

## UvA-DARE (Digital Academic Repository)

### Molecular simulations in electrochemistry

*Electron and proton transfer reactions mediated by flavins in different molecular environments*

Kılıç, M.

#### Publication date

2014

#### Document Version

Final published version

[Link to publication](#)

#### Citation for published version (APA):

Kılıç, M. (2014). *Molecular simulations in electrochemistry: Electron and proton transfer reactions mediated by flavins in different molecular environments*. [Thesis, fully internal, Universiteit van Amsterdam].

#### General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# 2

## Techniques

*In this chapter we give a short overview of the techniques which have been used for this thesis. We briefly review key concepts of statistical mechanics, and describe the computational techniques that we used to simulate flavin in gas, in aqueous and in protein environments.*

To understand solvent reorganization during electron transfer and calculate acidity constant, some background information on statistical thermodynamics and electronic structure calculations is essential. We will start with summary of Marcus Theory (Section 2.1), the calculation of  $pK_a$  (Section 2.2), elementary statistical thermodynamics (Section 2.3); then move on to introduce the basic electronic structure calculations and the combination of these two fields by the method of DFT–MD in Section 2.5 and QM/MM in Section 2.6.

## 2-1 Marcus theory

Rudolph Marcus published a mathematical model for redox reaction, based on classic theories of physical chemistry in 1956. In his theory, the size of the energy barrier could be calculated by simple quantities such as ionic radii and ionic charges. After extending his theory to electron transfer between different kinds of molecules, he formulated such mathematical expressions as ‘the quadratic equation’ and ‘the cross-equation’.

His theory was tested empirically and was challenged by experimental programs in whole field of chemistry. The Marcus Theory helps us to understand many phenomena, some of which include the capture of light energy in green plants, redox reactions in biological systems, oxidation/reduction processes, and photochemical electron transfer. Basically, Marcus Theory explains the rate of electron transfer reaction as a result of an electron moving from an electron donor to an electron acceptor in a polarizable medium [1–4, 6]. This theory has two assumptions: first, environment polarization has a linear response; and second, the amount of polarization is described by the vertical energy gap,  $\Delta E$ .

At this point, the vertical energy gap is the energy required to move an electron from donor to acceptor at fixed geometry without relaxing up the environment. Since the transfer of the electron –from the donor to acceptor– is much faster than solvent environment reorganization, the environment polarization slowly occurs after electron transfer process. Using the vertical excitation energy  $\Delta E$  as the reaction coordinate, Marcus theory models the electron transfer process with two parabolic curves that describe the donor/ acceptor states with the crossing point being the transition state. In this illustration, the reaction free energy, which is called redox potential, can be taken as the sum of the vertical energy gap and the reorganization free energy of the environment,  $\lambda$ . It should be kept in mind that since  $\Delta E$  is for fixed geometry, an entropic contribution is not included there.

There is a difficulty to the model of electron transfer process using a self consistent field electronic structure method such as density functional theory because imposing the electron transfer of an electron is far from trivial. The use of  $\Delta E$  as a reaction coordinate is not convenient. Recently, Sprik and coworkers addressed his problem

by focusing on a redox half reaction,



With this simplification, total DFT energies of donor (O, oxidant) and acceptor (R, reduce) can be computed individually in the ground state and the electron energy is taken as the chemical potential of an electron in a bath of electrons. In principle, the redox potential for reaction can be calculated using Reaction 2.1, but there are possible complications. First of all, there is the dependence on some of simulation parameters such as periodic boundary conditions, box-size dependency, self interaction energies, all of which cause an energy offset [22, 38]. Secondly, the free energy computation by molecular simulation doesn't take into account the electrode, which requires a reference potential for direct comparison with the experimental data. Therefore, the approach is limited to compute relative redox potentials between redox couple with the same box size and the same chemical environment.

Although there are this kind of restrictions, Sprik and coworkers [13, 14, 16–18, 20, 23, 39] demonstrated that the approach of half reactions is a very powerful tool to be employed in DFT-based molecular dynamics (DFT-MD). Marcus Theory can be applied for a half reaction 2.1. Besides, the free energy curves can be acquired from the distribution of the vertical gap energies. These distributions of vertical gap energies can be obtained from DFT-MD simulation of the system, which is either in an oxidized or reduced phase by computing the energy for certain geometries. With a sufficiently long simulation time, the histogram of  $\Delta E = E_R(\mathbf{r}^N) - E_O(\mathbf{r}^N)$  approaches the ensemble average of the probability density.

$$P_\eta(\Delta E') = Z^{-1} \int d\mathbf{r}^N e^{-\beta E_\eta} \delta(\Delta E - \Delta E') = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau \Delta E(\mathbf{r}^N(t)) dt, \quad (2.2)$$

in which  $P_\eta(\Delta E')$  is the probability to find  $\Delta E'$  in state  $\eta = O$  (oxidized) or  $\eta = R$  (reduced). Here,  $\beta = 1/k_B T$ , with  $k_B$  Boltzmann's constant and  $T$  the temperature.  $Z$  is the partition function and  $\mathbf{r}^N$  are the atom coordinates. The Landau free energy of the system as a function of the  $\Delta E$  coordinate is then

$$A_\eta(\Delta E) = k_B T \ln [P_\eta(\Delta E)] \quad (2.3)$$

Since the probability distribution ( $P_\eta(\Delta E)$ ) is Gaussian, the free energy curves are parabolic. One of the main assumptions of Marcus Theory about electron transfer stipulates  $\Delta E$  fluctuations producing interesting results of which we give a brief summary here.

Let's consider that the averaged energy gap of oxidized state,  $\langle \Delta E \rangle_O$ , at the center of the parabolic energy curve,

$$A_O = a(\Delta E - \langle \Delta E \rangle_O)^2, \quad (2.4)$$

in which we choose the curve minimum as the zero of our free energy scale, and  $a$  sets the curvature of the parabola and is related to the reorganization free energy,  $\lambda$ . Since the variable on the x-axis is the vertical energy gap, it should be the subtraction of  $A_O$  from  $A_R$ ,

$$A_R = A_O - \Delta E. \quad (2.5)$$

To find the position of minimum of the second curve, solving  $\frac{dA_R}{d\Delta E} = 0$  and using Equation 2.4, we can conclude that

$$\langle \Delta E \rangle_R = \langle \Delta E \rangle_O + \frac{1}{2a}. \quad (2.6)$$

the expression of the second curve can be a combination of Equation 2.4, Equation 2.6 and Equation 2.5

$$A_R = a(\Delta E - \langle \Delta E \rangle_R)^2 - \langle \Delta E \rangle_R + \frac{1}{4a}. \quad (2.7)$$

Thus the free energy profile of the reduced state is also a parabola with the same curvature parameter  $a$  as the curve of the oxidized state, but vertically shifted by  $-\langle \Delta E \rangle_R + \frac{1}{4a}$ .

Therefore, the free energy parabola of the reduced state is also the same as the oxidized state, which was shifted in y axis by  $-\langle \Delta E \rangle_R + \frac{1}{4a}$  as the difference between two curvatures. This shift is called redox potential or reaction free energy and using Equation 2.6, it can be written as

$$\Delta A = -\frac{1}{2}(\langle \Delta E \rangle_R + \langle \Delta E \rangle_O) \quad (2.8)$$

The reorganization free energy,  $\lambda$ , can now be defined as  $\lambda = \Delta E - \Delta A$ , which can be combined with Equation 2.8 to give:

$$\lambda = \frac{1}{2}(\langle \Delta E \rangle_R - \langle \Delta E \rangle_O). \quad (2.9)$$

By combining Equation 2.9 with Equation 2.6, we see that the parameter  $a$  is connected to the reorganization free energy by  $a = (4\lambda)^{-1}$ .

Finally, we can write for the probability to measure a certain  $\Delta E$ :

$$P_\eta(\Delta E) = Z^{-1} \exp \left[ -\frac{A_\eta(\Delta E)}{k_B T} \right] = Z^{-1} \exp \left[ -\frac{(\Delta E - \langle \Delta E \rangle_\eta)^2}{4\lambda k_B T} \right], \quad (2.10)$$

which is a Gaussian centered at  $\langle \Delta E \rangle_\eta$  with variance  $\sigma_\eta^2 = 2\lambda k_B T$ , or

$$\lambda_\eta = \frac{\sigma_\eta^2}{2k_B T}, \quad (2.11)$$

which again relates the reorganization free energy to the fluctuations in  $\Delta E$ . Equation 2.11 enables us to compute  $\lambda$  from a single simulation in the oxidized or reduced state and a value of the fluctuations  $\sigma^2$  in  $\Delta E$ .

After comparing each of these numbers using Equation 2.9, the validity of the linear response assumption underlying Marcus' theory and the accuracy of our simulations can be tested. Thus, the calculation of  $\Delta A_\eta$  is

$$\begin{aligned}\Delta A_{O \rightarrow R} &= -\langle \Delta E \rangle_O - \lambda_O \quad \text{and} \\ \Delta A_{R \rightarrow O} &= \langle \Delta E \rangle_R - \lambda_R,\end{aligned}\tag{2.12}$$

which is required for each sampling of a single oxidation state and it can be compared for other samplings via getting  $\Delta A$  obtained with Equation 2.8.

To sum up, two parabolic curves that make up the well-known free energy landscape from the definition for electron transfer of Marcus' theory are illustrated by their minima at  $\langle \Delta E \rangle_O$  and  $\langle \Delta E \rangle_R$  and curvature parameter  $a$ . To draw the free energy diagram, only two parameters are required when using Equation 2.7 (or Equation 2.9 as  $a = (4\lambda)^{-1}$ ).

As a result, the redox potential and the reorganization free energy can be calculated by Equation 2.12 and Equation 2.11 using the vertical energy gap from the single state or it can be calculated by Equation 2.8 and Equation 2.9 using the average vertical energy gaps of both states. The free energy curves can also be obtained as the logarithm of a histogram of the vertical energy gap (Equation 2.3). In this way, we don't need the assumption of linear response theory and Gaussian fluctuations of  $\Delta E$ .

## 2-2 Calculation of $pK_a$ using molecular simulation

The acidity constant can be obtained as the difference in the Helmholtz free energy,  $\Delta A$ , the protonated and deprotonated states of the molecule as:

$$pK_a = \frac{\Delta A}{k_B T \ln(10)}\tag{2.13}$$

with  $k_B$  the Boltzmann constant and  $T$  the absolute temperature. We calculate free energy difference by a constrained DFT-MD. The reaction coordinate, is a geometric function of atomic positions,  $q(\mathbf{r}^N)$ , that characterizes the progress of the protonation reaction from the reactant state to the product state. For each simulation, the sampling of the molecular coordinate is constrained to a certain fixed reaction coordinate value. Integration of the mean force of constraint, *i.e.*, the average force needed to maintain the reaction coordinate constraint during the simulation, results in a free

energy profile as a function of the reaction coordinate:

$$\Delta A(q') = - \int_{q_0}^{q'} dq f(q) \quad (2.14)$$

Here, we choose  $A(q_0) = 0$  and  $q_0$  to be at the free energy minimum of the reactant state. Using the method of Lagrange multipliers for each MD step, we enforce the reaction coordinate constraint. For more general coordinate types, the constraint force contains additional terms that unbiased the measured force to that of the unconstrained ensemble [40].

The efficiency of the reaction free energy profile depends on the description of reaction. The choice of reaction coordinate is entangled for protonation reaction. Typically, a proton transfer process is a non-local process, and involves various molecules: when the proton moves to a neighboring water molecule, another hydrogen of target water molecule jump to another water molecule which nearby the target water molecule and so on. The indication of previous works on this topic in the literature shows that the contribution of subsequent proton transfers to the acidity constant is quite small and it can be easily ignored.

In this thesis, in order to calculate acidity constant, we use two types of reaction coordinates:

- 1- A function that quantifies the coordination number ( $n_c$ )
- 2- The difference between the distance of the proton to the donating flavin nitrogen and the distance of the proton to the accepting water oxygen ( $\Delta d$ )

The number of hydrogens within a distance of the donor atom ( $d_0$ )

$$n_c = \sum_i \frac{1 - \left(\frac{d(X-H_i)}{d_0}\right)^n}{1 - \left(\frac{d(X-H_i)}{d_0}\right)^m}. \quad (2.15)$$

In Equation 2.15, the sum runs over all solvent hydrogen atoms in the system. The value of  $n_c$  is (close to) one in the protonated state and switches smoothly to zero as the distance  $d(X - H)$  increases. The smoothing parameters “n” and “m” were chosen to be proper number according to cut off radius of  $d_0$ . The coordination number coordinate is a commonly used reaction coordinate for this type of  $pK_a$  calculations. The advantage of this coordinate is that it leaves the choice of the solvent molecule that accepts the proton free. A disadvantage is that it is difficult with this coordinate to simulate the reverse proton transfer from the solvent to the molecule, because at very small values of  $n_c$ , the proton can escape into the bulk after which the coordinate does not anymore control the proton position. The series of constrained MD simulations is therefore setup by starting from the protonated state and stepwise decreasing  $n_c$  to generate initial conditions for the other simulations.

The second reaction coordinate types include also the distance of the proton to

the accepting water molecule oxygen:

$$\Delta d = d(X - H) - d(O_w - H). \quad (2.16)$$

Here,  $\Delta d$  is the difference between the distance of the proton to the donating flavin nitrogen and the distance of the proton to the accepting water oxygen. This value is negative in the protonated state, positive in the deprotonated state and zero when the proton is exactly equidistant from the donor and acceptor atoms. Using this reaction coordinate, a series of simulations starting from the deprotonated state can be generated. This deprotonated state can be constructed from a simulation of deprotonated state by adding a proton to a water molecule. Thus, this system can be equilibrated while maintaining the hydrogen bond using the  $\Delta d$  constraint and an addition coordination number restraint on the hydronium ion oxygen.

This constrained worked as a repulsive harmonic wall to neglect that either of the other two hydronium hydrogens would escape to a nearby water molecule. The wall potential on this coordination number was zero as long as its value was larger than 1.6.

## 2-3 Statistical thermodynamics

Statistical thermodynamics makes a connection between macroscopic and atomistic levels of description. It basically specifies to calculate macroscopic quantities from averages over microscopic ones. In the canonical ensemble (i.e. at fixed  $N, V$ , and  $T$ ), the probability  $P$  to find the system in some configuration ( $\mathbf{r}^N, \mathbf{p}^N$ ) depends on the total energy of the configuration given by the Hamiltonian  $\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)$  according to the Boltzmann distribution:

$$P(\mathbf{r}^N, \mathbf{p}^N) = Q^{-1} (h^{3N} N!)^{-1} \exp[-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)] \quad (2.17)$$

Here,  $\mathbf{r}^N$  is the number of different configurations of positions and momenta  $\mathbf{p}^N$  of  $N$  particles in a volume  $V$  at an absolute temperature  $T$ .  $\beta = \frac{1}{k_B T}$  is the inverse temperature and  $k_B = 1.38066 \cdot 10^{-23}$  J/K is the Boltzmann constant and Plank's constant is  $h = 6.62618 \cdot 10^{-34}$  J·s. Here, is the partition function given by

$$Q = (h^{3N} N!)^{-1} \int \int d\mathbf{r}^N d\mathbf{p}^N \exp[-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)] \quad (2.18)$$

and the parenthesized reciprocal term on the right-hand-side of these two equations (2.17 and 2.18) connects the classical distribution to the quantum mechanical distribution. The macroscopic observable  $A$  is obtained by taking the ensemble average of the microscopic property  $a$ :

$$A = \langle a \rangle = \frac{\int \int d\mathbf{r}^N d\mathbf{p}^N a(\mathbf{r}^N, \mathbf{p}^N) \exp[-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)]}{\int \int d\mathbf{r}^N d\mathbf{p}^N \exp[-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)]} \quad (2.19)$$

This expression can numerically be evaluated by sampling the configurations either stochastically, using the Monte Carlo method, or deterministically using Molecular Dynamics. In MD a representative part of all possible configurations should be sampled, so that the time average of property equals the naturally weighted ensemble average  $\langle a \rangle$ :

$$\langle a \rangle = \lim_{t \rightarrow \infty} \frac{1}{t} \int dt' a(\mathbf{r}^N, \mathbf{p}^N) \quad (2.20)$$

An advantage of MD over Monte Carlo is that also dynamical properties can be evaluated.

## 2-4 Molecular dynamics

Computational studies play an crucial role in many fields of chemistry. Computational chemistry is not restricted to small organic molecules in the gas phase. Nowadays large and complex systems, like solutions, proteins, or solids, can be modeled by computational chemistry with many sophisticated methods. These complex systems require methods that thoroughly sample configuration space, such as molecular dynamics (MD) [41].

Newton's law of motion is used for all molecular dynamics simulations in this thesis.

$$\mathbf{F} = m \cdot \mathbf{a} \quad (2.21)$$

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} \quad (2.22)$$

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} \quad (2.23)$$

Here,  $\mathbf{F}$  is the force on a particle,  $\mathbf{a}$  its acceleration,  $\mathbf{v}$  is velocity,  $\mathbf{r}$  is position,  $m$  is the mass of a particle and  $t$  is for time. A different formulation of the equations of motion, which is widely used in DFT-based MD, is based on the Lagrangian function. For a system of  $N$  particles of mass  $m_i$  with kinetic energy  $U_{kin}$  and potential energy  $V$ , the Lagrangian  $\mathcal{L}$  is

$$\mathcal{L}(\mathbf{r}^N, \dot{\mathbf{r}}^N) = U_{kin}(\dot{\mathbf{r}}^N) - V(\mathbf{r}^N) = \sum_{i=1}^N \frac{1}{2} m_i \dot{\mathbf{r}}_i^2 - V(\mathbf{r}^N) \quad (2.24)$$

giving the equations of motion by applying the Euler-Lagrange equation

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i} = \frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} \quad (2.25)$$

This leads to

$$\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i} = m_i \dot{\mathbf{r}}_i \quad (2.26)$$

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i} = m_i \ddot{\mathbf{r}}_i \quad (2.27)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = -\frac{\partial V}{\partial \mathbf{r}_i} \quad (2.28)$$

therefore

$$\mathbf{F}_i = \frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} = m_i \ddot{\mathbf{r}}_i \quad (2.29)$$

which for each particle  $i$  corresponds to Equation 2.21. The force on a particle is equal to minus the derivative of the potential energy. The potential energy can be computed using the presence of other particles and/or the presence of an external field. The force changes the velocities  $\mathbf{v}$  of the particles, and after that, their positions  $\mathbf{r}$ . In a new position, they get new forces and the whole cycle starts again. Integrating the equations of motion gives the new positions. This is done for each time step in the MD simulation. Here the assumption is that the force does not change when the system evolves from  $t$  to  $t+\Delta t$ . A simple (though not numerically stable) way is

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \mathbf{a}\Delta t \quad (2.30)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\Delta t \quad (2.31)$$

To start the simulation, the data for positions are often taken from a structural database (pdb). The general approach for initializing velocities is drawing them from a Maxwell-Boltzmann distribution.

## 2-5 DFT-based molecular dynamics

Force field Molecular Dynamics can not describe systems that involve excited states, some dipole-dipole interactions or chemical reactions. The forces between atoms depend on many parameters such as positions of atoms, polarization of chemical groups, electric field etc. So, using fixed charges on atoms (or residues) can not give proper answers for the specific questions; in those cases the electrons that surround the nuclei, need to be taken into account. In DFT-based Molecular Dynamics molecular dynamics (also known as *ab initio* molecular dynamics), the forces are calculated *ab initio*<sup>†</sup>,

<sup>†</sup> *ab initio* is a Latin term meaning "from the beginning" and is derived from the Latin *ab* ("from") + *initio*, ablative singular of *initium* ("beginning").

which means that without any force field parameters. The theoretical calculations of quantum chemistry describes how the electron density is influenced by the nuclei in the system. In this technique, apart from the starting positions of the atomic cores and the laws of quantum mechanics, nothing else is needed to perform the simulation. A system of nuclei and electrons is described by the Schrödinger Equation Solving this equation for each snapshot is computationally very demanding, as solving the quantum mechanical equations is intrinsically difficult and computationally demanding. The calculation has to be done for degrees of freedom for every atom and every electron.

The starting point is the time-independent Schrödinger equation:

$$\hat{H}\psi(\mathbf{r}_i, \mathbf{R}_j) = E\psi(\mathbf{r}_i, \mathbf{R}_j) \quad (2.32)$$

In the right part of the equation, the wavefunction  $\psi(\mathbf{r}_i, \mathbf{R}_j)$  describes the whole system via spatial coordinates of ions ( $\mathbf{R}_j$ ) and electrons ( $\mathbf{r}_i$ ). The Hamiltonian operator on  $\psi$  gives the eigenvalue  $E$ , which is total energy of the complete system.

The Hamiltonian can be written,

$$\begin{aligned} \hat{H} = & -\frac{\hbar^2}{2m_e} \sum_i \nabla_i^2 - \sum_{i,I} \frac{Z_I e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \\ & - \sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J e^2}{4\pi\epsilon_0 |\mathbf{R}_I - \mathbf{R}_J|} \end{aligned} \quad (2.33)$$

with  $\hbar=2\pi\hbar$  Planck's constant,  $m_e$  and  $M_I$  the masses of an electron and ion  $I$  respectively and  $Z_I$  the atomic number. The first term is the kinetic energy of the electrons, the second term is the electron-ion Coulomb interaction, the third one is the electron-electron Coulomb energy, the fourth term is the ionic kinetic energy and finally the fifth term is the potential energy due to ion-ion Coulomb interaction. Equation 2.32 can be solved analytically up to the system of two protons with one electron. We need to use approximations and numerical methods for larger systems. There are many approximations to address this problem. The reader interested in the technical details of these approximations might wish to read references [42–46].

A good approximation in almost every ground state system is the Born–Oppenheimer (which is also called adiabatic approximation). It is meaningful to assume that the positions of the nuclei are fixed in order to calculate the energy of electrons in the ground state, and so, the kinetic energy of the nuclei can be eliminated from the Hamiltonian. The basic idea of the approximation is that ions are so heavy with respect to electrons, so you can decouple their motion. This leaves us with the Hamiltonian for the electronic structure,

$$\hat{H} = \hat{T} + \hat{V}_{ext} + \hat{V}_{int} + E_{II} \quad (2.34)$$

$\hat{T}$  is the operator giving the kinetic energy of the electrons,  $\hat{V}_{ext}$  gives their potential energy due to the external field of the ions,  $\hat{V}_{int}$  the electron-electron interaction and the constant term  $E_{II}$  is the classical Coulomb interaction of the nuclei.

In DFT-MD, the method of choice to calculate the electronic structure is DFT which is explained in the next section.

### 2-5.1 Density Functional Theory

Density functional theory (DFT) was proven in 1964 by Hohenberg and Kohn [47] using the variational principle for systems with a non-degenerate ground-state. DFT is based on the notion that for a many-electron system there is a one-to-one mapping between the external potential and the electron density:  $v(\mathbf{r}) \leftrightarrow \rho(\mathbf{r})$ . All properties are a functional of the electron density, because the electron density determines the potential, which determines the Hamiltonian, which determines the energy ( $E[\rho]$ ) and the wave function ( $\psi[\rho]$ ), from which all physical properties can be determined.

Hohenberg and Kohn formulated DFT, using the following theorems.

1. In a system of interacting electrons there is an external potential  $V_{ext}$  which is determined uniquely, by the ground state electron density  $\rho_0$ ;
2. There is a universal energy functional  $E[\rho]$ , which depends on electron density.

Applying the variational principle, the ground state energy can be calculated. The Hohenberg-Kohn energy functional has the following form;

$$\begin{aligned} E_{HK}[\rho] &= T[\rho] + E_{int}[\rho] + \int d\mathbf{r} V_{ext}(\mathbf{r})\rho(\mathbf{r}) \\ &= F_{HK}[\rho] + \int d\mathbf{r} V_{ext}(\mathbf{r})\rho(\mathbf{r}) \end{aligned} \quad (2.35)$$

The electron density is a function of  $\mathbf{r}$ , with  $\rho(\mathbf{r}) \geq 0$ . Integrated over the entire space, it must be equal to the total number of electrons.  $F_{HK}[\rho]$  is universal which means the position of the nuclei is not important. The interaction of electrons with the external potential of the nuclei is shown by the integral term in Equation 2.35.

According to second theorem, the ground state energy  $E_0$  is

$$E_0 = E[\rho_0(\mathbf{r})] \leq E[\rho(\mathbf{r})] \quad (2.36)$$

Here,  $\rho_0$  is the ground state density. Since the total number of electron is constant, using variational principle, we can get the minimum energy. The result is given by Euler-Lagrange equation, with Lagrange multiplier  $\mu$  being the chemical potential of the electrons

$$\mu = \frac{\partial E[\rho(\mathbf{r})]}{\partial \rho(\mathbf{r})} = V_{ext}(\mathbf{r}) + \frac{\partial F_{HK}[\rho(\mathbf{r})]}{\partial \rho(\mathbf{r})} \quad (2.37)$$

The Hohenberg-Kohn equation, which is given in Equation 2.35, is exact. However in practice it is not useful because the precise energy functional is not known. This

is addressed by the Kohn–Sham method [48] that provides an alternative description of the functional. It, basically, introduces a reference system, within an exchange correlation functional that combines the differences between the real and the reference system [48].

The density of the auxiliary system is built up by the squares of the so called Kohn–Sham orbitals  $\phi_i$ . Although these orbitals are not the ‘real’ orbitals, they are often used as a good representation and add up to the real density

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2 \quad (2.38)$$

where  $2N$  is the number of electrons in the system ( $2N$  electrons fill  $N$  orbitals). In a simulation, the orbitals will be built up from a suitable basis set. The kinetic energy of these these non-interacting electrons is

$$T_S = -\frac{\hbar^2}{2m_e} \sum_{i=1}^N \langle \phi_i | \nabla_i^2 | \phi_i \rangle \quad (2.39)$$

The Coulomb interaction between the electrons is calculated as the Hartree energy

$$E_{Hartree}[\rho] = \frac{e^2}{2} \int d^3\mathbf{r} d^3\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{4\pi\epsilon_0|\mathbf{r}-\mathbf{r}'|} \quad (2.40)$$

Note that this is not exactly the same as the  $V_{int}$  term in Equation 2.34, as Equation 2.40 includes an extra self–interaction term. Because the integration runs over the complete density twice, all electrons necessarily interact with themselves. The potential field from the nuclei now interacts with the electron density instead of the separate electrons. For any point in an external field of  $N$  ions,  $V_{ext}$  is

$$V_{ext}(\mathbf{r}) = -\sum_{I=1}^N \frac{Z_I e}{4\pi\epsilon_0|\mathbf{r}-\mathbf{R}_I|} \quad (2.41)$$

Hence, the energy functional is

$$E_{KS} = T_S + \int d^3\mathbf{r} V_{ext}(\mathbf{r})\rho(\mathbf{r}) + E_{Hartree}[\rho] + E_{XC}[\rho] \quad (2.42)$$

The exchange correlation functional energy  $E_{XC}$  is the difference between all other terms and the universal Hohenberg–Kohn functional (Equation 2.35), so

$$E_{XC} = F_{HK}[\rho] - (T_S + E_{Hartree}[\rho]) \quad (2.43)$$

$E_{XC}$  includes the difference in the kinetic energy of the real and the auxiliary system and the difference between the Hartree energy and the real electron–electron interaction. If  $E_{XC}$  is known exactly, then  $E_{KS}$  is exact. However, in practice it must be estimated. When this estimation is sufficiently good, then the Kohn–Sham approach

provides reasonable results. Easy and widely-used is the Local Density Approximation (LDA),

$$E_{XC}^{LDA} = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{XC}([\rho(\mathbf{r})], \mathbf{r}) \quad (2.44)$$

with  $\epsilon_{XC}$  the exchange-correlation energy for each electron in a homogeneous electron gas at constant density. This  $\epsilon_{XC}$  is calculated for many densities with very accurate quantum Monte Carlo simulations. The simple LDA calculations works for solid state applications. On the other hand, a quantitative study of chemical reactions is problematic, as a high accuracy of a few kcal/mol is needed. To get better results, a generalized gradient approximation (GGA) improves the LDA. A commonly used and reasonably accurate example of this is the BLYP functional which combines the exchange energy approximation of Becke [49] with the correlation energy approximation of Lee, Yang and Parr [50]. Another important example is Perdew-Burke-Ernzerhof (PBE) [51] functional which we used for this thesis.

## 2-5.2 Born–Oppenheimer molecular dynamics

The potential energy is the Kohn–Sham energy, leading to the Lagrangian for the dynamics of the atomic nuclei.

$$\mathcal{L}_{BO}(\mathbf{R}^N, \dot{\mathbf{R}}^N) = \sum_{I=1}^N \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 - \min_{\{\phi_i\}} E^{KS}(\{\phi_i\}; \mathbf{R}^N) \quad (2.45)$$

Under the condition that the minimization of  $E^{KS}$  leads to orthogonal wavefunctions  $\phi_i$ , the Euler–Lagrange equation gives

$$M_I \ddot{\mathbf{R}}_I = -\nabla_i (\min_{\{\phi_i\}} E^{KS}(\{\phi_i\}; \mathbf{R}^N)) \quad (2.46)$$

This result is same as Hellmann-Feynman theory [52] that gives the force on ion  $I$  for a system a known wavefunction  $\psi$

$$F_I = -\frac{\partial E}{\partial \mathbf{R}_I} = -\langle \psi | \frac{\partial \hat{H}}{\partial \mathbf{R}_I} | \psi \rangle \quad (2.47)$$

As orthogonality constraints ( $\Lambda_{ij}$ ) are used for the orbitals, this leads to an extended energy functional to be minimized (only required if  $\psi$  is a dynamical degree of freedom, e.g. in Car Parrinello dynamics)

$$\epsilon^{KS} = E^{KS} + \sum_{ij} \Lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}) \quad (2.48)$$

After minimization, the forces on the ions can be calculated by combining Equation 2.48 and Equation 2.46. Here, the new ionic velocities, new ion positions, new ground state energy, new force are calculated from the force.

## 2-6 Hybrid quantum mechanics / molecular mechanics

In this thesis, hybrid Quantum Mechanics / Molecular Mechanics (QM/MM) method, which is the combination of quantum mechanics and molecular mechanics, also referred to as refer to force fields, is also applied. Although the convenience of Force Field Molecular Dynamics methods for much larger system with longer simulation times is evident, it can be difficult to design accurate force fields, in particular for chemically reactive systems. The general idea of the QM/MM approach is large chemical systems may be partitioned into an chemically active region which requires a quantum mechanical treatment (QM level) and a remaining part which can be modeled by a force field (MM level).

QM/MM, which was introduced by Warshel and Levitt [53], enables to perform virtual experiments for realistic large systems, like reactive biomolecular structures at a reasonable computational effort while providing sufficient accuracy. The method induced the development of new research areas [54–62, 62–67], open up new branch of biological chemistry and even gave rise to define new terminology ‘in silico’.

One of the main goals of QM/MM is to make able to select a specific, small QM region, that is handled at the QM level, in a large biomolecular system that are set up using experimental data, e.g. from the Protein Data Bank [68]. The selection of the partitioning requires care and is dependent on the nature of system, e.g. the extend of the region that is involved in charge transfer or in a rearrangement of chemical bonding. The accessible time scale in QM/MM simulations is set by that of the QM part and is of the order of 10-100 ps. To address processes that occur on a longer time scale one can employ biased simulation methods such as metadynamics [69, 70], blue moon [71, 72]. These methods enable the enhancement of the sampling of rare events thereby expanding the simulations not only with respect to the size of the system but also with respect to the timescale.