Ab-Initio Molecular Dynamics Study of Structural Modifications at Aromatic Aminoacids Triggered by Al(III)

Julen Larrucea

14 May 2007

Contents

1 Introduction **5**

2 An introduction to Molecular Dynamics **7**
2.1 Introduction to The Equations of Motion 7
2.1.1 The Lagrangian formalism 7
2.1.2 Propagation algorithms 8
2.2 Ab-initio Molecular Dynamics 10
2.2.1 Born-Oppenheimer Molecular Dynamics 12
2.2.2 Car-Parrinello molecular dynamics 12

3 Setting up the CPMD simulations **15**
3.1 General parameters 15
3.2 Choosing of the pseudopotentials and the optimum cutoff 15

4 Results of the Molecular dynamics simulations **17**
4.1 Molecular Dynamics of the aminoacids interacting with Al(III) . 17
4.1.1 Phenylalanine (Phe) 17
4.1.2 Tryptophan (Trp) 20
4.1.3 Tyrosine (Tyr) 20
4.1.4 Discussion 24
4.1.5 Conclusions 25

5 Acknowledgements **29**

A Car-Parrinello Molecular Dynamics using CPMD **31**

Chapter 1

Introduction

This project aims at help in the process to understanding the protein modification in presence of Al(III), for this task we have chosen to study the system by computational methods. Previous theoretical studies have been done in order to conclude the most probable binding site for the aluminium [1]. As Al(III) or (Al^{+3}) has positive charge, it will produce stronger bonds if it gets bonded to the aminoacid side chains with large electronic density, namely highly base centers. From the 21 aminoacids in ature there are just three of them which are aromatic [2], and the aromatic aminoacids (AAA) are going to be more interesting than the rest because they can in adition to the interactions all aminoacids produce, they produce also some other ones, like cation-π interactions [3]. These three aromatic aminoacids are Phenylalanine (Phe), Tryptophan (Trp) and Tyrosine (Tyr) and are the ones that we are going to study here.

In the previous studies, from where structural and stability information was obtained, were performed by using static calculations, therefore there was not any information about the structural propagation of the system. This work will use ab-initio molecular dynamics to describe the dynamics of different systems that in all the cases consist of an aminoacid (one of the three above mentioned), one Al(III) atom and several structural water molecules.

Ab-initio molecular dynamics is needed in order to be able to describe with enough accuracy bonding patterns and rearrangements in molecular systems. The method used for these calculations is the Car-Parrinello Molecular Dynamics, because of its accuracy and efficient use of the computational facilities. The Car-Parrinello method can be described shortly as a molecular dynamics method that calculates the valence electrons by pure ab-initio techniques, describes the inner electrons by pseudopotentials and treats classically the motion of the nuclei.

Chapter 2

An introduction to Molecular Dynamics

2.1 Introduction to The Equations of Motion

2.1.1 The Lagrangian formalism

In classical mechanics natural phenomena are governed by Newton's equations. For dynamical systems one choice for handling the system is to build a generalized function that describes concisely the equations of motion and some other important properties of the system. The function that does this job is called the Lagrangian.

We can easily build a simple classical dynamic Lagrangian function by substracting the kinetic energy of a system to its potential energy.

$$\mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) = \mathcal{T} - \mathcal{V}. \tag{2.1}$$

Let's use the Lagrangian formalism in order to make a little simulation on a two body system with a spring like potential. A much better classical description could be done, but I would like to minimize the amount of variables in order to increase the speed of understanding. So, let it be a system where two spheres of 1 kg mass each are bonded by a spring with spring constant of 5 N/m and the two bodies are placed to a distance of 8 meters apart.

The kinetic energy for the two particle system is straightfordwardly

$$T = \frac{1}{2} m_1 \dot{\mathbf{x}}_\mathbf{1} + \frac{1}{2} m_2 \dot{\mathbf{x}}_\mathbf{2},$$

and the potential energy is given by Hooke's law

$$V = \frac{1}{2} k (\mathbf{x_2} - \mathbf{x_1})^2.$$

So the Lagrangian for this system will look like

$$\mathcal{L}(\mathbf{x_1}, \mathbf{x_2}, \dot{\mathbf{x}}_\mathbf{1}, \dot{\mathbf{x}}_\mathbf{2}) = \frac{1}{2} m_1 \dot{\mathbf{x}}_\mathbf{1} + \frac{1}{2} m_2 \dot{\mathbf{x}}_\mathbf{2} - \frac{1}{2} k (\mathbf{x_2} - \mathbf{x_1})^2 \tag{2.2}$$

Let us apply the Euler-Lagrange equations in order to derive the equations of motion from the Lagrangian. Euler-Lagrange equations are

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}_i}\right) - \frac{\partial \mathcal{L}}{\partial \mathbf{x}_i} = 0 \quad , \quad i = 1, \ldots, N. \tag{2.3}$$

We now just need to substitute the different elements in these equations by deriving the classical Lagrangian of equation 2.1, so we first derive all the elements independently, namely $\mathbf{x_i}$ and $\dot{\mathbf{x}}_i$ are treated as independent variables.

$$\begin{array}{l} \frac{d\mathcal{L}}{d\mathbf{x_1}} = -\frac{dV}{d\mathbf{x_1}} = -k(\mathbf{x_2} - \mathbf{x_1})(-1) = k(\mathbf{x_2} - \mathbf{x_1}) \\ \frac{d\mathcal{L}}{d\mathbf{x_2}} = -\frac{dV}{d\mathbf{x_1}} = -k(\mathbf{x_2} - \mathbf{x_1}) = k(\mathbf{x_1} - \mathbf{x_2}) \\ \frac{d\mathcal{L}}{d\dot{\mathbf{x}}_1} = m_1\dot{\mathbf{x}}_1 \\ \frac{d\mathcal{L}}{d\dot{\mathbf{x}}_2} = m_2\dot{\mathbf{x}}_2 \end{array}$$

And substituting these elements in the eq. 2.3 we get the equations of motion for this simple system.

$$\begin{array}{l} \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}_1}\right) - \frac{\partial \mathcal{L}}{\partial \mathbf{x}_1} = m_1\ddot{\mathbf{x}}_1 - k(\mathbf{x_2} - \mathbf{x_1}) = 0 \rightarrow m_1\ddot{\mathbf{x}}_1 = k(\mathbf{x_2} - \mathbf{x_1}) \\ \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}_2}\right) - \frac{\partial \mathcal{L}}{\partial \mathbf{x}_2} = m_2\ddot{\mathbf{x}}_2 - k(\mathbf{x_1} - \mathbf{x_2}) = 0 \rightarrow m_2\ddot{\mathbf{x}}_2 = k(\mathbf{x_1} - \mathbf{x_2}) \end{array} \tag{2.4}$$

There are many different algorithms to handle the propagation of the system by calculating the trajectories of the particles.

2.1.2 Propagation algorithms

For this simple example we take the Euler's propagation algorithm because it is the simplest one. The new positions ($\mathbf{x}$) and the velocities ($\dot{\mathbf{x}}$) of the particles are given by

$$\begin{array}{l} \mathbf{x}(t_0 + \delta t) = \mathbf{x}(t_0) + \dot{\mathbf{x}}(t_0)\delta t \\ \dot{\mathbf{x}}_i(t_0 + \delta t) = \dot{\mathbf{x}}_i(t_0) + \ddot{\mathbf{x}}_i(t_0)\delta t \end{array} \tag{2.5}$$

If we substitute the acceleration ($\ddot{\mathbf{x}}_\mathbf{i}$) from eq. 2.4 in the Euler's integrator (eq. 2.5) we get the following equations for the velocities:

$$\begin{array}{l} \dot{\mathbf{x}}_1(t_0 + \delta t) = \dot{\mathbf{x}}_1(t_0) + \frac{k}{m_1}\left(\mathbf{x_2} - \mathbf{x_1}\right)\delta t \\ \dot{\mathbf{x}}_2(t_0 + \delta t) = \dot{\mathbf{x}}_2(t_0) + \frac{k}{m_2}\left(\mathbf{x_1} - \mathbf{x_2}\right)\delta t \end{array}$$

It is necessary to set initial positions and velocities for the particles 1 and 2 in order to start the propagation of the trajectories. Let them be

$$\begin{array}{ll} \mathbf{x}_1(t_0) = (-4.0)m & \dot{\mathbf{x}}_1(t_0) = (1.0)m/s \\ \mathbf{x}_2(t_0) = (4.0)m & \dot{\mathbf{x}}_2(t_0) = (-1.0)m/s \end{array}$$

The expression δt in eq. 2.5 determines how often we want to get new positions and recalculate the parameters. Let us recalculate the positions and parameters every 1 second. In molecular dynamics δt is known as "**time step**". So our time step (ts) will be 1 second and the new positions and velocities starting from $t_0 = 0$ will be

$$\begin{array}{l} \mathbf{x_1}(1) = (-4.0) + (1,0)1 = (-3.0)m \\ \dot{\mathbf{x}}_\mathbf{1}(1) = (1.0) + (5\frac{(4.0)-(-4.0)}{1})1 = (41.0)m/s \\ \mathbf{x_2}(1) = (4,0) + (-1,0)1 = (3,0)m \\ \dot{\mathbf{x}}_\mathbf{2}(1) = (-1.0) + (5\frac{(-4.0)-(4.0)}{1})1 = (-41.0)m/s \end{array}$$

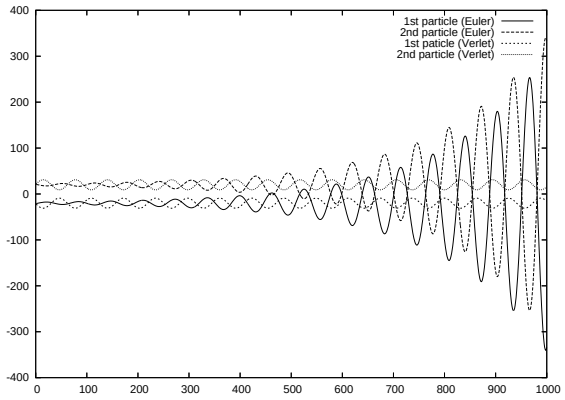


Figure 2.1: Comparison of the motion of two particles with two different propagation algorithms: Euler and Verlet. Euler algorithm produces a feedback and the oscillation of the particles keep getting bigger and bigger amplitude. On other hand Verlet algorithm keeps the particles oscillation in the with a constant amplitude

Now that we have already got the positions and velocities for the first step. let us make a second one:

$$\begin{aligned}
\mathbf{x_1}(2) &= (-3.0) + (1.0)1 = (-2.0)m \\
\dot{\mathbf{x}}_\mathbf{1}(2) &= (1.0) + (5\tfrac{(3.0)-(-3.0)}{1})1 = (31.0)m/s \\
\mathbf{x_2}(2) &= (3.0) + (-1.0)1 = (2.0)m \\
\dot{\mathbf{x}}_\mathbf{2}(2) &= (-1.0) + (5\tfrac{(-3.0)-(3.0)}{1})1 = (-31.0)m/s
\end{aligned}$$

... and so on.

So this way the system evolves, and we get new positions for the particles each time step. In the same way, it is possible to calculate not just the positions and velocities, but also many other properties, like the energy (by time derivative to the Lagrangian), frequencies (from harmonic oscillator model), ... and when we use this method for quantum calculations, it is possible to get most of the molecular properties by using statistical mechanics.

Instead of Euler algorithm it is possible some other propagation algorithms can be applied as well. One of the most common ones in molecular dynamics is the Verlet algorithm. Verlet algorithm is written as

$$\begin{aligned}
\mathbf{x}(\mathbf{t_0} + \delta\mathbf{t}) &= \mathbf{2x_0} + \mathbf{x}(\mathbf{t} - \delta\mathbf{t}) + \dot{\mathbf{x}}\delta\mathbf{t^2} \\
\ddot{\mathbf{x}}(\mathbf{t_0}) &= \frac{\mathbf{x}(\mathbf{t_0} + \delta\mathbf{t}) - \mathbf{x}(\mathbf{x}(\mathbf{t_0} + \delta\mathbf{t})}{\mathbf{2}\delta\mathbf{t}}.
\end{aligned}$$

The Verlet algorithm offers a much better improvement with respect to Euler, since as we can see in Fig. 2.1 it does not produce a feedback error.

2.2 Ab-initio Molecular Dynamics

The simulation of a molecular system requires a different mathematical description than the one in previous section, in order to get reliable results. In an Ab-Initio Quantum Mechanics description of the dynamics, as a difference with the previous one governed by classical (Newtonian) equations, the equations of motion are going to be determined by the time-dependent Schrodinger equation [4]

$$i\hbar\frac{\partial}{\partial t}\Phi(\{\mathbf{r}_i\},\{\mathbf{R}_i\};t) = \mathcal{H}\Phi(\{\mathbf{r}_i\},\{\mathbf{R}_i\};t). \tag{2.6}$$

In this equation one must define a Hamiltonian operator ($\mathcal{H}$) that acts on the particle's wavefunction (Φ) the same way as the Lagrangian operator acts on the position in the previous section. An standard Hamiltonian operator for a molecular system can be built as the sum of the kinetic and potential energy for nuclei and electrons and a correlation term between nuclei and electrons [5]

$$\begin{aligned}
\mathcal{H} &= -\sum_I \frac{\hbar^2}{2M_i}\nabla^2{}_I - \sum_i \frac{\hbar^2}{2m_e}\nabla^2{}_i + \sum_{i<j}\frac{e^2}{|\mathbf{r_i}-\mathbf{r_j}|} + -\sum_{I,i}\frac{e^2Z_I}{|\mathbf{R_I}-\mathbf{r_i}|} + \sum_{I<J}\frac{e^2Z_IZ_I}{|\mathbf{R_I}-\mathbf{R_J}|} \\
&= -\sum_I \frac{\hbar^2}{2M_i}\nabla^2{}_I - \sum_i \frac{\hbar^2}{2m_i}\nabla^2{}_i + V_{n-e}(\{\mathbf{r_i}\},\{\mathbf{R_I}\}) \\
&= -\sum_I \frac{\hbar^2}{2M_i}\nabla^2{}_I + \mathcal{H}_e(\{\mathbf{r_i}\},\{\mathbf{R_I}\}).
\end{aligned} \tag{2.7}$$

where $\{\mathbf{r_i}\}$ are the electronic and $\{\mathbf{R_I}\}$ the nuclear degrees of freedom respectively.

The wave function Φ, needs to have both nuclear and electronic coordinates separated, and this can be approached in a simple way, just by writing them separately and adding a phase factor, as

$$\Phi(\{\mathbf{r_i}\},\{\mathbf{R_I}\};t) \approx \Psi(\{\mathbf{r_i}\};t)\chi(\{\mathbf{R_I}\};t)exp\left[\frac{i}{\hbar}\int_{t_0}^{t} dt'\tilde{E}_e(t')\right], \tag{2.8}$$

where the nuclear and electronic wavefunction are separately normalized to unity at every instant of time, i.e. $\langle\chi;t|\chi;t\rangle = 1$ and $\langle\Psi;t|\Psi;t\rangle = 1$, respectively. In addition, a convenient phase factor was introduced as well such that the final equation looks more compact, so that

$$\tilde{E}_e = \int d\mathbf{r}d\mathbf{R}\Psi^*(\{\mathbf{r_i}\};t)\chi^*(\{\mathbf{R_I}\};t)\mathcal{H}_e\Psi(\{\mathbf{r_i}\};t)\chi(\{\mathbf{R_I}\};t),$$

what is known as one determinant or single-configuration ansatz for the total wavefunction. This ansatz differs from the Born-Oppenheimer one (excepting the phase factor) in the fact that it separates the fast and slow variables. Born-Oppenhemier ansatz is written as

$$\Phi_{BO}(\{\mathbf{r_i}\},\{\mathbf{R_I}\};t) = \sum_{k=0}^{\infty}\tilde{\Psi}_k(\{\mathbf{r_i}\},\{\mathbf{R_I}\};t)\tilde{\chi}_k(\{\mathbf{R_I}\};\mathbf{t}). \tag{2.9}$$

After introducing the expression 2.7 and 2.8 in the equation 2.6, multiply the left part by $\langle\Psi|$ and $\langle\chi|$ respectively and imposing energy conservation $d\langle\mathcal{H}\rangle/dt \equiv 0$ the result is:

$$\begin{aligned} i\hbar\tfrac{\partial\Psi}{\partial t} &= -\textstyle\sum_i \tfrac{\hbar^2}{2m_e}\nabla^2{}_i\Psi + \left\{\int d\mathbf{R}\chi^*(\{\mathbf{R_I}\};t)V_{n-e}(\{\mathbf{r_i}\},\{\mathbf{R_I}\})\chi(\{i\mathbf{R_I}\ t)\right\}\Psi \\ i\hbar\tfrac{\partial\chi}{\partial t} &= -\textstyle\sum_i \tfrac{\hbar^2}{2M_I}\nabla^2{}_i\chi\left\{\int d\mathbf{r}\Psi^*(\{\mathbf{r_i}\};t)\mathcal{H}_e(\{\mathbf{r_i}\},\{\mathbf{R_I}\})\Psi(\{\mathbf{r_i}\};t)\right\}\chi \end{aligned} \tag{2.10}$$

what defines the basis of the time-dependent self-consistent field (TDSCF), because both electrons and nuclei move quantum-mechanically in time-dependent effective potentials (or self-consistently obtained average fields).

At this point, an approximation is needed in order to describe nuclei as classical particles. This is achieved by rewriting the nuclear wavefunction

$$\chi(\{\mathbf{R_i}\};t) = A(\{\mathbf{R_I}\};t)e^{\left[\frac{iS(\{\mathbf{R_I}\};t)}{\hbar}\right]} \tag{2.11}$$

in terms of amplitude factor **A** and phase **S**. So introducing this new ansatz for the nuclear wavefunction in TDSCF equation 2.10

$$\tfrac{\partial S}{\partial t} + \textstyle\sum_I \tfrac{1}{2M_I}(\nabla_I S)^2 + \int d\mathbf{r}\Psi^*\mathcal{H}_e\Psi = \hbar^2\sum_I \tfrac{1}{2M_I}\tfrac{\nabla_I^2 A}{A} \tag{2.12}$$

$$\tfrac{\partial A}{\partial t} + \textstyle\sum_I \tfrac{1}{M_I}(\nabla_I S) + \sum_I \tfrac{1}{2M_I}A(\nabla_I^2 S) = 0 \tag{2.13}$$

we get what is known as "quantum fluid dynamical representation" from where time-dependent Schrödinger equation can be solved. If the classical limit is taken as $\hbar \to 0$ the term on the right of the eq. 2.12 disappears, hence

$$\frac{\partial S}{\partial t} + \sum_I \frac{1}{2M_I}(\nabla_I S)^2 + \int d\mathbf{r}\Psi^*\mathcal{H}_e\Psi = 0$$

which it looks like the Hamilton-Jacobi formulation

$$\Psi = \frac{\partial S}{\partial t} + \mathcal{H}(\{\mathbf{R}\}_\mathbf{I}, \{\nabla_\mathbf{I}\mathbf{S}\}) = \mathbf{0}$$

of classical dynamics with classical Hamilton function

$$\mathcal{H}(\{\mathbf{R}\}_I, \{\mathbf{P}\}_I) = T(\{\mathbf{P}\}_I) + V(\{\mathbf{P}\}_I)$$

where $\mathbf{P} \equiv \nabla_I S$ and the Newtonian equation of motion corresponding to eq. 2.12

$$\frac{d\mathbf{P_I}}{dt} = -\nabla_I \int dt\Psi^*\mathcal{H}_e\Psi$$

or written in another way

$$\begin{aligned} M_I\ddot{\mathbf{R}}_\mathbf{I}(t) &= -\nabla_I \int d\mathbf{r}\Psi^*\mathcal{H}_e\Psi \\ &\quad -\nabla_I V_e^E(\{\mathbf{R_I}\}(t)), \end{aligned} \tag{2.14}$$

so, nuclei move according to classical mechanics in an effective potential V_e^E due to the electrons. This effective potential can be developed starting from eq. 2.10 so that

$$V_e^E = \int d\mathbf{r}\Psi_0^*\mathcal{H}_e\Psi_0 \equiv E_0(\{\mathbf{R_I}\}) \tag{2.15}$$

and the nuclei will move under it. The potential in eq. 2.15 can be computed from the time-independent Schrödinger equation, which is usually approached in terms of truncated expansion of many-body contributions

$$V_e^E \approx V_e^{approx}(\{\mathbf{R_i}\}) = \sum_{I=1}^{N} v_1(\mathbf{R_I}) + \sum_{I<J}^{N} v_2(\mathbf{R_I}, \mathbf{R_J}) + \sum_{I<J<K}^{N} v_3(\mathbf{R_I}, \mathbf{R_J}, \mathbf{R_K}) + \dots \tag{2.16}$$

2.2.1 Born-Oppenheimer Molecular Dynamics

Born-Oppenheimer approach uses classical equations for describe the nuclei and adds the electronic structure by solving statically (time-independent) Schrödinger equation for each time step. The resulting molecular dynamics method is defined by

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I min_{\{\Psi_0\}}\{\langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle\} \tag{2.17}$$

$$E_0 \Psi_0 = \mathcal{H}_e \Psi_0. \tag{2.18}$$

So the electronic ground state has to be reached every step. After applying Hartree-Fock approximation

$$\begin{array}{c} M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I min_{\{\psi_i\}} \left\langle \Psi_0 | \mathcal{H}_e^{HF} | \Psi_0 \right\rangle \\ 0 = -\mathcal{H}_e^{HF}\psi_i + \sum_i \Lambda_{i,j}\psi_j. \end{array} \tag{2.19}$$

This can be also written according to Lagrange's formalism

$$\mathcal{L} = -\langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle + \sum_{i,j} \Lambda_{i,j}(\langle \psi_i | \psi_j \rangle - \delta_{i,j}), \tag{2.20}$$

where $\Lambda_i j$ are the associated Lagrangian multipliers. And deriving this Lagrangian with repect to the orbitals

$$\frac{\delta \mathcal{L}}{\delta \psi_i^* = 0}$$

leads to Hartree-Fock equations

$$\mathcal{H}_e^{HF}\psi_i = \sum_j \Lambda_{ij}\psi_j.$$

2.2.2 Car-Parrinello molecular dynamics

Car-Parrinello method proposes a special Lagrangian for the motion of the nuclei, which in addition to the terms of potential energy and constraints in eq. 2.20 contains also two terms for the kinetic energy of nuclei and electrons [6]. This last one is described in a pseudoclassical way by using a "fictitious" mass μ term for the electrons and a time derivative of the electronic wavefunctions for describe some kind of "velocity"

$$\mathcal{L}_{CP} = \sum_i \frac{1}{2}\mu_i \left\langle \dot{\psi}_i | \dot{\psi}_i \right\rangle + \sum_i \frac{1}{2} M_i \dot{\mathbf{R}}_\mathbf{i}^\mathbf{2} - \langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle + \sum_{i,j} \Lambda_{i,j} \left(\langle \psi_i | \psi_i \rangle - \delta_{i,j} \right). \tag{2.21}$$

In the same way as in the classical model, the Newtonian equations of motion can be obtained by Euler-Lagrange equation (Eq. 2.3)

$$\begin{array}{c} \frac{d}{dt}\left(\frac{\partial \mathcal{L}_{CP}}{\partial \dot{\mathbf{R}}_i}\right) - \frac{\partial \mathcal{L}_{CP}}{\partial \mathbf{R}_i} = 0 \\ \frac{d}{dt}\left(\frac{\partial \mathcal{L}_{CP}}{\partial \dot{\psi}_i^*}\right) - \frac{\partial \mathcal{L}_{CP}}{\partial \psi_i^*} = 0 \end{array} \tag{2.22}$$

and after solving and applying Hartree-Fock approximation, the Car-Parrinello equations of motion are

$$\begin{array}{c} M_I \ddot{\mathbf{R}}_I(t) = -\frac{\delta E}{\delta bf R_I} \\ \mu_i \ddot{\psi}_i(t) = -\frac{\delta E}{\delta \psi_i^*} + \sum_i \Lambda_{i,j}\psi_j \end{array} \tag{2.23}$$

which at variance with eq. 2.20 does not need to reach the ground state every step.

In order to solve the 2^{nd} and 3^{rd} terms in eq. 2.21 it is necessary to solve the electronic Schrödinger equation [4] but this becomes too expensive so it is necessary to make approximations and one of the main ones is given by the Density Functional Theory.

Chapter 3

Setting up the CPMD simulations

3.1 General parameters

The value of the Electronic fictitious mass (EMASS) was left as default at 400 a.u. and the time step was choose to be 3 a.u. (0.072 fs) for ensure the adiabatic decoupling of the ions and electrons, and therefore I got almost no drift of the Electronic Kinetik Energy along the simulations, keeping the electrons "cold" every moment. The temperature of the system (temperature of Ions $\sum_I M_I \dot{\mathbf{R}}_\mathbf{I}^2$) was controlled by scaling at 280 ±20 K, for more realistic simulation on biologic environment.

The global system's charge was set to +3, so that we assume no counterions around Al^{+3} .

Calculations were performed at SGI and I2bask [7].

The starting molecular structures where obtained from previously made Gaussian 03 calculations [2] at B3LYP/6-311++G(2df,2p).

3.2 Choosing of the pseudopotentials and the optimum cutoff

There are many different kind of pseudopotentials available for describing the atoms that form our system (H, O, C, N, Al), but I choose the Vanderbilt's (VDB) ones [8] with the implementation for ultrasoft fitting to the All-electron shape by Laasonen et al. [9], because they make possible to perform calculations by using lower plane wave basis cutoff and reducing therefore the computational cost significantly without loosing accuracy [10].

After choosing the type of pseudopotentials, I choose the most interesting ones between all the available Vanderbilt Ultrasoft Pseudopotential database possibilities by using as a criteria the theoretical method behind and also the ability to get higher accuracy at lower plane wave cutoffs. In figure: 3.1 we see how the energy tends to converge while increasing the cutoff. We can see that

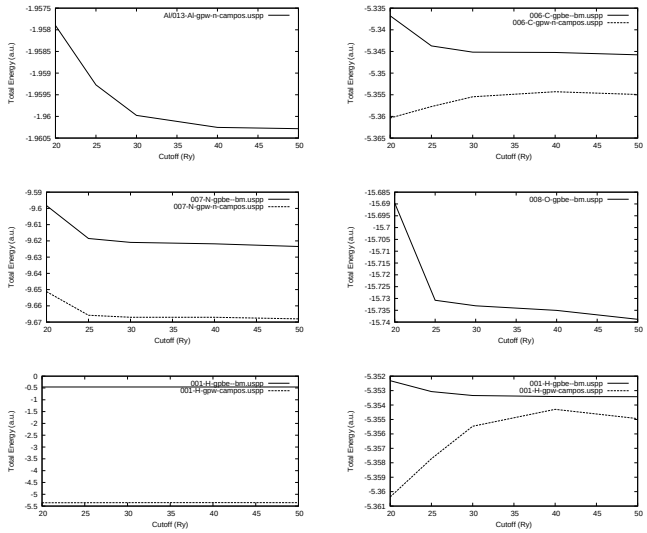


Figure 3.1: Graphical representation of the Total Energy vs plane wave cutoff for some Vanderbilt pseudpotentials. Last figure is scaled for appreciate the convergency curve.

the "gpbe" (PBE 96) pseudopotentials converge at lower cutoffs in the case of H and O , so I will use them for this atoms. On other hand, for the case of Al, N and C, the Campos's ones (PW 91) converge faster, therefore, I will use the them for Al, N and C atoms. In all the cases the energy is already well converged for cutoff of 40 Ry. In some cases, it get's converged already for 25 Ry (for example C, N and O), but in order to minimize computational cost without loosing accuracy, I will choose a plane wave cutoff of **30 Ry**.

Chapter 4

Results of the Molecular dynamics simulations

In this chapter the results and the analysis of the simulations is presented. The Analysis will be carried out with "on the fly" data, because there is not almost any averaged data for any interesting parameter, as Energies, ion Diplacements, Gradients, ... this is because it was impossible to finish any simulation correctly (due to computing clusters walltimes and queuing systems). But anyway, most of available interesting data is presented in the present chapter.

4.1 Molecular Dynamics of the aminoacids interacting with Al(III)

In this section the results of the performed CPMD simulations are presented. The simulation time is shorted as "s.t.". Graphics are generated with Gnuplot [11] and molecular structures with xmakemol [12]. [1]

4.1.1 Phenylalanine (Phe)

Phenylalanine + 1 H_2O: `s.t.`$\rightarrow$ `6.5 ps`

In this first simulation we can see how during the simulation the Al(III) looses the bent conformation centered at aluminium's β conformation and starts to open. As we can see in Fig. 4.1 [2] the molecule does not suffer big energy changes except for a slightly lower level for the straight conformation. The energy fluctuates around a $\pm$ 0.02 a.u. interval, so all the structures are very similar energetically.

[1] A proper simulation should be over 20 ps long (usually something between 20 and 100 ps + many picoseconds at the beginning for stabilization of the system), but due to the very limited computational resources some simulations are much shorter. This means that the simulations might not show all of the possible structures, but at least all structures present are correct.

[2] The position of the figures is just "more or less" around the time the are happening, because sometimes the don't fit otherwise.

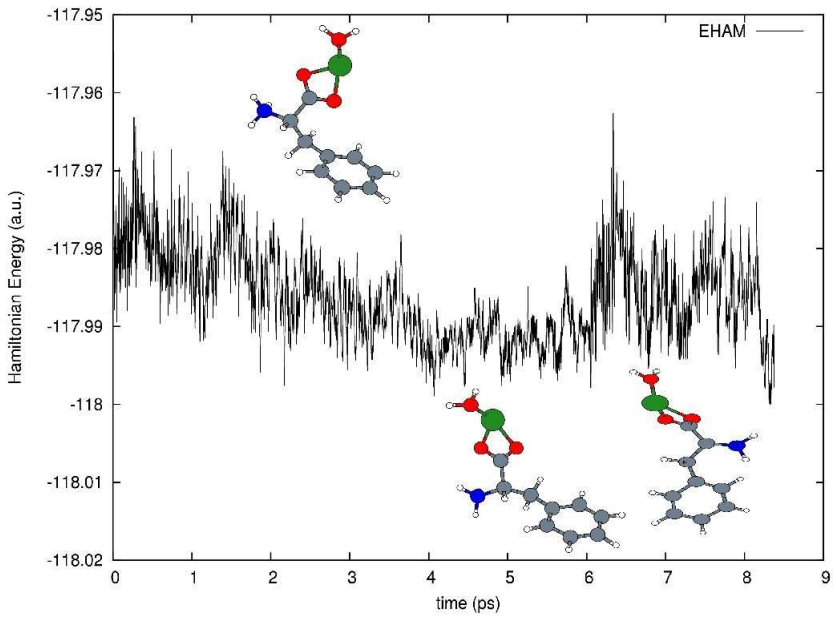


Figure 4.1: Phenylalanine + 1 H_2O: The structure tends to open.

Phenylalanine + 2 H_2O: `s.t.→ 21.6 ps`

For the case of two structural waters, the predicted minima were the ones corresponding to the initial conformation, and another one corresponding to an structure where the Al^{+3} is bonded to the nitrogen.

After the simulation we find how one of the phenylalanine oxygens bonded to the Al^{+3} breaks the bond and get's a proton from the structural waters next to it (Fig. 4.2) lowing this way the energy meaningfully (almost 0.5 a.u.). In a second phase, the aluminium moves to the top of different carbons around the ring, prefferently to one of the *ortho* ones.

Phenylalanine + 3 H_2O: `s.t.→ 5.9 ps`

In this simulation there's first the breaking of the O-Al-O triangle and then, the system get's an straight conformation. The lower minimum is reached after both O-Al-O triangle breaking and straighting are already achieved. Energy difference between open and close conformation is around 0.02 a.u.

Phenylalanine + 4 H_2O: `s.t.→ 11.65 ps`

Even this simulation is relatively long, it seems it is still too short for noticing any phenomena. It seems that the system just got stabilized in a stable minimum. In this case with four "structural" water molecules around the aluminium, the previously predicted structure [2] with one of the waters in the second solvation shell, keeps oscillating without any mentionable change. We

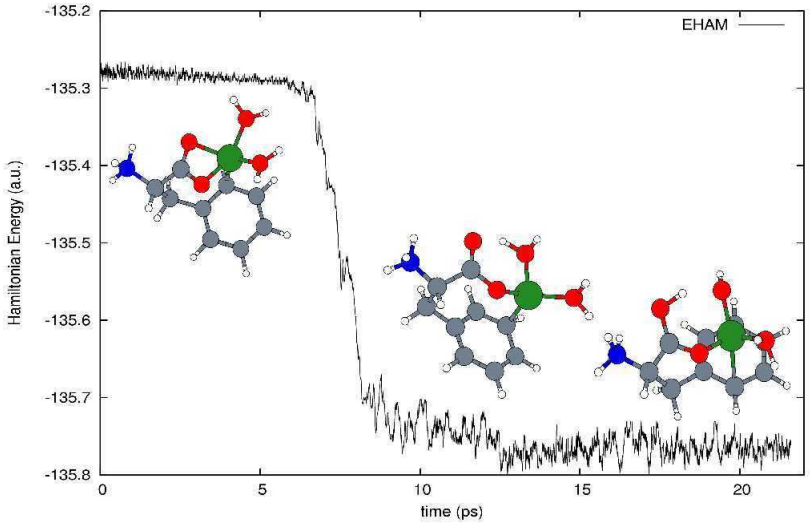


Figure 4.2: Phenylalanine + 2 H_2O: The three main structures around the 21 ps simulation. One of the bridge oxygens of the phenilalanin tends to brake the bond with the aluminium where it reaches another energy minimum. Finally there is also a proton transfer from structural water to bridge oxygen.

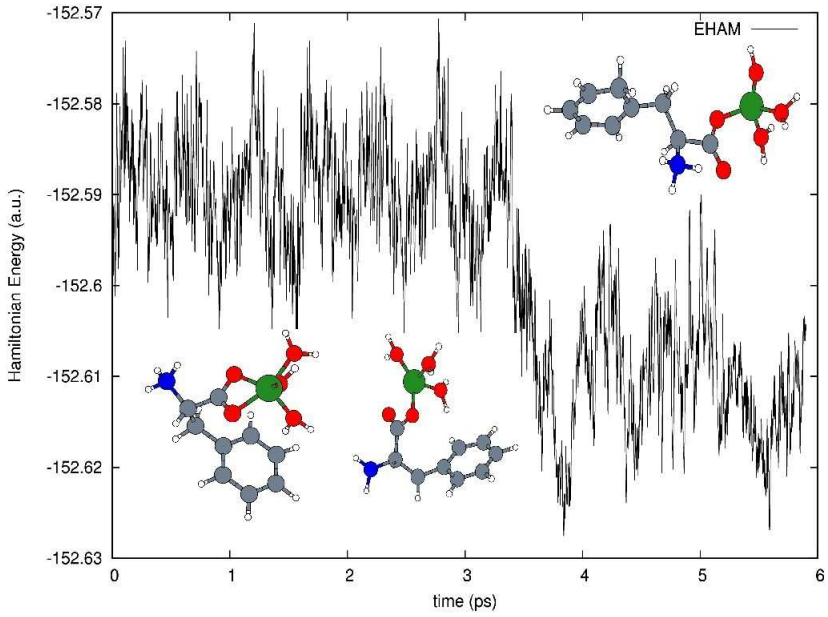


Figure 4.3: Phenylalanine + 3 H_2O: first the O-Al-O triangle is broken and then the system goes for an straight conformation.

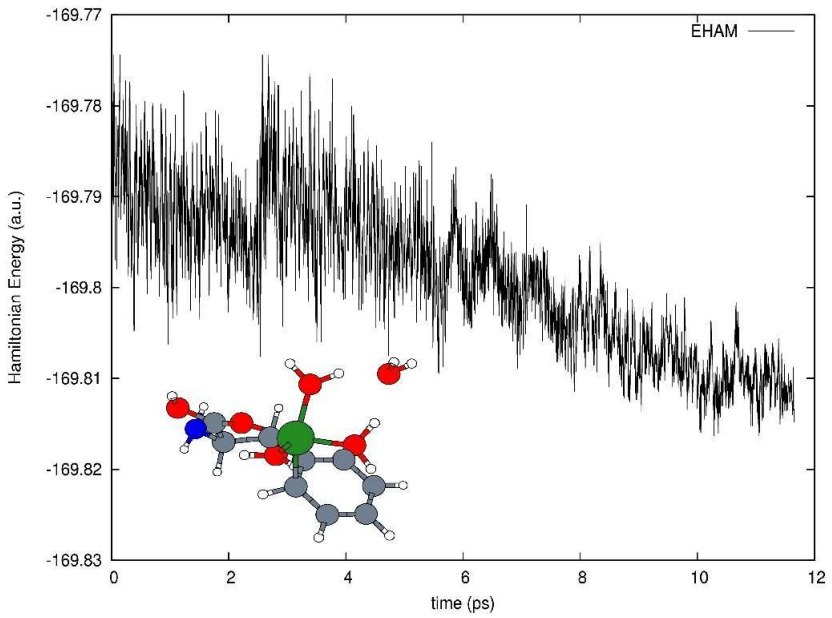


Figure 4.4: Phenylalanine + 4 H_2O: the system keeps it's original structure.

can see a little abnormal behavior of the energy at around 2.2 ps at fig. 4.4, but it is just one of the (usual) restarting points of the calculation.

4.1.2 Tryptophan (Trp)

Tryptophan + 3 H_2O: `s.t.→ 14.7 ps`

The system get's stability when the Al(III) starts moving along the pentagonal ring from the top of a carbon to the top of the next one, until the Al(III) breaks the bond with nitrogen and leaves the zone of the pentagonal ring, then starts to move on the top of the carbons on the hexagon, where the energy minimum is located. In fig. 4.5 can be appreciated how the energy drops suddenly (up to 0.7 a.u.) after the Al(III) starts to move.

Tryptophan + 4 H_2O: `s.t.→`

In this simulation, the system just keeps in a similar conformation to the initial one. As seen in fig. 4.6 there are not meaningful energy changes either.

4.1.3 Tyrosine (Tyr)

Tyrosine + 1 H_2O: `s.t.→ 2.13 ps`

This simulation is really short, but at least we can see that the molecule oscillates from the curved to the straight conformation, The simulation is too short, but most likely it would behave in a similar way as the case of "Phenylalanine + 1 H_2O".

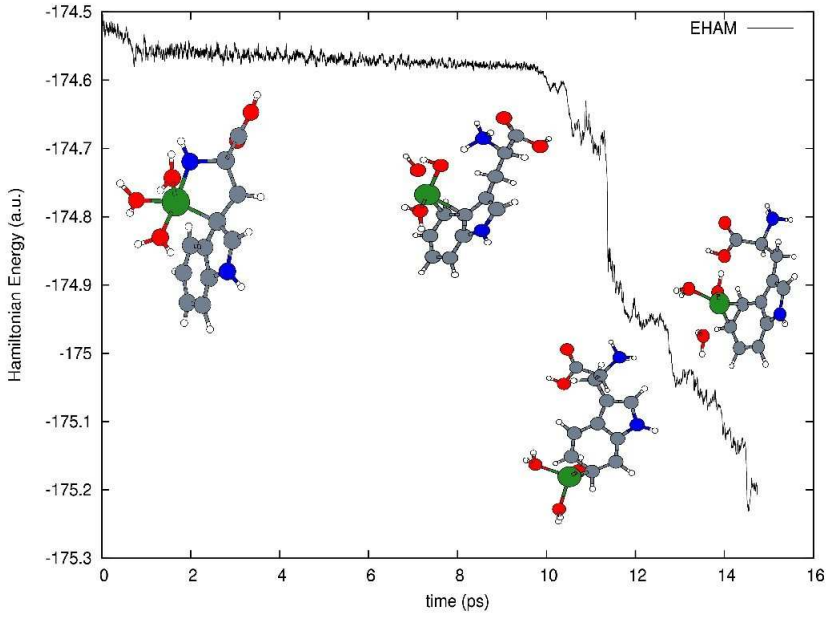


Figure 4.5: Tryptophan + 3 H_2O: The energy drops suddenly when N-Al bond is broken.

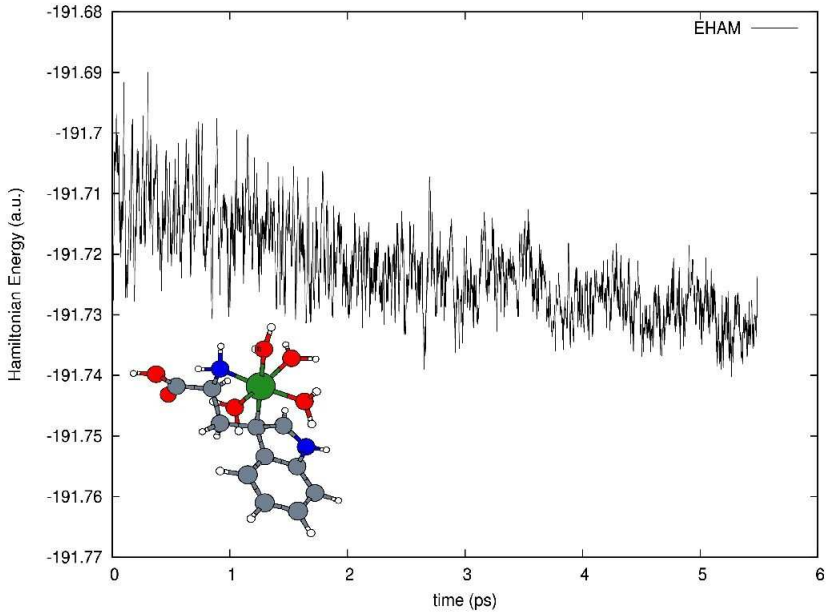


Figure 4.6: Tryptophan + 4 H_2O: keeps the initial conformation.

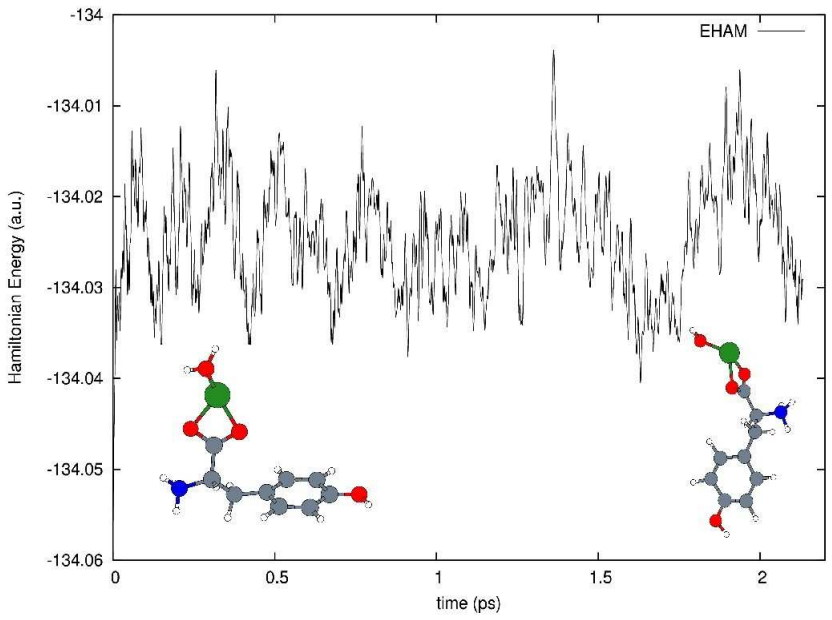


Figure 4.7: Tyrosine + 1 H_2O: System goes from curved to straight

Tyrosine + 2 H_2O: `s.t.→ 4.32 p.s`

The O-Al-O triangle starts to open after 800 steps (0.057 ps). In Fig. 4.8 we can see that the simulation is quite short so there is not enough time for the molecule to move, so this is why the Hamiltonian Energy is pretty constant and there is not increasing or decreasing due to structural modification.

Tyrosine + 3 H_2O: `s.t.→ 3.36 ps`

In this system, as the simulation shows, the system changes the conformation to lower energy minima after only 1500 steps (or around 0.5 ps) (Fig. 4.9). The more stable conformation is the one where the molecule is straight and the Al(III) is around the same plane as the hexagon. After this conformational change, the molecule continues with a similar shape oscillating around this straight conformation.

Tyrosine + 4 H_2O: `s.t.→ 8.39 ps`

In this case, as in the case of the Phenylalanine + 4 H_2O, the second solvation layer will be described as well. In this case there will be 2 water in the second solvation layer, but the system will not suffer meaningful changes during the simulation time, just the normal oscillations.

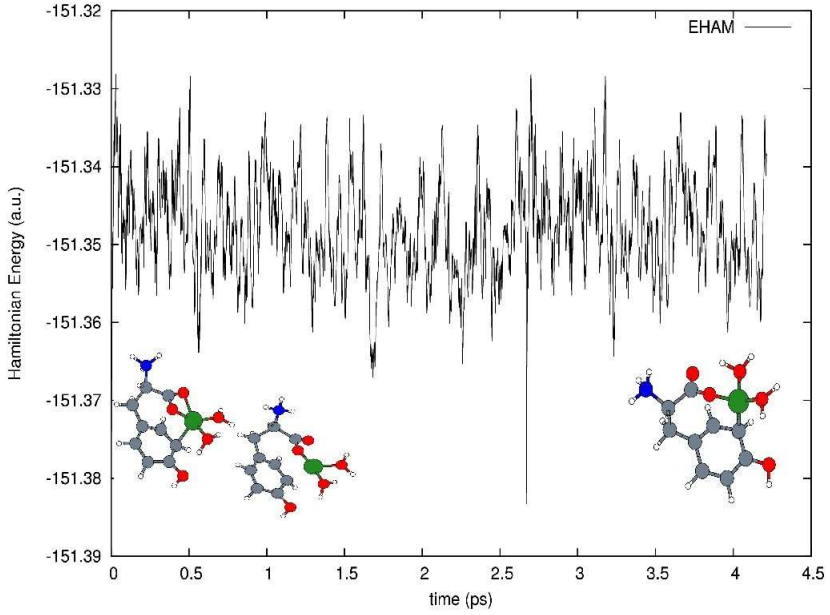


Figure 4.8: Tyrosine + 2 H_2O: The O-Al-O triangle opens, but it does not cause a big energetic change in the system

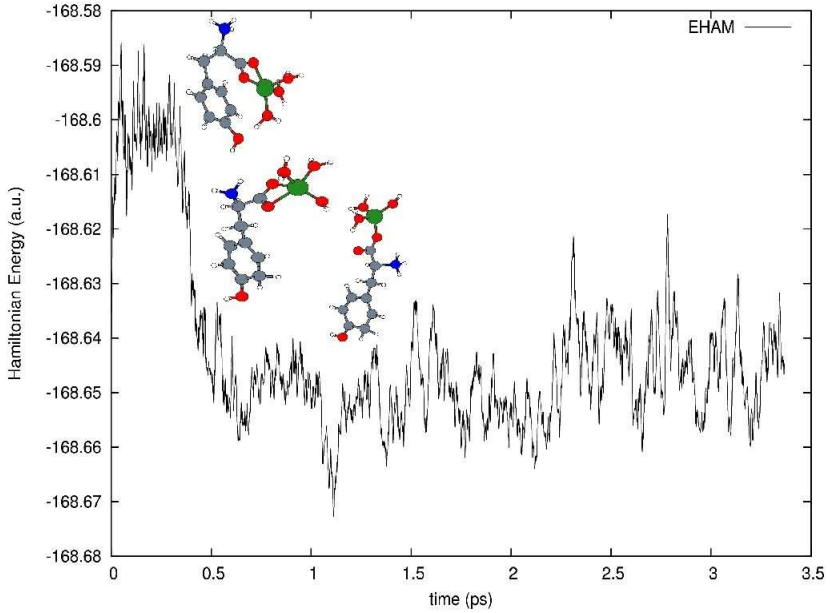


Figure 4.9: Tyrosine + 3 H_2O:The system opens very rapidly after 0.5 ps.

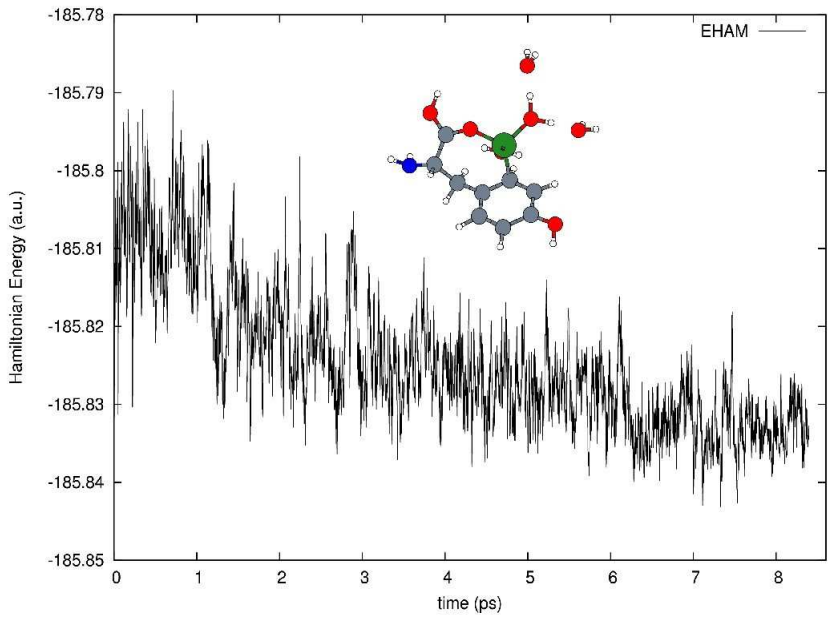


Figure 4.10: Tyrosine + 4 H_2O: keeps initial structure.

4.1.4 Discussion

RDF comparison

First of all I will compare the RDFs between different combinations inside the same aminoacid so that possible influence of the system on the bond lengths could be seen. In the graphics $g(r)$ is the radial distribution function (RDF) and the R is the distance in Å. The graphics are not very accurate, because the simulations are not enough long for produce a proper statistical average. RDFs where generated with VMD [13] and plotted by gnuplot [11].

The first comparison is about aluminium with respect to oxygen. As we can see in Fig. 4.11 (left column) for all the graphics the curves show a peak around 2 Å, this means that the main distance between Al and O will be 2 Å. The graphic on the top shows the case phenylalanine and it describes 4 peaks at the same length, this is how it should be. In the graphics in the middle and bottom, we have more and much more deviation of the peaks from the top of each other. In the case of tryptophan, this can can be because in the tryptophan the the aluminium is not directly bonded to the tryptophan's oxygens and the first peak in the RDFs belong to the distance with the waters oxygens. But again, the deviation is also affected for short length of the simulations.

In the figure showing the O-C RDFs (Fig. 4.11 right column) there are 3 peaks, first one at around 1.4 Å is the one that represents the carbon with the two oxygens bonded also to aluminium. It is a sharp peak because the atoms don't oscillate very much due to the strong covalent bond. The second peak at around 2.5 Å belongs to the same oxygens with the carbon next to the last one,

and finally there is all the "noise" of the water oxygens with previous carbons and interactions with the carbons from the rest of the aminoacid.

About the O-H RDF in Fig. 4.12 (left column), there is just a peak at around 1.2 Å, and obviously its the one belonging to the waters O-H covalent bond, and this is also so sharped (notice that x scale is wider). The second little peak at around 1.4 Å in the case of Tyrosine and Tryptophan, can be because of the OH radical.

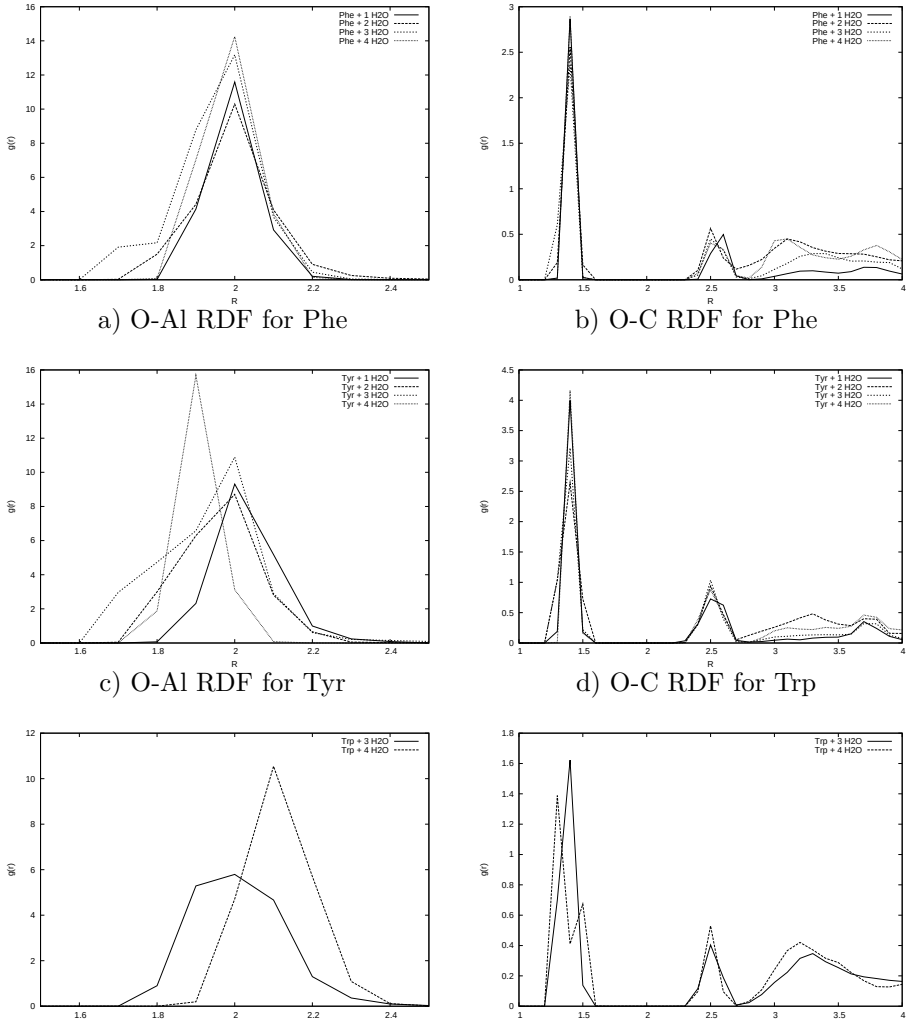
Finally, the Fig. 4.12 shows the O-H RDFs for the different systems (right column). There is the first peak at around 1.2 Å for the covalent C-H bond. Inside the aminoacid, the C-H is minimum in the *orto*, medium in *meta* and maximum (up to 1.3 Å) in *para* positions of the hexagonal ring. Second peak at around 2.3 Å belongs to the distance between hydrogens and the carbon that is bonded to the one the hydrogen is bonded to. The rest of the peaks it is just difficult to say, they are just distances between the rest of different C-H combinations.

4.1.5 Conclusions

The structures open for the cases of one and three waters for the case of phenylalanine and tyrosine. In the case of tryptophan and three waters, in the shown case (aluminium bonded to nitrogen), the aluminium displaces with its bonded water molecules around the rings.

In the cases with 2 waters (phenylalanine and tyrosine), the structure does not open, but the O-Al-O triangle does, causing in the case of phenylalanine a bigger stabilization.

The case of 4 waters, none of the 3 simulations says nothing important. The only one where the four waters are bonded to the aluminium is the tryptophan one, in the case of phenylalanine one of the four waters is at the second solvation shell and at tyrosine there are two waters in the second solvation shell. But in any of the cases it is possible to see any conformation change. In Fig. 4.13 we can see that the atoms are to very similar distances independently of the aminoacid, except for the Al-O case where the Al-O distance is slightly bigger in Tryptophan than in Tyrosine and slightly shorter for the Phenylalanine.



a) O-Al RDF for Phe

b) O-C RDF for Phe

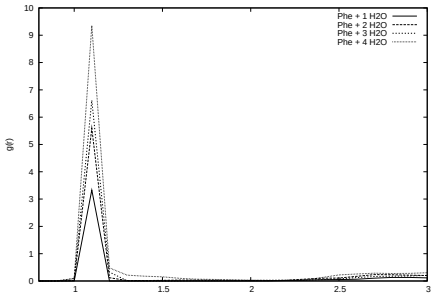
c) O-Al RDF for Tyr

d) O-C RDF for Trp

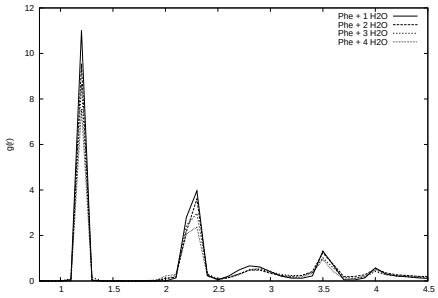
e) O-Al RDF for Trp

e) O-C RDF for Trp

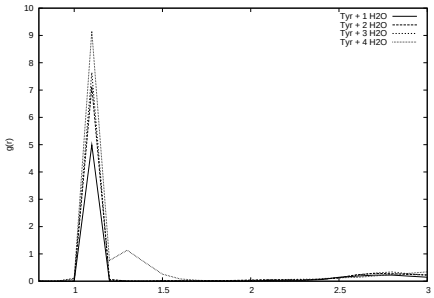
Figure 4.11: RDF comparison for different amount of water molecules on the same aminoacid: O-Al (left column) and O-C (right column).



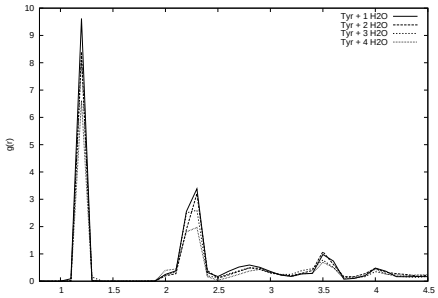
a) O-H RDF for Phe



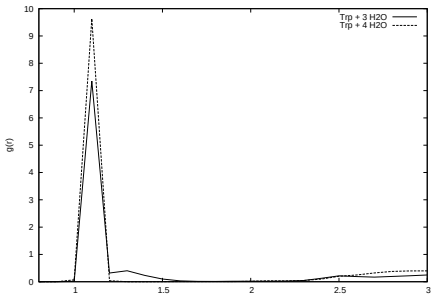
b) C-H RDF for Phe



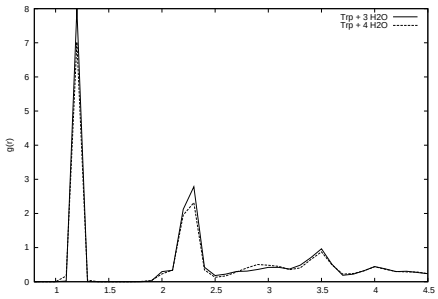
c) O-H RDF for Tyr



d) C-H RDF for Trp



e) O-H RDF for Trp



e) C-H RDF for Trp

Figure 4.12: RDF comparison for different amount of water molecules on the same aminoacid: O-H and C-H.

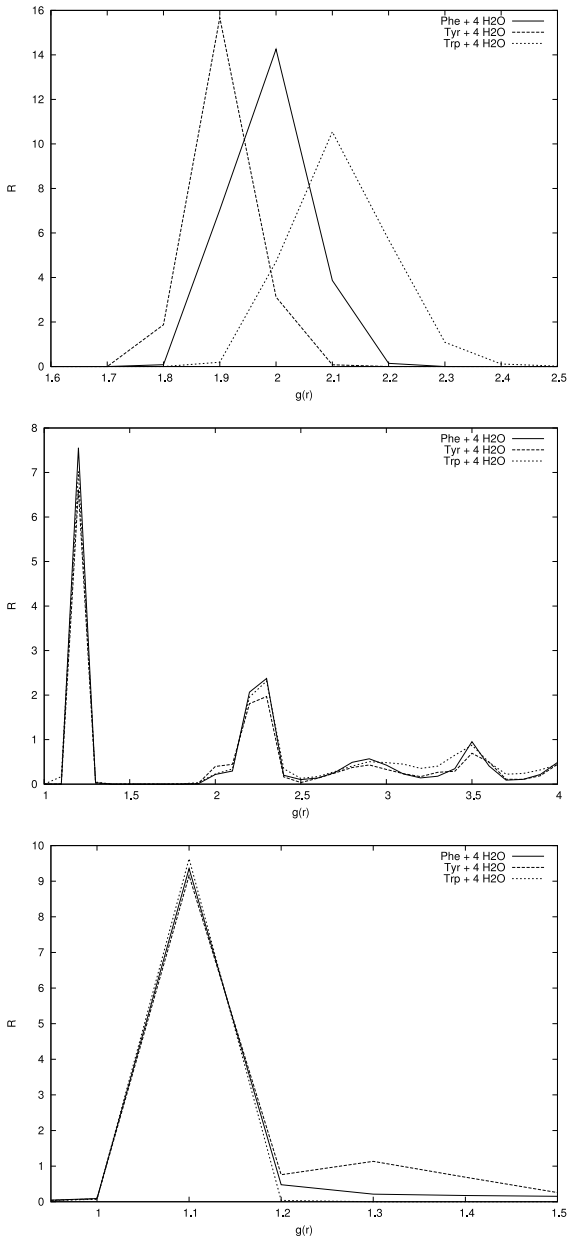


Figure 4.13: Comparison between the same pair RDFs for the three aminoacids for the case of 4 waters. The atom pairs are (from top to bottom): O-Al, C-H and O-H RDFs. First one shows a different behavior depending on the aminoacid, but the other two examples show very similar behaviors.

Chapter 5

Acknowledgements

This work has been done in the Theoretical Chemistry group of the Faculty of Chemistry in Donostia under the supervision of Prof. Jesus M. Ugalde, to whom I should acknowledge first.

The research was carried out by using Euskal Herriko Unibertsitatea, Gipuzkoa Foru Aldundia and Eusko Jaurlaritza's funding.

I have to thank also the SGI/IZO-SGIker UPV/EHU (Supported by the National Program for the Promotion of Human Resources within the national Plan of Scientific Research, Development and Innovation- Fondo Social Europeo MCyT) for the computational resources in cluster Arina, where half of the calculations where done. The other half of the calculations where done at I2BASK to whom I should acknowledge as well for the computing resources, and the efficient technical support of Francisco Javier Ridruejo.

I would like to acknowledge also to Dr. Txema Mercero, Dr. Mario Piris and Dr. Xabi Lopez for the valuable advising in and also the rest of the co-workers Dr. Txoni, Dr. Iñaki, Eli, Eider and Elena.

I have to acknowledge the support given by my Master's thesis director, Prof. Kari Laasonen (University of Oulu), for so many advises given in a totally altruist way.

And finally I want to give an special acknowledge to my parents Encarni and Gabi, for been always there and also to Marika. Eskerrik asko!

Appendix A

Car-Parrinello Molecular Dynamics using CPMD

System's information on the outputs The CPMD [14] produces many different outputs [15]. The most common ones (at least in my case):

- `*.out`: **is the standard output of CPMD. At the beginning prints all the information about the system and calculation parameters and then gives information about the computed task. On molecular dynamics calculations the progressing part is equivalent to** `ENERGIES`

- `RESTART. / RESTART.x`: **binary file with information in order to restart a calculation from the latest point.** `RESTART.x` **is used in the case we want to produce many restarts.**

- `LATEST`: **the file that will say which is the latest** `RESTART.x` **that was saved.**

- `GEOMETRY / GEOMETRY.xyz`: **the geometry of the system in the last saved step, the second one in xyz format.**

- `TRAJECTORY / TRAJEC.xyz`: **binary and ascii files for the trajectories in molecular dynamics simulations.**

- `ENERGIES`: **a file containing data concerning to the various energies of the system at each time step.**

In the `ENERGIES` part the columns mean respectively:
- NFI: Number of step
- EKINC: Fictitious electronic kinetic energy of the electrons (a.u), first term in eq. 2.21

- TEMPP: Temperature the ions, calculated from the kinetic energy of the ions (EKIONS = $5^{th} - 4^{th}$ columns or 2^{nd} term in eq. 2.21)
- EKS: Kohn-Sham Energy. 3^{rd} term in eq. 2.21
- ECLASSIC: Classical Energy (EKS+EKIONS)
- EHAM: Hamiltonian Energy (ECLASSIC+EKINC). Namelly, the total ENERGY
- DIS: mean square displacement of the ions with respect to the initial position. Gives some information on the diffusion
- TCPU. Central Prosessing Unit's requiered time for completing the task (in this case the task is one step)

Bibliography

[1] Rezabal, E.; Mercero, J. M.; Lopez, X.; Ugalde, J. M. *Journal of Inorganic Biochemistry*, **2006**, *100*, 374–384.

[2] Rezabal, E.; Marino, T.; Mercero, J.; Russo, N.; Ugalde, J. M *Inorganic Chemistry*, **2007**, In Press.

[3] Rezabal, E.; Marino, T.; Mercero, J. M.; Russo, N.; Ugalde, J. M. *Journal of Inorganic Biochemistry*, **2007**, In Press.

[4] Schrdinger, E. *Physical Review*, **1926**, *28*, 1049–1079.

[5] (John von Neumann Institute for Computing, Jüulich.). volume 1 of *NIC*. pp 3001–449. 2000.

[6] Car, R.; Parrinello, M. *Physical Review Letters*, **1985**, *55*, 2471–2474.

[7] About the machines, the half of simulations where performed at UPV/EHU's cluster "Arina" with up to 16 Itanium 2 1.6 GHz or 16 Opteron 2.4 GHz / Infiniband network for each simulation. The other half of the simulations where carried out at I2BASK's computational infrastructure with Intel (R) Xeon (TM) 2.4 GHz / Gigabit, shared in 3 different clusters located at the three main Campuses of the UPV/EHU. Optimum parallel performance was achieved by using OPENMP (www.openmp.org) inside the multiprocessor motherboard nodes and MPICH (www-unix.mcs.anl.gov/mpi/mpich/) message passing interfaces for internodal communications trough LAN network.

[8] Vanderbilt, D. *Physical Review B*, **1990**, *41*, 7892–7895.

[9] Laasonen, K.; Car, R.; Lee, C.; Vanderbilt, D. *Physical Review B*, **1991**, *43*, 6796–6799.

[10] Laasonen, K.; Pasquarello, A.; Car, R.; Lee, C.; Vanderbilt, D. *Physical Review B*, **1993**, *47*, 10142–10153.

[11] GnuPlot: http://www.gnuplot.info/

[12] xmakemol: Matthew P. Hodges,
http://www.nongnu.org/xmakemol/

[13] Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics, 1996.

[14] CPMD: Copyright IBM Corp 1990-2006, Copyright MPI fr Festkrperforschung Stuttgart 1997-2001.

[15] CPMD manual: Copyright IBM Corp 1990-2001, Copyright MPI für Festkörpeforschung Stuttgart 1997-2005.