d\lambda' \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda'} \tag{2.1}$$

where F is the Helmholtz free energy, λ_A and λ^* represents respectivly start state and local free energy maximum (free energy barrier). The brackets refers to an ensemble average, H is the Hamiltonian for the complete system, λ is the reaction coordinate (varying parameter) and the subscript indicates fixed values for reaction coordinate λ'[9].

2.1.2   Umbrella sampling

Another way to sample the free energy barrier is a method called umbrella sampling. This method is meant to overcome the issue of insufficient sampling of energetically unfavorable states. It is possible to perform umbrella sampling using either MC or MD. Umbrella sampling is based on limiting the possible moves a particle in the system can make by hindering *ergodicity* from the system[13]. Ergodicity is a concept that impiles it should be possible to visit all the conformation points in the system.

By removing ergodicity the system is forced into unfavorable positions, so it is possible to sample these conformations. It is then possible to limit the conformations to a certain area, so we can sample the energy maxima in a transition state. By sampling the probability distribution of the conformations within the umbrella sampling window, we can obtain the local probability distribution. By repeating this procedure for different regions along the reaction coordinate, using different umbrella sampling windows, we get a set of probability distributions that can be matched to get the overall probability distribution[14].

2.2   Path sampling methods

In addition to free energy based methods, there are path sampling methods. These methods use MC sampling on MD generated paths. This paper will not go into great detail about how these methods work, but will explain the general. Because it is time consuming to generate many trajectories, therefore it is important to have an efficient algorithm to generate these trajectories.

2.2.1   Transition path sampling

Transition path sampling (TPS) is a path sampling method, in which the rare occurence where a system moves between stables states is heavily sampled. TPS is nowadays considered just as a method for sampling trajectories from a state A to state B. The original article also included an approach to calculate the probability of passing from state A to state B (which is essencial to compute the reaction rate[15]. This approach was later improved by other methods such as transition interface sampling (TIS) and replica exchange transition interface sampling (RETIS)[12].

There are two types of MC moves in TPS, *shooting move* and *shifting move*[15]. In the shooting, move a point in a trajectory is selected at random. As shown in fig. 2.1, a point with particles having the momenta P on trajectory (o) is selected, and a random change (ΔP) is applied to the momenta of all the particles at the specific
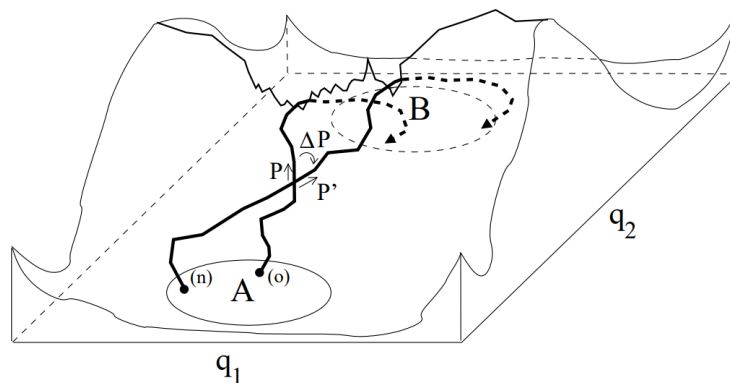
Figure 2.1: Illustration of a shooting move. A trajectory is generated between state A and state B, where a random change (ΔP) is made to the momenta P of all particles in a given point, to generate the new trajectory with momenta P'. Afterwards we integrate forwards and backwards in time to generate the trajectory from state A to state B. This figure taken from ref. [9].

time slice. This creates the new momenta P' from the old point. Afterwards we integrate forwards and backwards in time to generate a new trajectory (n) to complete the shooting move. In TPS the shifting move is also employed, but we will not discuss this, as this move is redundant in the more advanced path sampling simulations techniques such as TIS and RETIS[16]. There also exist other moves such as *web throwing* and *stone skipping*, which can increase the efficiency of factor 10 in some cases[17], but the explanation of these moves involves complex mathematical exposure which is beyond the scope of this thesis.

2.2.2 Transition interface sampling

TIS is a development of the TPS method. Here a set of *interfaces* (λ_0, λ_1,...,λ_n) are defined, between the two states, A and B. Using these interfaces, we can use the same algorithms as in TPS (e.g. shooting move) to attempt to generate trajectories[17]. For example, an ensemble of 1 000 trajectories can be generated, where the trajectories are required to pass from state A through at least λ_0, where we reject every move that does not pass the interface. We refer to the collection of paths that we generate in this way as $[0^+]$ ensemble, which imples all the trajectories that have been sampled cross the λ_0 interface. As a general expression, we can refer to the $[i^+]$ ensemble as the collection of paths which we can generate, that crosses the λ_i interface[12].

To generate trajectories in the $[i^+]$ ensemble, we reject any path that fail to cross

λ_i, so by definition the paths that sampled in this ensemble will all cross λ_i. From the generated trajectories in the $[0^+]$ ensemble, we can calculate the probabilities of passing through the next interface, λ_i, given that it passed λ_0. The trajectories that were sampled in the $[0^+]$ ensemble all cross λ_0, but some might go further along the reaction coordinate. The fraction that cross λ_1 will be the conditional crossing probability (crossing λ_1 given that is crosses λ_0). Once we have simulated 1 000 trajectories in the $[0^+]$ ensemble we will move on and try to simulate 1 000 trajectories in the $[1^+]$ ensemble. To initiate this simulation in the $[1^+]$ ensemble, we need to have an initial trajectory that crosses λ_1. We can pick this initial trajectory by picking a trajectory from the $[0^+]$ ensemble that also crossed λ_1. Once we have a trajectory we use the MC moves in the same way as in TPS and generate trajectories. Any trajectory not passing through λ_1 will be rejected. All trajectories generated in the simulation for $[1^+]$ will then have crossed the interface λ_1. This simulation of $[1^+]$ allows us to generate the probability of going from state A to λ_2 given it crosses λ_1.

By getting these conditional probabilities of crossing a given interface from another interface, it is possible to compute the overall probability of going from state A to state B. The TIS method was shown to be more efficient than TPS by a factor of 5 for a simple two-state dimer molecule immersed in a solvent by allowing flexible path langths and the introduction of interface path ensemble averages.[18]. The relative efficiency increase is expected to be even higher for complex systems.

2.2.3   Replica exchange transition interface sampling

Replica exchange transition interface sampling (RETIS) is a combination of TIS and the replica exchange method[19]. This method gave a factor 20 improved in efficiency compared to TIS in studying denaturation of a mesoscopic DNA model[20]. The method is more efficient than TIS because it still generates the trajectories with MD and the ensemble using MC (mainly shooting move), but it also uses special MC move called *swapping move*, also known as *replica exchange*[12]. This move attempts to swap two generated paths into different path ensembles as shown in fig 2.2 along with the shooting move and time reversal/shifting move from TPS.

Consider the interfaces λ_1 and λ_2, and the assosiated path ensembles $[1^+]$ and $[2^+]$. If we were to simulate both path ensembles at the same time, and updated each path ensemble by a full move, it is then possible to attempt to swap these moves betweeen these two ensembles. If the path from the $[1^+]$ ensemble crosses λ_2 it is a valid path for the $[2^+]$ ensemble and the move is accepted. If it does not pass λ_2, it is rejected. The path in the $[2^+]$ ensemble is required to cross λ_1, so this swap is automatically accepted. Swapping these two paths gives us two *new* paths with very little computational cost. In addition, this method helps making the moves
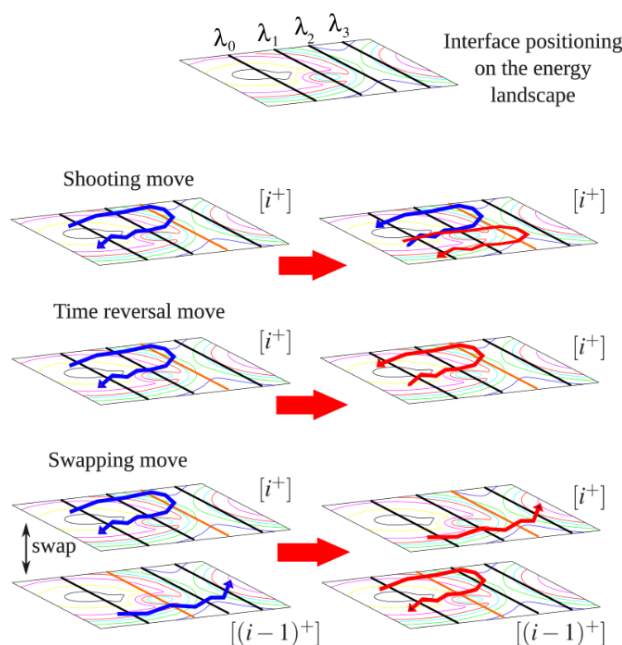
Figure 2.2: Illustration of the moves for generating trajectories with RETIS. First the MC shooting move, then the time reversal/shifting move and lastly the swapping move. This figure taken from ref. [12]

uncorrelated. TIS generates new paths by making small changes to a previous accepted path, therefore these paths will in general look similar. With RETIS it is possible to swap two uncorrelated paths. Because the set of the trajectories in each path ensemble are more decorrelated this gives a lower statistical error after generating a fixed number of trajectories compared to the TIS method.

The downside to RETIS is that since you are doing the swapping moves, it is hard to parallellize the simulations[12]. The RETIS algorithm is also more difficult to implement than standard TIS, therefore it is not as frequently used as it should be considering the performance upgrade from regular TIS. Recently the two groups in NTNU and University of Amsterdam has developed codes, PyRETIS[21] and open path sampling (OPS)[22] that should make this easier.
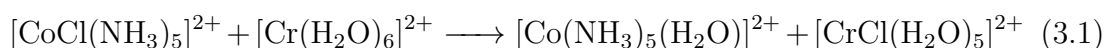
2.3 Other path sampling methods

There are several other path sampling methods that are worth mentioning. These include forward flux sampling (FFS)[23], adaptive multiple splitting (AMS)[24], milestoning[25] and partial path transition interface sampling (PPTIS)[16]. These are all methods for sampling the reaction paths, however some of these methods

use approximations to increase the efficiency, but in turn makes them no longer an exact method (milestoning and PPTIS). Other methods such as FFS and AMS only work for special types of dynamics (stochastic dynamics). For this reason the paper will not focus on these methods.
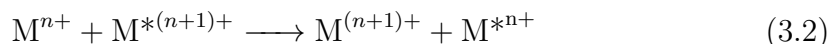
3 Reactions theory

3.1 Redox reactions

Redox reactions are *electron transfer reactions*. These are reactions in which electrons moves from an electron *donor* (oxidized part) to an electron *acceptor* (reduced part). These types of reactions can be divided into two categories called **inner sphere electron transfer** and **outer sphere electron transfer**[26]. The inner sphere reactions occur between a donor and an acceptor that is covalently bonded, for example

$$[\text{CoCl(NH}_3)_5]^{2+} + [\text{Cr(H}_2\text{O})_6]^{2+} \longrightarrow [\text{Co(NH}_3)_5(\text{H}_2\text{O})]^{2+} + [\text{CrCl(H}_2\text{O})_5]^{2+} \quad (3.1)$$

where Co goes from Co(III) to Co(II) and Cr goes from Cr(II) to Cr(III), meaning Co gets reduced and Cr gets oxidized. In addition H_2O goes from being bonded to Cr to being bonded to Co and Cl goes from being bonded to Co to being bonded to Cr.

The outer sphere electron transfer does NOT involve the rearrangement of bonds. This transfer can be observed in for example two metal ions in aqueous solution.

$$\text{M}^{n+} + \text{M}^{*(n+1)+} \longrightarrow \text{M}^{(n+1)+} + \text{M}^{*n+} \quad (3.2)$$

Here we can see that M donates one electron to M*, and the oxidation states of both M and M* are changed[26].

3.2 Chemical kinetics

The reaction rate is *"the change in concentration of a reactant or a product with time"*[27]. For a simple equation

$$\text{A} \longrightarrow \text{B} \quad (3.3)$$

the reaction rate is given by

$$rate = -\frac{\Delta[A]}{\Delta t} = \frac{\Delta[B]}{\Delta t} \quad (3.4)$$

where Δt is the change in time and $\Delta[A]$ and $\Delta[B]$ are the changes in the concentration over the given period of time. If we have a more general reaction given as

$$aA + bB \longrightarrow cC + dD \tag{3.5}$$

where a, b, c and d are coefficients and A, B, C and D are atoms or molecules in a reactions. The reaction rate is given by

$$rate = -\frac{1}{a}\frac{\Delta[A]}{\Delta t} = -\frac{1}{b}\frac{\Delta[B]}{\Delta t} = \frac{1}{c}\frac{\Delta[C]}{\Delta t} = \frac{1}{d}\frac{\Delta[D]}{\Delta t} \tag{3.6}$$

A rate constant k is introduced to explain the proportionality between the reaction rate and the concentration of the reactants. This gives us

$$rate = k[A]^x[B]^y \tag{3.7}$$

where [A] and [B] are the concentration of the reactants and x and y is the reaction order[27]. Eq. 3.7 reflects how many transition events are expected per second, per mol reactants. The rate can be translated into the rate of a specific system, with a specific number of molecules. In that case, the rate of the system will have the unit of s^{-1}. The inverse of the rate of the system will then be the expected time for a reactive event.

4 Comparing methods

Choosing correct a method is important as it can differ in months or years in simulation time. The method can also affect the accuracy of the results. In this section the applicability of the different of the different methods will be analyzed for redox reactions.

4.1 Classical MD vs. ab initio MD

As explained in chapter 1.1 and 1.1.1 we have two types of MD, classical MD and ab initio MD. So what dictates the choice of method? Classical MD uses a premade potential. This makes it very hard to simulate bond breaking and bond formation because it is difficult to parametrize these cases. A possible way to use classical MD for these reactive events, would be to use ReaxFF force field, which is designed for describing reactive systems. This method is faster than quantum-chemical and semi-empirical quantum methods currently used[28]. However even using a specialized force field such as ReaxFF it is still difficult to parameterize and the optimal parameterized potential can often give unreliable results[29]. Redox reactions are even harder to simulate than normal breaking and forming of bonds.

This is because the transfer of electrons happends very quickly, so there is hardly any change in the coordinates of the nuclei.

Because of the aforementioned reasons, it is not possible to use classical MD to accurately study redox reactions. It is therefore required to use ab initio MD. Because ab initio MD is more computationally expensive, it requires a significant reduction, meaning only small systems with short simulation time can be studied. Systems for studying redox reactions do not have to be as big as other systems, for example protein folding[19]. Hence, redox reactions needs to use ab initio MD over classical MD.

4.2   Studying redox reactions with RETIS

The electron transfer reactions behave similar to other chemical reactions. They both can be described by concepts such as free energy barriers, reaction coordinates etc. Since redox reactions are often rare events, it should be possible to use RETIS to study these reactions. Redox reactions have been studied with several computational methods[18, 26, 30]. One of them is very relevant for us, which is *Toward accurate simulation of electrocatalyzed water splitting* by Tiwari[26], where the reaction

$$Ru^{2+} + Ru^{3+} \longrightarrow Ru^{3+} + Ru^{2+} \tag{4.1}$$

is studied using ab initio MD with TPS. This approach does not compute any quantitative properties regarding reaction rates, because TPS is not able to compute the probability for the rare event. This could have been done with TIS or RETIS. The latter is preferred as it allows you to get more efficient calculations for the reaction rates than with TIS.

As stated earlier, it should be possible to study redox reactions using RETIS, but it is beneficial to make an analysis of the computational cost for doing a RETIS reaction with ab initio MD and a brute force MD simulation with ab initio MD. Since we already found out it is not possible to use classical MD for these types of reactions.

4.3   Brute force MD vs. RETIS at the ab initio level

A specific water molecule in liquid water takes 10 hours before it dissociates to hydronium (H_3O^+) and hydroxide (OH^-) ions[31]. Using brute force ab initio MD, with time steps of 1 femtosecond these 10 hours would require $3.6 \cdot 10^{19}$ steps. This is an astronomical amount of steps, however because over 99.9999% of the these steps give no information related to the dissosiation of water into H_3O^+ and OH^-, a smarter algorithm can reduce the computational cost without loss of information, which is why RETIS can be used.

To give an estimate of the CPU time needed to simulate $3.6 \cdot 10^{19}$ steps with present day computational resources. For a similarly sized system with 64 water molecules and 1 chloride ion, using a cluster of 4 nodes with 32 cores, it took about 45 minutes to simulate 1 ps (30. April 2020. Personal information. Moqadam, M.). 1 ps of simulation would then require 1 000 MD steps. This means if we assume that the system for water dissosiation is similar to the system above, it would take $1.62 \cdot 10^{18}$ minutes to simulate the occurence of the water dissosiation, which is roughly equal to 3 trillion years. The dissociation of water molecules is a rare case among rare cases, therefore this is not the most representative result, therefore we will give another example where we compare brute force MD and RETIS at the ab initio level.

Another article compared path sampling using RETIS with ab initio MD to investigate mechanicsms and rate of silicate dimerization reactions in aqueous phase[18]. It was found that the rare event occured once per 50 000 ps of simulation time, this means a brute force simulation would have to simulate 50 000 ps of simulation time to observe the event once. This would require 2 250 000 minutes of CPU time, which is slightly more than 4 years. Using RETIS it was possible to generate 2 000 reactive trajectories, which would have required to simulate 100 µs using brute force ab initio MD.

Using the same system as previously it would require $4.5 \cdot 10^9$ minutes, which equates to 8562 years of simulation, just to get the same amount of trajectories as with RETIS. Comparing brute force ab initio MD using a timestep of 1 fs to RETIS, it would require $2.0 \cdot 10^{11}$ force evaluation with brute force ab initio MD. According to the article RETIS requires a number of force evaluations equating to between 1 000 and 1 400 ps of MD simulation, which equates to between 1 000 000 and 1 400 000 force evaluations. Comparing the efficiencies in force evaluations RETIS produces roughly a 100 000 times speed increase, meaning it would take RETIS close to 31 days instead of 8562 years.

Now let us look at how RETIS and brute force MD compares for redox reactions. Given the redox reaction

$$\text{Fe}^{2+} + \text{OH}^\bullet \longrightarrow \text{Fe}^{3+} + \text{OH}^- \tag{4.2}$$

where the reaction rate constant, k, is $3 \cdot 10^8 \ \frac{L}{mol \cdot s}$[32]. In a system with one Fe(II) ion and one $\text{OH}^\bullet$ radical in a water solution consisting of 64 molecules it is possible to calculate the reaction rate using eq. 3.7 to get

$$rate = (3 \cdot 10^8 \frac{L}{mol \cdot s})[Fe^{2+}][OH^\bullet] \tag{4.3}$$

To get the rate for this system it is necessary to take into account the size of the

system, so for the proposed system the rate becomes

$$rate_{Sys} = (3 \cdot 10^8 \frac{L}{mol \cdot s})[Fe^{2+}][OH^\bullet] \cdot V_{Sys} \cdot N_A \tag{4.4}$$

where V_{Sys} is the volume of the system and N_A is avogadro's number. Using the density of water[33] and disregarding repulsion forces and the volume occupied by OH and Fe, we get

$$rate_{Sys} = (3 \cdot 10^8 \frac{L}{mol \cdot s}) \cdot (0.87 mol L^{-1})^2 \cdot 1.92 \cdot 10^{-24} L \cdot 6.022 \cdot 10^{23} mol^{-1} = 2.60 \cdot 10^8 s^{-1} \tag{4.5}$$

The expected time for the event to happend is given by the inverse of the rate of the system, which then becomes $3.84 \cdot 10^{-9}$ seconds, which is about 3840 ps of simulation time. Using the same system as assumed earlier where it took 45 minutes to simulate 1 ps, it would take about 172 800 minutes to run these simulations, which equate to about 120 days. This is a long time, but simulation time of 4 months are not unusual in the field of computational modelling. However there is no reason to waste computational resources in a situation like this, if better algorithms are available and provide the same information.

It is difficult to know how much RETIS would have sped up these calculations because it differs depending on the type of system being sampled. Previous it was 100 000 times faster, but RETIS gives better results compared to brute force MD when the event sampled is more rare. In the case of 120 days brute force ab initio simulation a RETIS simulation of several weeks to generate enough trajectories to be able to calculate e.g. the rate constant (30. April 2020. Personal communication. van Erp, T.S).

Conclusion

There are many approaches to simulating different properties of a chemical reactions. Choosing the correct method for the right application is crucial as in worst case scenarios simulation can take trillions of years, however with a smart method this can be dramatically reduced as these standard methods would sample mostly the stable states, and not the transition events. For reactions such as breaking and forming of bonds, or redox reactions it is necessary to use ab initio MD over classical MD because even specialized force fields still struggle to produce accurate results compared to ab initio MD. For redox reactions it is even more important to use ab initio MD over classical MD because the coordinates of the nuclei only changes slightly during the electron transfer, which makes it nearly impossible to parameterize reactive classical force fields for this event.

For simulating the transition state between redox reactions it is both possible, and recommended to use RETIS because of the short time scale of brute force ab initio MD and because redox reactions spend most of the time in either reactant or product state.

We have seen for standard reactions in comparing brute force ab initio MD with RETIS on ab initio level. The brute force MD would require 8 562 years of CPU time in one of the examples, while RETIS at ab initio level would only require 31 days. For redox reactions we did a similar calculation which proved that brute force ab initio MD would require 120 days to observe a single event, it is likely that RETIS could have simulated enough trajectories to calculate properties such as rate constant in a couple of weeks.

References

[1] J. L. Beck and K. M. Zuev, *Rare-Event Simulation*, pp. 1075–1100. Cham: Springer International Publishing, 2017.

[2] A. Clauset and R. Woodard, "Estimating the historical and future probabilities of large terrorist events," *Ann. Appl. Stat.*, vol. 7, pp. 1838–1865, 12 2013.

[3] S. Juneja and P. Shahabuddin, "Chapter 11 rare-event simulation techniques: An introduction and recent advances," in *Simulation* (S. G. Henderson and B. L. Nelson, eds.), vol. 13 of *Handbooks in Operations Research and Management Science*, pp. 291 – 350, Elsevier, 2006.

[4] F. Jeltsch, M. S. Müller, V. Grimm, C. Wissel, and R. Brandl, "Pattern formation triggered by rare events: lessons from the spread of rabies," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 264, no. 1381, pp. 495–503, 1997.

[5] A. R. Leach, *Molecular Modelling Principles and Applications*. Pearson Education Limited, 2001.

[6] T. S. van Erp, "Dynamical rare event simulation techniques for equilibrium and non-equilibrium systems," 2011.

[7] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, and B. Towles, "Millisecond-scale molecular dynamics simulations on anton," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, SC '09, (New York, NY, USA), Association for Computing Machinery, 2009.

[8] J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," *Current Opinion in Structural Biology*, vol. 19, no. 2, pp. 120 – 127, 2009. Theory and simulation / Macromolecular assemblages.

[9] T. S. van Erp, "Solvent effects on chemistry with alcohols. an ab initio study," 2003.

[10] J. Marchi, *Biocompatible Glasses From Bone Regeneration to Cancer Treatment*. Springer, 2018.

[11] M. E. Tuckerman, "Reaction coordinates." $https$: $//www.nyu.edu/classes/tuckerman/stat.mechII/lectures/lecture_9/node5.html$. Accessed: 13/04-2020.

[12] R. Cabriolu, K. M. Skjelbred Refsnes, P. G. Bolhuis, and T. S. van Erp, "Foundations and latest advances in replica exchange transition interface sampling," *The Journal of Chemical Physics*, vol. 147, no. 15, p. 152722, 2017.

[13] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 1011–1021, 1992.

[14] W. You, Z. Tang, and C.-e. A. Chang, "Potential mean force from umbrella sampling simulations: What can we learn and what is missed?," *Journal of Chemical Theory and Computation*, vol. 15, no. 4, pp. 2433–2443, 2019. PMID: 30811931.

[15] C. Dellago, P. G. Bolhuis, and P. L. Geissler, *Transition Path Sampling*. Springer International Publishing, 2001.

[16] T. S. Van Erp, *Dynamical Rare Event Simulation Techniques for Equilibrium and Nonequilibrium Systems*, ch. 2, pp. 27–60. John Wiley & Sons, Ltd, 2012.

[17] E. Riccardi, O. Dahlen, and T. S. van Erp, "Fast decorrelating monte carlo moves for efficient path sampling," *The Journal of Physical Chemistry Letters*, vol. 8, no. 18, pp. 4456–4460, 2017. PMID: 28857565.

[18] M. Moqadam, E. Riccardi, T. T. Trinh, A. Lervik, and T. S. van Erp, "Rare event simulations reveal subtle key steps in aqueous silicate condensation," *Phys. Chem. Chem. Phys.*, vol. 19, pp. 13361–13371, 2017.

[19] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physics Letters*, vol. 314, no. 1, pp. 141 – 151, 1999.

[20] T. Dauxois, M. Peyrard, and A. R. Bishop, "Entropy-driven DNA denaturation," *Phys. Rev. E*, vol. 47, pp. R44–R47, Jan 1993.

[21] E. Riccardi, A. Lervik, S. Roet, O. Aarøen, and T. S. van Erp, "Pyretis 2: An improbability drive for rare events," *Journal of Computational Chemistry*, vol. 41, no. 4, pp. 370–377, 2020.

[22] D. W. H. Swenson, J.-H. Prinz, F. Noe, J. D. Chodera, and P. G. Bolhuis, "Openpathsampling: A python framework for path sampling simulations. 2.

building and customizing path ensembles and sample schemes," *Journal of Chemical Theory and Computation*, vol. 15, no. 2, pp. 837–856, 2019.

[23] C. Dellago and P. Bolhuis, *Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events*, vol. 221, pp. 167–233. 12 2008.

[24] I. Teo, C. G. Mayne, K. Schulten, and T. Lelièvre, "Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time," *Journal of Chemical Theory and Computation*, vol. 12, no. 6, pp. 2983–2989, 2016. PMID: 27159059.

[25] P. Májek and R. Elber, "Milestoning without a reaction coordinate," *Journal of Chemical Theory and Computation*, vol. 6, no. 6, pp. 1805–1817, 2010. PMID: 20596240.

[26] A. Tiwari, *Towards Accurate Simulation of Electrocatalyzed Water Splitting.* PhD dissertation, University of Amsterdam, 2019.

[27] R. Chang and K. A. Goldsby, *General Chemistry.* McGraw-Hill Education, 2014.

[28] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "Reaxff: A reactive force field for hydrocarbons," *The Journal of Physical Chemistry A*, vol. 105, no. 41, pp. 9396–9409, 2001.

[29] M. Moqadam, E. Riccardi, T. T. Trinh, P.-O. Åstrand, and T. S. van Erp, "A test on reactive force fields for the study of silica dimerization reactions," *The Journal of Chemical Physics*, vol. 143, no. 18, p. 184113, 2015.

[30] T. van Erp, "Reaction rate calculation by parallel path swapping," *Physical review letters*, vol. 98, p. 268301, 07 2007.

[31] M. Eigen, L. De Maeyer, and J. D. Bernal, "Self-dissociation and protonic charge transport in water and," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 247, no. 1251, pp. 505–533, 1958.

[32] S. Ermolin, I. Ivanova, D. Knyazev, S. Trofimova, and I. Piskarev, "Mechanism of water luminescence upon radiolysis under the effect of background radiation," *Russian Journal of Physical Chemistry A*, vol. 86, 06 2012.

[33] A. Blackman, *Aylward and Findlay's SI Chemical Data.* John Wiley and Sons, 2014.