Tutores

Dra. Carme Rovira Virgili
Departament de Química Inorgànica i Orgànica

Dra. Alba Nin-Hill
Departament de Química Inorgànica i Orgànica

Grau de Química

Treball Final de Grau

Computational Study of the Conformational Free Energy Landscape of β-D-Glucopyranose

Estudi Computacional del Mapa Conformacional d'Energia Lliure de la β-D-Glucopiranosa

Adrià Olives Salmerón

June 2021

Universitat de Barcelona

B:KC Barcelona Knowledge Campus
Campus d'Excel·lència Internacional

If you think you understand quantum mechanics,
you don't understand quantum mechanics.

Richard P. Feynman

M'agradaria expressar el més profund agraïment cap a les meves tutores, les doctores Carme Rovira i Alba Nin-Hill. En primer lloc, vull donar-te les gràcies, Carme, per ajudar-me a encaminar aquest treball i adaptar-lo al meu gust des del primer dia que vam reunir-nos, per confiar en mi des del principi i per oferir-me un lloc al teu grup de recerca. En segon lloc, vull donar-te les gràcies a tu Alba, per tota la paciència que has tingut amb mi i amb els meus dubtes, per la dedicació a explicar-me les coses i a fer les reunions i per ensenyar-me aquesta part de la química computacional que pràcticament desconeixia i que és meravellosa.

M'agradaria també donar les gràcies als meus amics, per ajudar-me quan ho he necessitat.

Finalment, m'agradaria agrair als meus pares, Esther i Jaume, a la meva germana, Mireia, i a la meva parella, Ari, pel vostre suport durant aquests quatre anys i per confiar tant en mi.

REPORT

CONTENTS

1. SUMMARY

A computational study of the possible β-D-glucopyranose conformations and energies has been performed using Density Functional Theory (DFT). Three different density functionals have been used; PBE, BLYP and HCTH. The calculations have been performed using the metadynamics method combined with Car-Parrinello Molecular Dynamics (CPMD). The results obtained show that the most stable conformation of β-D-glucopyranose corresponds to a chair conformation, $^4C_1$, with $B_{3,O}$ as the second most stable conformation. The most reliable results are obtained using the PBE functional.

Keywords: β-D-glucopyranose, Density Functional Theory, sugar conformations, metadynamics, Car-Parrinello Molecular Dynamics.

2. RESUM

S'ha realitzat l'estudi computacional de les possibles conformacions i energies de la β-D-glucopiranosa mitjançant la Teoria del Funcional de la Densitat (DFT). S'han usat tres funcionals diferents per a fer els càlculs; PBE, BLYP i HCTH. Els càlculs s'han dut a terme mitjançant el mètode de metadinàmica combinat amb la dinàmica molecular de Car-Parrinello (CPMD). Els resultats obtinguts mostren que la conformació més estable per a la β-D-glucopiranosa es correspon a la conformació cadira, $^4C_1$, sent la conformació $B_{3,O}$ la segona més estable. Els resultats més fiables s'obtenen utilitzant el funcional de PBE.

Paraules clau: β-D-glucopiranosa, Teoria del Funcional de la Densitat, conformacions dels sucres, metadinàmica, dinàmica molecular de Car-Parrinello.

3. INTRODUCTION

A general introduction of β-D-glucopyranose, sugar ring conformations and the carbohydrate-active enzymes glycoside hydrolases is done in the following sections.

3.1. β-D-GLUCOPYRANOSE

Carbohydrates, also called sugars or saccharides, are biomolecules composed of carbon (C), hydrogen (H) and oxygen (O) and sometimes also contain nitrogen (N), phosphorous (P) or sulfur (S). They form polyhydroxy aldehydes/ketones, and if the number of C is five or more, they tend to cyclize. These biological molecules are the most abundant on Earth and they play a very important role in life. [1] Carbohydrates can be classified in function of the monomer units as:

- monosaccharide (single polyhydroxy aldehyde/ketone)
- oligosaccharides (short monosaccharide chains bonded by glycosidic bonds)
- polysaccharides (more than 20 monosaccharides)

Probably the most important carbohydrate is the monosaccharide D-glucose. The empirical formula of this sugar is $C_6H_{12}O_6$. D-glucose exists mainly in a cyclic form, also called ring form. When D-glucose gets cycled can appear two anomers of glucose, as can be seen in **Figure 1**, depending on the relative positions between hydroxyl group on the carbon 1 (also called anomeric carbon) and CH_2OH group on the carbon 5.

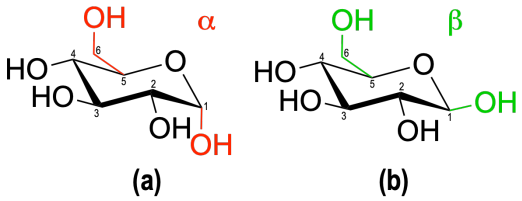

Figure 1. Possible D-glucose anomers: (a) α-D-glucopyranose (b) β-D-glucopyranose.

In this project we will focus on the β-D-glucopyranose molecule. Hereafter, its simplified name, β-glucose, will be used.

3.2. SUGAR RING CONFORMATIONS

It has just been seen that β-glucose is a six-membered ring sugar, but what shape this ring adopts? Is this molecule always like the one depicted in **Figure 1b**? Absolutely not.

There are many possible ways in which β-glucose organizes in space. These different structures of the same molecule are called conformations and each of them can be interconverted into another if it has enough energy (without breaking any bond). These sugar ring conformations were first classified into families using the Schwartz nomenclature[2] (resumed in **Figure 2**) and later this classification was accepted and approved by the International Union of Pure and Applied Chemistry (IUPAC). For pyranoses, 38 main conformations are found, called canonical conformations.

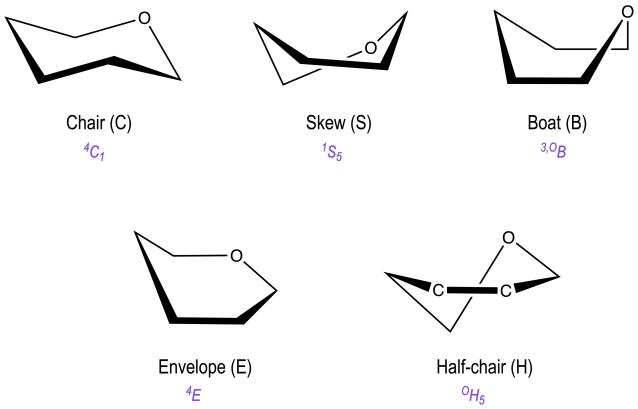

Figure 2. Family of pyranose ring conformations. Purple labels refer to the specific conformation within a given type (chair, skew, boat, envelope or half-chair).

Stoddart proposed a schematic[3] diagrams to interconnect all conformations. These are called Stoddart's diagrams (**Figure 3**), and they are very useful to understand the interconversion, i.e., all the possible conformational pathways.[4] Nevertheless, Stoddart's diagram gives no information about the relative energies of each canonical conformation.

One of the interests of this project is to complement Stoddart's diagram with energy values to understand the amount of energy involved in these conformational pathways and the relative differences among conformations.

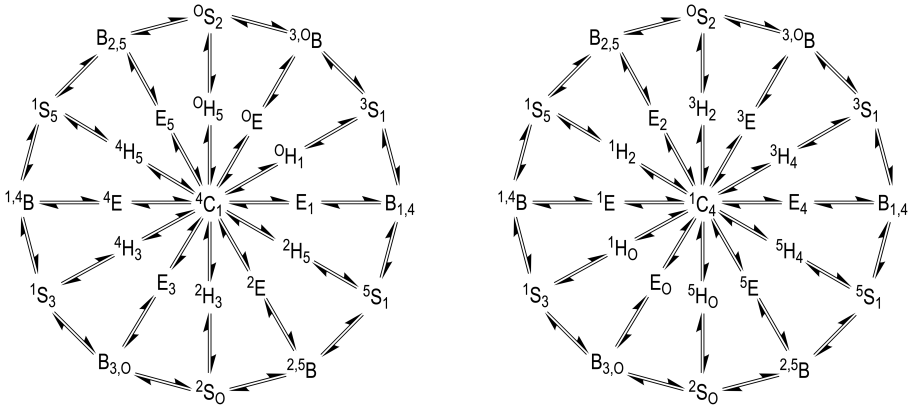

Figure 3. Stoddart's diagram.

3.3. GLYCOSIDE HYDROLASES

Glycoside hydrolases or glycosidases are enzymes that catalyze the hydrolysis of the glycosidic bonds, i.e., depolymerize polysaccharides into smaller sugar chains. These enzymes are very important in biology, medicine and in many other fields of research[5], so it is important to understand the mechanistic action of glycoside hydrolases and study the interaction between them and their corresponding sugar substrates such as a glucose polysaccharide (**Figure 4**).

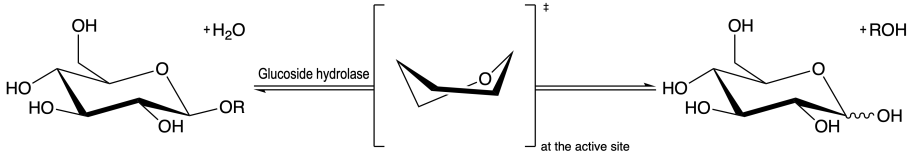

Figure 4. Hydrolysis of a glycoside. Note that the reactive sugar adopts a skew conformation at the Transition State (TS) of the reaction.

Although glycoside hydrolases are not analyzed in this project, there is a clearly close relationship between the activity of these enzymes and the conformations that the reactive sugar adopts during catalysis. In fact, glycosidases have been often seen to recognize distorted conformations in its active site.[6] Knowing the intrinsic conformational properties of glucose helps to predict the catalytic conformational pathway in the active site of the glycoside hydrolase. Some authors have demonstrated that the small structural and electronic changes observed upon distortion seem to be an intrinsic property of the substrate and the enzyme has

probably evolved to use these properties for a more efficient catalysis. Therefore, understanding the catalytic conformational itineraries is of great importance when designing selective inhibitors (for possible malfunctioning enzymes) as the usually more powerful inhibitors are the ones mimicking the properties of the TS.[29]

4. OBJECTIVES

The first goal of this project is to use the metadynamics (MTD) technique combined with Car-Parrinello Molecular Dynamics (CPMD), a method based on Density Functional Theory (DFT), to study the conformational Free Energy Landscape (FEL) of the isolated β-glucose in the gas phase.

The second goal of this project is to compare the results obtained with three different density functionals (PBE, BLYP and HCTH) in order to determine which of them is more suitable for sugars systems.

The third goal is to compare the results obtained in this project with the results obtained in previous works.

As a summary of the objectives:

1. build the FEL of the β-glucose using CPMD-based MTD.
2. compare between PBE, BLYP and HCTH functionals used in DFT calculations.
3. compare the results obtained in this project with previous works.

5. METHODS

In these sections, the basic concepts to understand the calculations of this project and the simulation parameters are presented. Broadly speaking, a brief introduction to *Ab Initio* Molecular Dynamics (AIMD) is given, focusing on Car-Parrinello Molecular Dynamics (CPMD) and the uses of Density Functional Theory (DFT). Next, a technique that enhances the sampling of molecular dynamics, called metadynamics (MTD), is discussed. Then, the coordinates necessary to describe each conformation of the β-glucose ring (explained in **3.2**) called sugar puckering coordinates are presented and plotted geometrically. Finally, the computational details and some simulation parameters are described.

5.1. *AB INITIO* MOLECULAR DYNAMICS

Ab Initio Molecular Dynamics (AIMD) (*ab initio* means "from first principles") combine classical Molecular Dynamics (classical MD) with *ab initio* electronic structure, i.e., Newton's equation of motion and Schrödinger equation (or some approximate method in order to obtain the electronic energy) must be solved simultaneously (**Figure 5**). From another point of view, it is like a classical MD where electronic energy is calculated at each simulation step.

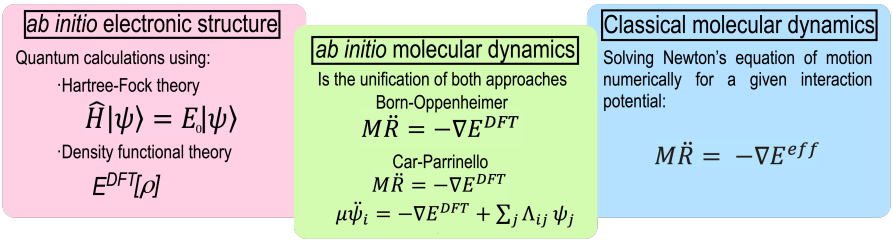

Figure 5. AIMD combine *ab initio* electronic structure and classical MD.[7]

The difference between AIMD and classical MD is that the potential energy due to electronic structure in AIMD replaces the analytical functions used in classical MD, called force field, in order to approximate the interatomic energy, i.e., the interatomic energy is obtained via a quantum treatment of electrons in AIMD, while classical MD uses some approximate functions associated with bonds, angles, dihedrals, inversions, electrostatic potential and Van der Waals potential. Furthermore, AIMD is more accurate than classical MD but spends much more time.[11]

AIMD makes use of the following approximations[14]:

- Born-Oppenheimer approximation; it allows to separate the treatment of electrons dynamics of nuclei dynamics.
- nuclei are treated classically.

The first AIMD methods were Ehrenfest Molecular Dynamics (EMD) and Born-Oppenheimer Molecular Dynamics (BOMD). As explained later, both methods have drawbacks and in 1985, Roberto Car and Michele Parrinello developed a new procedure to enhance AIMD; Car-Parrinello Molecular Dynamics.[12] Car-Parrinello method or Car-Parrinello Molecular Dynamics (CPMD) is an AIMD based on the Density Functional Theory (DFT).

An overview of DFT, a general explanation of BOMD and EMD and the improvement by CPMD are presented below.

5.1.1. Density Functional Theory

DFT allows approximating the ground state energy, electronic distribution and many other molecular properties,[9] working with an electronic density functional instead of the electronic wavefunction, which avoids having to solve the Schrödinger equation. It is necessary to know atomic coordinates and minimize the energy functional with respect to the electronic density.[10] Working with the density functional is easier than working with the wavefunction because it only needs 3 coordinates, while a wavefunction that describes N electron system needs 3N coordinates.[8], [9], [10]

A general DFT energy equation is expressed in equation (1), where $T_S[\rho]$ is the kinetic energy calculated from the Slater determinant using non-interacting electrons, the second and third terms are potential energy and $E_{xc}[\rho]$ is a correction term (see the following paragraph). Then (1) must be minimized; equation (2).[15]

$$E[\rho] = T_S[\rho] + E_{ne}[\rho] + E_{ee}[\rho] + E_{xc}[\rho] \quad (1)$$

$$E^{DFT} = \min_{\rho(\vec{r})} E\left[\rho(\vec{r}), \vec{R}_N\right] \quad (2)$$

In DFT, the correlation and exchange energy ($E_{xc}[\rho]$) are calculated by using some functionals. There are many kinds of functionals classified in some families, but the main ones are Local Density Approximation (LDA), Generalized Gradient Approximation (GGA) and hybrid functionals. LDA is restricted to the simple situation where orbitals are doubly occupied (nevertheless, there is an extension to the unrestricted case called Local Spin-Density Approximation, LSDA) and have limitations in describing a number of chemical bonds. Because LDA or LSDA approximation cannot be applied in most chemical and biological systems, the better approximation GGA takes into account for functional expression not only the density but also the gradient of the density. Hybrid functionals (mixt of GGA and Hartree-Fock exchange) can also be used and it is especially suited for radical system.[9]

DFT was developed for the time-independent problem, but it was later extended for the time-dependent problem. The corresponding theory is called Time-Dependent Density Functional Theory (TDDFT), which is typically applied to describe certain excited states.[13] However, this project does not use the TDDFT.

5.1.2. Born-Oppenheimer Molecular Dynamics

BOMD can be resumed by the following flow chart:

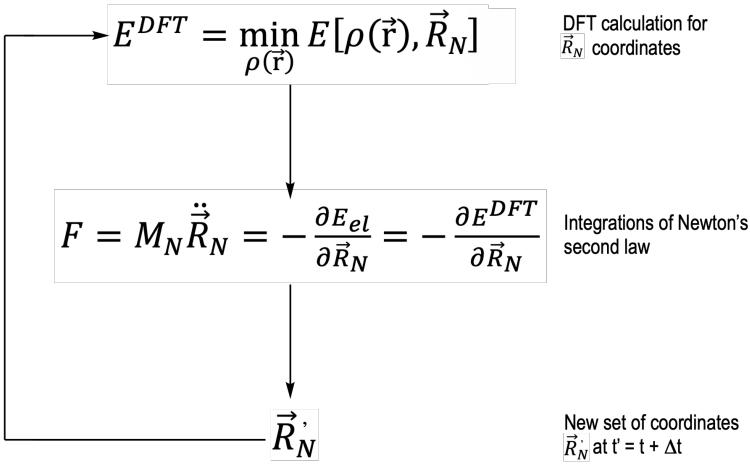

Figure 6. Flow chart of BOMD, an *ab initio* molecular dynamics based on DFT.[9]

Firstly, the electronic energy (E_{el}) can be obtained by quantum calculations like solving the Time-Independent Schrödinger Equation (TISE) or the minimization of density functional using DFT at fixed atomic coordinates (BOMD is a time-independent technique). Secondly, assuming that E_{el} is the interatomic energy, Newton's equation of motion is solved (M_N are the nuclei masses) to obtain a new set of atomic positions ($\vec{R}_N'$) after a time increment or time step (Δt). Finally, it is necessary to repeat this procedure using $\vec{R}_N'$ instead of $\vec{R}_N$.

The equations of motion can be derived by a Lagrangian formalism. The Lagrangian is a scalar function that contains the properties of some physical system, and the expression is the difference between kinetic and potential energy, as can be seen in equation (3).

$$\mathcal{L} = T - V \tag{3}$$

For a given Lagrangian, the solutions of Euler-Lagrange equations give, e.g., the equations of motion of the system we are interested in.[16] The BOMD Lagrangian has the following form[17]:

$$\mathcal{L}_{BOMD} = E_N^{kin} - E^{el} = \sum_N \frac{1}{2} M_N \dot{\vec{R}}_N^{\,2} - E^{DFT} \tag{4}$$

NOTE: remember that Born-Oppenheimer approximation is applied.

5.1.3. Ehrenfest Molecular Dynamics

EMD can be resumed by the following flow chart:

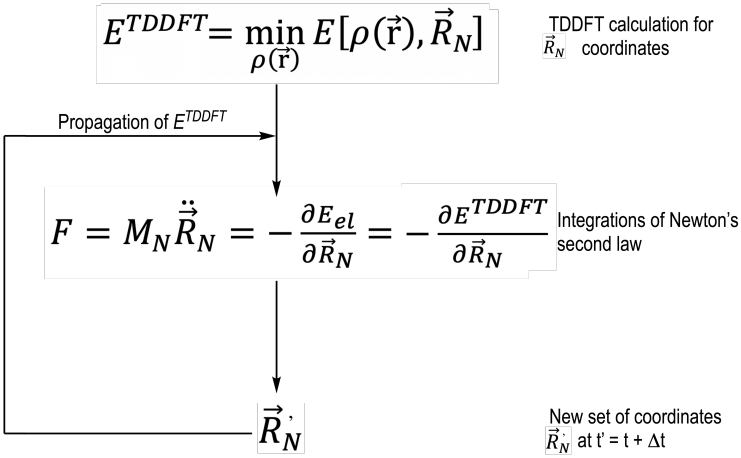

Figure 7. Flow chart of EMD, an *ab initio* molecular dynamics based on TDDFT.[9]

Firstly, E_{el} can be obtained by quantum calculations like solving the Time-Dependent Schrödinger Equation (TDSE) or the minimization of density functional using TDDFT at fixed atomic coordinates (EMD is a time-dependent technique). Secondly, assuming that E_{el} is the interatomic energy, Newton's equation of motion is solved to obtain the new set $\vec{R}_N'$ after a time step Δt. Finally, this last step is repeated but using $\vec{R}_N'$ instead of $\vec{R}_N$ and propagating E_{el} (also called E^{TDDFT}) instead of minimizing again the density functional.

The EMD Lagrangian has the following form[18]:

$$\mathcal{L}_{EMD} = E_N^{kin} - E^{el} = \sum_N \frac{1}{2} M_N \dot{\vec{R}}_N^{\,2} - E^{TDDFT} \tag{5}$$

5.1.4. Car-Parrinello Molecular Dynamics

In EMD, only an optimization of the initial wave function is necessary because, for the rest of the simulation, the movement of the electrons is given by the propagation of this initial wave function by applying the Hamiltonian. In BOMD, there is no electron dynamics because this is a time-independent technique, which means that wave function must be found at each MD step.

In contrast, the EMD time step must be much smaller than BOMD because electronic motion is much faster than nuclei motion, so BOMD allow us to do a larger MD time step.

CPMD mix both methods in order to[7]:

1. profit from the EMD advantage that only one wave function must be optimized (or the equivalence for the density functional) and its subsequent propagation.
2. profit from the BOMD advantage of integrating the equations of motion in a longer MD time step.

Car and Parrinello considered the parameters R_N and ψ_i in the energy functional to be time-dependent and proposed the following Lagrangian[7], [9], [12]:

$$\mathcal{L}_{CPMD} = E_N^{kin} + E_{el}^{kin} - E^{el} + \sum_{ij} \Lambda_{ij}\left(\langle\psi_i|\psi_j\rangle - \delta_{ij}\right) = \tag{6}$$

$$= \sum_N \frac{1}{2} M_N \dot{\vec{R}}_N^{\,2} + \sum_i \frac{1}{2}\mu\langle\dot{\psi}_i|\dot{\psi}_i\rangle - E^{KS} + \sum_{ij} \Lambda_{ij}\left(\langle\psi_i|\psi_j\rangle - \delta_{ij}\right)$$

If we take a look at BOMD and EMD Lagrangians, equations (4) and (5) respectively, the CPMD Lagrangian has similar terms like the kinetic energy of nuclei and potential energy of electrons (in this case called as E^{KS}) but incorporate some new terms, as can be seen in equation (6). The last terms are the Lagrangian multipliers, and they are introduced in order to satisfy the orthonormal condition of orbitals. However, the most important difference is the second term, corresponding to fictitious kinetic energy associated with electronic functions, where ψ_n are the Kohn-Sham (KS) orbitals and μ is the fictitious electronic mass. This fictitious kinetic energy does not have a real physical meaning like nuclei kinetic energy term, but it is an excellent mathematical tool to perform the MD because using it, the simulation allows electrons and nuclei to evolve simultaneously. The small electronic mass and the corresponding small fictitious energy ensures the adiabatic behavior of the nuclei-electrons system, i.e., there hardly any interaction between them, avoiding significant energy transfer from nuclei to electrons that could affect the forces on the atoms. If there were an energy transfer, electrons would be excited, and the simulation would not describe the ground state.[19]

Solving the Euler-Lagrange equations with CPMD Lagrangian, the following equations of motion are obtained[14]:

$$M_N \ddot{\vec{R}}_N = -\frac{\partial E^{KS}}{\partial \vec{R}_N} \tag{7}$$

$$\mu\ddot{\psi}_i = -\frac{\partial E^{KS}}{\partial \psi_i^*} + \sum_j \Lambda_{ij}\,\psi_j \tag{8}$$

NOTE: if E^{KS} is minimized with respect to single-electron orbitals ψ_i, E^{DFT} is obtained. CPMD uses DFT, so in the following equations, E^{DFT} is used instead of E^{KS}.

CPMD can be resumed by the following flow chart:

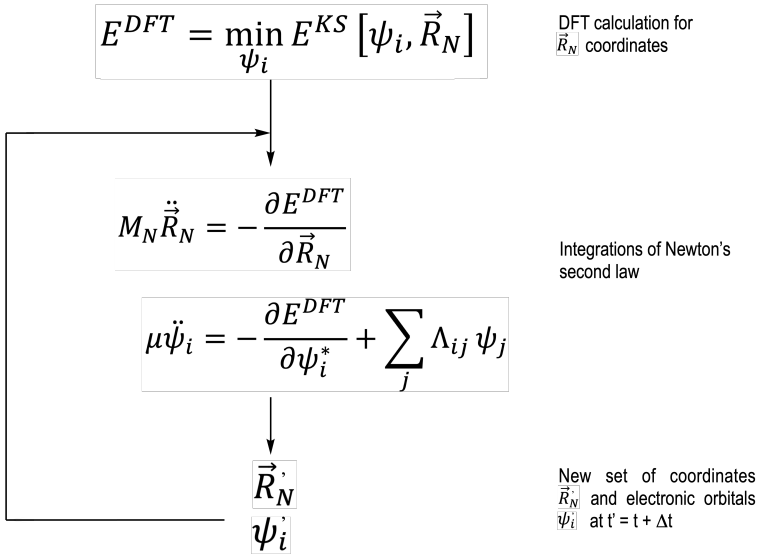

Figure 8. Flow chart of CPMD, an improved *ab initio* molecular dynamics over BOMD and EMD and based on DFT.[8]

Firstly, E_{el} is obtained by minimization of the Kohn-Sham equations using DFT at fixed atomic coordinates. Secondly, assuming that E_{el} is the interatomic energy, Newton's equation of motion is solved for nuclei and electrons to obtain the new set $\vec{R}_N'$ and a new set of orbitals (ψ_i') after a time step Δt. Finally, repeat this procedure from solving Newton's second law for nuclei and electrons but using $\vec{R}_N'$ instead of $\vec{R}_N$ and ψ_i' instead of ψ_i.

Now, it can be understood the improvement that CPMD brings with respect to the other AIMDs: CPMD "propagates" wave function throughout the simulation (like EMD) and the electronic energy is calculated by using DFT instead of time-dependent technique in order to increase the time step (more close to the one used in BOMD).

5.2. METADYNAMICS

MTD is an enhanced-sampling MD technique that allows exploring some regions that would not be possible to explore, or it would spend much time, using conventional MD, i.e., it allows to accelerate MD processes and study rare (i.e., short-lived) states like TSs or high energy conformations. In addition, it allows obtaining the Free Energy Landscape (FEL) of our chemical system.[20] The FEL, sometimes called Free Energy Surface (FES) despite not being entirely correct, is the set of free energy values as a function of some coordinates.

Two important aspects of MTD are[8], [19]:

1. dimension reduction.
2. forcing the system to explore not yet visited states.

Dimension reduction means finding some coordinates as a function of system coordinates that can correctly describe the desired process and thus work in a low-dimensional space. These coordinates are called Collective Variables (CVs) and they must be selected in such a way that they differentiate the states of the system that we are interested in (e.g., the conformations). In the next point (**5.3**), it is seen which CVs are chosen in this project.

Forcing the system to explore means forcing the CVs to have other values and thus explore the FEL. The problem is that there are energetic barriers that prevent FEL exploration. So, this procedure is accomplished by adding a biasing potential in the form of small repulsive Gaussian-like potentials functions, called hills, every certain time intervals. When a minimum is completely filled with the additional Gaussians, the molecule can "fall" into a new minimum, where Gaussians are placed again. This is schematically depicted in **Figure 9**.

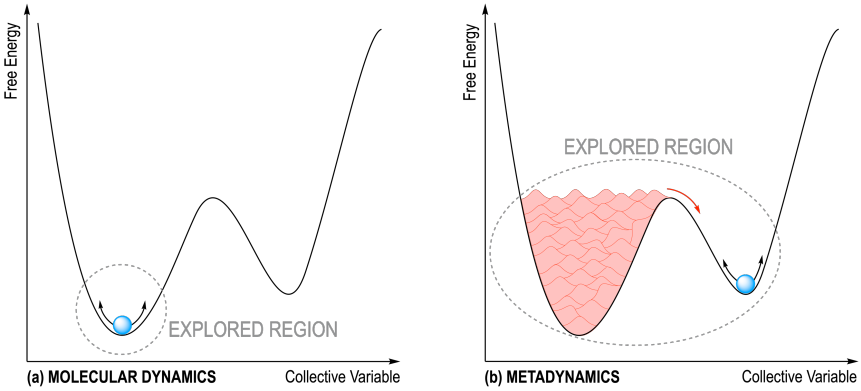

Figure 9. MD (a) explores only one minimum while MTD (b) explores all of them.

It is important to use hills of appropriate size, i.e., the height (w) and the width (σ) (**Figure 10**) of these should be carefully selected because, e.g., if the Gaussian is too high, the FEL will lose resolution and it will probably be wrong but, if the Gaussian height is too low, the simulation time will be very long despite its high resolution.

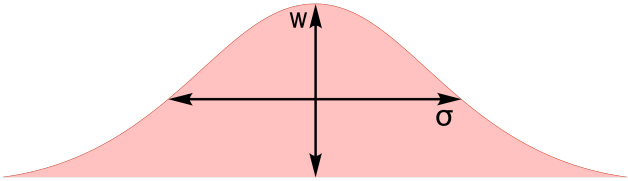

Figure 10. Gaussian potential or hill; w is the height and σ is the width.

This process is repeated until all minima are filled. The FEL is considered converged when the system can change states without passing through an energetic barrier (diffusion regime). When this "scanning" through the FEL is finished, all the Gaussians can be converted to free energy and the FEL can be obtained.

From a mathematical point of view, we are adding an extra potential called bias potential (V_{bias}) to the electronic energy E^{el} obtaining a new total potential, equation (9), i.e., CPMD Lagrangian (6) is changing and therefore equations of motion too, (7) and (8). [19]

$$V_T = V_{bias} + E^{el} \tag{9}$$

The expression of bias potential is shown in equation (10) and has a Gaussian form (s is the CV position where a hill is added).

$$V_{bias}(s,t) = \sum_{t_i} we^{-\frac{\left(s(t)-s(t_i)\right)^2}{2\sigma^2}} \tag{10}$$

Because we are adding hills and therefore modifying the total potential, MTD is not an equilibrium process (it is an out-of-equilibrium dynamics), so it is necessary to wait some time before a new Gaussian is added (deposition time) to help the system to return to the equilibrium. Moreover, electronic thermostats are added to ensure that the velocities distribution (related with the temperature) remains constant at the simulation temperature[8] (see **5.4.2**).

Once the final value of $V_{bias}(s,t)$ is computed, the FEL can be obtained from (11). If the correct values of w and σ have been chosen, the free energy is related to the sum of all hills as:

$$\lim_{t\to\infty} V_{bias}(s,t) \approx -\Delta G(s) \tag{11}$$

5.3. SUGAR PUCKERING COORDINATES

We previously saw in section **3.2** that distinct families of pyranose conformations are interconnected through Stoddart's diagram. Nevertheless, if we take a look at **Figure 3**, there are two different diagrams that coincide by the edges (like two hemispheres), so if we overlap both diagrams and separate the centers of each one away, Cremer and Pople's sphere is obtained (**Figure 11**).

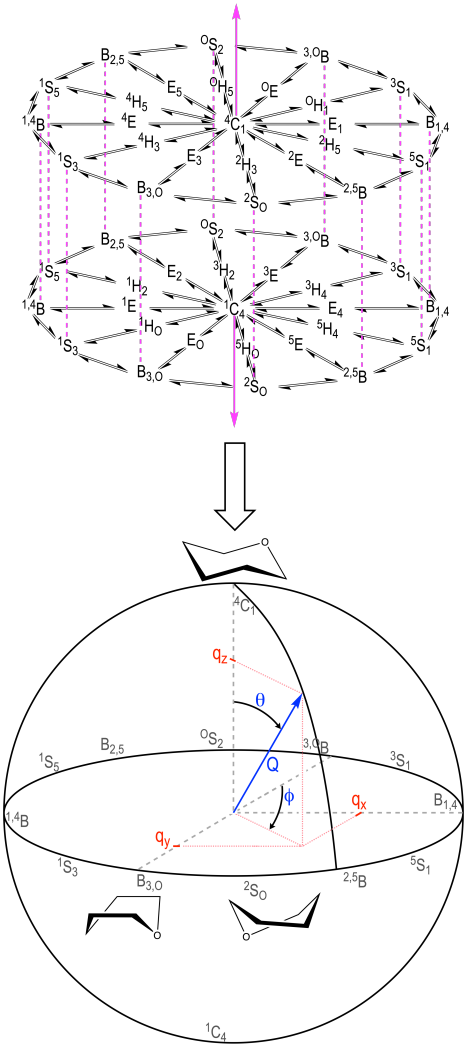

Figure 11. Cremer and Pople's sphere.

Cremer and Pople introduced this concept and the mathematical description in 1975 to describe these sugar conformations quantitatively using only the atomic coordinates of ring atoms.[21] All conformations lie on the surface of a sphere of radius Q and can be described by spherical coordinates Q, θ, and ϕ. So, to each sugar conformation, it can be assigned a given value of (Q, θ, ϕ), that are a function of the coordinates of the atoms that form the sugar ring.

A mathematical description of a pyranose ring is presented below[8], [21]:

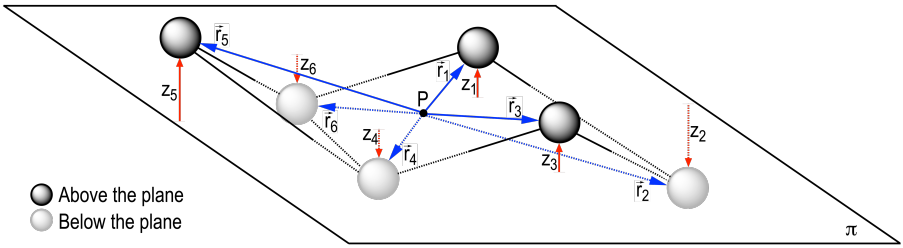

Figure 12. Pyranose ring system.

Firstly, all nuclei of the atoms ring are characterized by Cartesian coordinates X_j, Y_j and Z_j (where $j \in \{1,2,\cdots,6\}$ for pyranoses). Furthermore, position vectors can be defined respect some origin and the most comfortable to work is the geometrical center, point P in **Figure 12**. So, position vectors are r_j and they satisfy the following condition:

$$\sum_{j=1}^{6} \vec{r}_j = 0 \tag{12}$$

Secondly, to define a system of puckering coordinates is better to specify each nuclei displacement with respect to a mean plane, the plane π in **Figure 12**. Plane π passes through the geometrical center and new Cartesian coordinates can be defined as x_j, y_j and z_j, where z axis is perpendicular to π and y axis passes through the nuclei 1 projection over π. Now, a new condition appears,

$$\sum_{j=1}^{6} z_j = 0 \tag{13}$$

but more conditions are needed in order to define and fix the mean plane in the space. These conditions are:

$$\sum_{j=1}^{6} z_j \cos\frac{2\pi(j-1)}{6} = 0 \tag{14}$$

$$\sum_{j=1}^{6} z_j \sin\frac{2\pi(j-1)}{6} = 0 \tag{15}$$

Now, we can express π orientation as a function of $\vec{r}_j$ defining two new vectors (two linearly independent vectors define a plane in space), (16) and (17), and obtaining the unit vector of their cross product (18), which is perpendicular to π.

$$\vec{R}' = \sum_{j=1}^{6} \vec{R}_j \sin\frac{2\pi(j-1)}{6} \tag{16}$$

$$\vec{R}'' = \sum_{j=1}^{6} \vec{R}_j \cos\frac{2\pi(j-1)}{6} \tag{17}$$

$$\hat{n} = \frac{\vec{R}' \times \vec{R}''}{\left|\vec{R}' \times \vec{R}''\right|} \tag{18}$$

Now, we can choose this vector $\hat{n}$ to define z axis. The displacements of nuclei from the π are given by the scalar product of each $\vec{R}_j$ with $\hat{n}$ (19). Equation (19) satisfy conditions (13), (14) and (15).

$$z_j = \vec{R}_j \cdot \hat{n} \tag{19}$$

Finally, as puckering coordinates are a function of z_j, these ones can be obtained by solving the following system of equations:

$$\begin{cases} Q\sin\theta\cos\phi = \dfrac{1}{\sqrt{3}}\displaystyle\sum_{j=1}^{6} z_j \cos\left[\frac{2\pi}{6}2(j-1)\right] \\ Q\sin\theta\sin\phi = \dfrac{1}{\sqrt{3}}\displaystyle\sum_{j=1}^{6} z_j \sin\left[\frac{2\pi}{6}2(j-1)\right] \\ Q\cos\theta = \dfrac{1}{\sqrt{6}}\displaystyle\sum_{j=1}^{6} (-1)^{j-1} z_j \end{cases} \tag{20}$$

Q is the total puckering amplitude and is also defined as the equation (21). The amplitude is greater than 0 and tells us about how much flattened our system is.

$$Q = \left(\sum_{j=1}^{6} {z_j}^2 \right)^{1/2} \tag{21}$$

θ and ϕ are the coordinates that differentiate between possible conformations, and they can be obtained by solving (20).

Furthermore, we can express these coordinates in the Cartesian system:

$$q_x = Q \sin\theta \sin\phi \tag{22}$$

$$q_y = Q \sin\theta \cos\phi \tag{23}$$

$$q_z = Q \cos\theta \tag{24}$$

These last coordinates (q_x, q_y *and* q_z) are the ones that are used in this project as CV. Note that doing this procedure, our initial system of 6·3=18 coordinates are transformed to a 3 coordinates problem, that is easier to apply MTD.

To do metadynamics, the CVs used must fulfill certain conditions[19]:

1. they must be an explicit function of the atomic positions.
2. they must be able to distinguish each relevant state of the system, i.e., all conformations of Stoddart's diagram.
3. they must include the slow movements (low frequency modes) of the system. This means that the faster movements can be described as long as the slow movements are well described by CVs (they quickly follow the movements of these slow CVs).
4. the number of CVs used must be small in order to decrease the computational cost.

The selected CVs (puckering coordinates) fulfill all the above requirements and therefore are good CVs to do metadynamics.

5.4. COMPUTATIONAL DETAILS

The procedure and its different computational parameters are detailed below and resumed in **Table 1**. The procedure of this project is divided into system preparation, equilibration and production. The software used have been: CPMD 3.15.1[22], Plumed driver[23], VMD[24], Gnuplot[25] and matplotlib[26].

5.4.1. System preparation

System preparation consist on defining and minimizing the system to be ready for the next steps, i.e., minimizing the density functional and then performing a geometry optimization (this simulation is in the gas phase, so no solvent is needed to add). Density functional minimization (also called "wavefunction optimization" in CPMD jargon) is used to obtain the electronic structure of a given system with fixed nuclei. Of course, since the nuclei are fixed, this structure is not relaxed (ring tension…), but this is the first step needed for the description of our system. Geometry optimization is used to "relax" a little bit the structure obtained in the wavefunction optimization, and it consists of searching the positions of the nuclei that have the minimum energy.

The molecule initial geometry used is shown in **Appendix 1**, **Table 6**, and has $^4C_1$ conformation. In this project, only GGA-type functionals have been used, particularly PBE, BLYP and HCTH; therefore, the entire procedure has been done with each functional. Regarding the basis set size, a large and conservative kinetic energy, E_{cut}, of 70 Ry (Rydberg) has been taken because it is the largest E_{cut} value needed for any of the atoms in each functional. The simulation box has been calculated as the difference between the maximum and minimum value of each axis plus a margin of 7.5 Å (angstrom) and the result has been 15.2 Å x 14.1 Å x 10.3 Å (rounded values). The system has been considered as isolated, i.e., no periodic boundary conditions have been applied.

5.4.2. Equilibration

To relax the structure completely and bring it to desired conditions for production, a few picoseconds (ps) of MD were done changing these conditions slowly otherwise the system can adopt undesired and unrealistic geometry or, in extreme cases, the simulation can crash. This procedure is commonly referred as system equilibration.

NOTE: picoseconds is the typical AIMD simulation timescale.

As we are using CPMD, new parameters must be introduced. We defined 250 steps as 1 frame of the simulation. Also, the MD time step has been defined as 5 a.u. (atomic units) or 0.12 fs (femtoseconds). The equilibration time has been taken as 13 ps (see next paragraph). Finally, the fictitious electron mass needed for the electronic equation of motion, equation (8), is taken as 850.0 a.u. The timestep and fictitious electron mass values should be tested and see their fluctuations in order to determine them, but these values have been taken as reference values from a previous study of the group on the same system.[8]

The equilibration time has been taken as 13 ps. This value can be justified if we analyze the Root-Mean-Square Deviation (RMSD), which is an average value of the distance that each atom of the molecule has moved with respect to initial position and helps us to see if the simulation has been able to reach an equilibrium position. Therefore, if the average RMSD remains stable, it means that we have already reached the equilibrium. E.g., equilibration using the PBE functional has given the RMSD of **Figure 13**. It can be seen that the RMSD value does not change too much (around 0.15 Å), thus 13 ps is enough. Using BLYP and HCTH, the obtained RMSD are shown in **Appendix 1**, **Figures 31** and **32**.

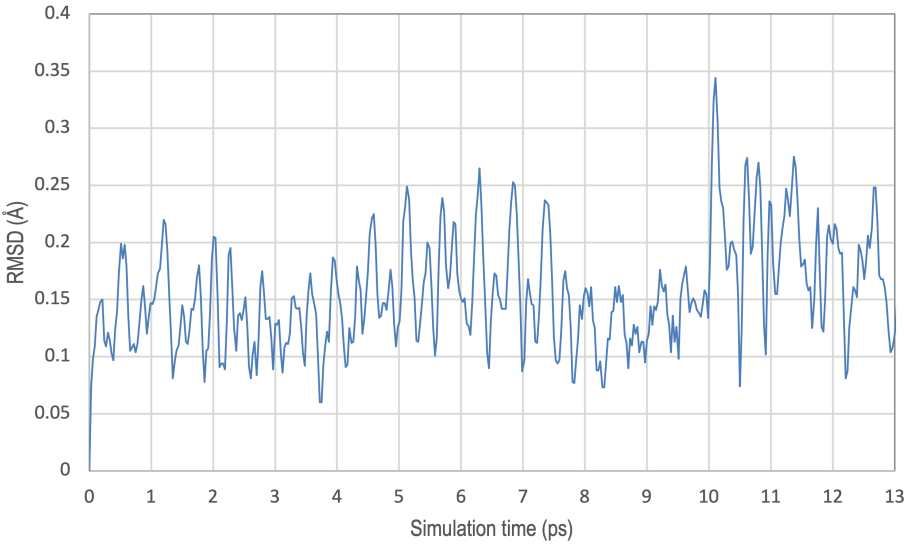

Figure 13. RMSD of the 13 ps equilibration using the PBE functional.

Since we plan to work at 300 K (Kelvin), equilibration must be done bringing the molecule from 0 K to 300 K. In order to stabilize this temperature and prevent it from fluctuating too much,

a Nosé thermostat has been added for nuclei at 300 K, with a frequency of 1000 cm^{-1} (this last value has been taken from previous work[8]).

Another thing that must be avoided is a large fluctuation of the fictitious electronic kinetic energy of the CPMD Lagrangian. Because as CPMD assumes that the electron system is like fictitious particles (propagating electronic orbitals via Newton's law), this energy must be (approximately) the same throughout the MD. To stabilize the electronic kinetic energy and prevent it from fluctuating too much, another Nosé thermostat has been added for electrons. Its parameters were determined as described below. The electronic thermostat has been added from picosecond 9 onwards. Therefore, the electronic kinetic energy of the first 9 ps is less stable than that after 9 ps. In order to know the energy value “preferred" by the system to introduce it in the thermostat, reoptimization of the wavefunction at the beginning of MD is done, i.e., a quench of electronic wavefunction onto the Born-Oppenheimer surface.[22] Applying this trick, one can see that just when the reoptimization is done, the electronic kinetic energy reaches its minimum, the stable value that is wanted. To do this, two wavefunction reoptimization have been necessary, at times 4 ps and 8 ps. The electronic kinetic energy evolution during this first 9 ps of equilibration is shown in **Figure 14** for PBE functional. For BLYP and HCTH functionals, the obtained results are shown in **Appendix 1**, **Figures 33** and **34**.

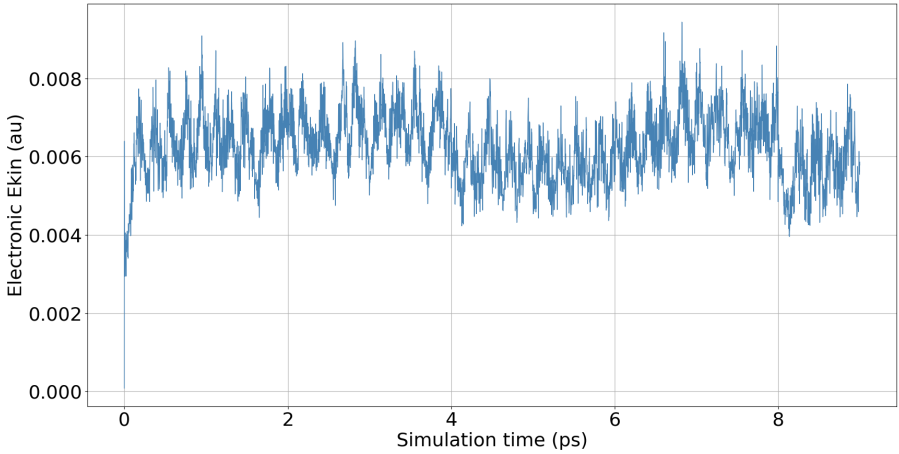

Figure 14. Electronic kinetic energy of the first 9 ps equilibration using the PBE functional.

It can be seen in **Figure 14** that when a quench of the wavefunction is done, the energy stabilizes. Taking the last quench (picosecond 8) and doing a linear regression of this last picosecond to see its destabilization, a $9.09816 \cdot 10^{-8}$ a.u.·step^{-1} slope and 0.005 a.u. y-intercept is obtained. On the one hand, if the slope is of the order of $10^{-5} - 10^{-4}$ a.u.·ps^{-1}·atom^{-1}, it can be said that the equilibration is succeeding; the smaller, the better. On the other hand, the y-intercept tell us about the energy minimum of the quench and thus, this value is the energy of the electronic thermostat. For PBE, the slope is $3.2 \cdot 10^{-5}$ a.u.·ps^{-1}·atom^{-1} (with the corresponding units) and the interception is 0.005 a.u. or E_h (Hartree). For BLYP, $2.8 \cdot 10^{-5}$ a.u.·ps^{-1}·atom^{-1} and 0.005 a.u. And for HCTH, $2.6 \cdot 10^{-5}$ a.u.·ps^{-1}·atom^{-1} and 0.007 a.u. Finally, a 10500 cm^{-1} frequency for the electronic thermostat has been chosen because it is the value that is usually used in sugar simulations.

5.4.3. Production

Production refers to the MTD simulation (described in **5.2**).

The simulation time chosen has been 60 ps, based on the time simulated by Biarnés et al.[4] in 2007. To assess whether 60 ps time is enough for convergence, the explored regions must be analyzed, as is done in **6.1**. In the simulation of Biarnés et al.[4], only q_x and q_y were used as CVs, thus only the northern hemisphere of the puckering sphere was sampled. In this work, we are extending the phase space to include also the southern hemisphere. For this reason, the CVs used are q_x, q_y *and* q_z (see **5.3**). Gaussians height and width have been taken from Biarnés et al.[4]: 0.3 kcal·mol^{-1} and 0.15 Å respectively. To know the values of width, we should do tests doing a few ps of MD to see the oscillation of each CV, e.g., if q_x oscillate between -0.075 and 0.075 Å, so σ_x=0.15 Å. The height value is usually a tenth of barrier energy, e.g., $^4C_1$ to S energy is around 3 kcal, so w=0.3 kcal. Finally, for Gaussian pace or deposition has been used a larger (more conservative) value than the value used by Biarnés et al.[4] because of the current computational power.

	Parameter	Value/Description
System preparation	Initial geometry	$^4C_1$ conformation (**Appendix 1**, **Table 6**)
	Functional	PBE/BLYP/HCTH
	Basis size	70 Ry
	Box size	15.2 Å x 14.1 Å x 10.3 Å (rounded values)
Equilibration	1 step	5 a.u. or 0.12 fs
	Equilibration time	13 ps
	Fictitious electron mass	850 a.u.
	Temperature	300 K
	Thermostats for nuclei	Nosé, 300 K, 1000 cm^{-1}
	Quench of wavefunction	At 4 ps and 8 ps
	Thermostats for electrons	Nosé, 0.005 a.u., 10500 cm^{-1} (PBE) Nosé, 0.005 a.u., 10500 cm^{-1} (BLYP) Nosé, 0.007 a.u., 10500 cm^{-1} (HCTH)
Production	MTD time	60 ps
	CVs	q_x, q_y, q_z
	Gaussian height	0.3 kcal·mol^{-1}
	Gaussian width	0.15 Å
	Gaussian pace	250 steps
	Deposited Gaussians	1980

Table 1. Summary of the computational details.

6. RESULTS AND DISCUSSION

In these sections, the metadynamics trajectory and conformation space sampled, some β-glucose properties and the conformational FELs obtained using different functionals are discussed.

Firstly, the metadynamics trajectory and the conformational FEL results are analyzed using the PBE functional, as it is the one used in previous works of the group.[4], [29] Within the trajectory analysis, puckering amplitude is analyzed. Intramolecular hydrogen bonds are also analyzed to justify the relative stability of each conformation. Then, a comparison between the conformational FEL using PBE and those obtained using two other functionals, BLYP and HCTH, is done. The energetic results obtained are compared with those of others works.

6.1. TRAJECTORY ANALYSIS

To analyze the trajectory, Stoddart's and Mercator's representations are used to observe the explored regions and when they have been explored. These types of plots are very useful to evaluate if our simulation needs to be extended in order to reach convergence. Mercator's representation is a cylindrical projection of the Cremer and Pople's sphere (equivalent to the typical representation of the Earth maps).

6.1.1. Explored conformations using the PBE functional

Stoddart's and Mercator's representations are respectively shown in **Figures 15** and **16**.

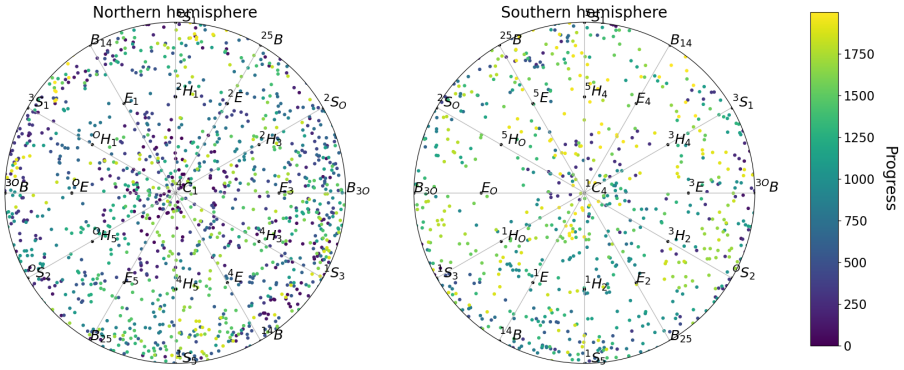

Figure 15. Conformations sampled in the MTD simulation in Stoddart's representation (PBE, 60 ps).

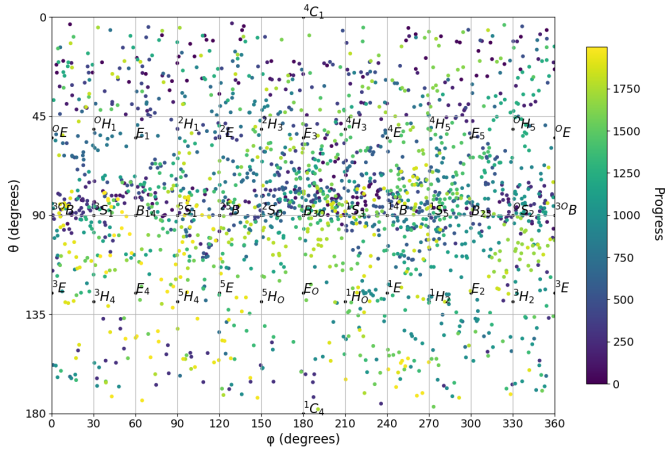

Figure 16. Conformations sampled in the MTD simulation in Mercator's representation (PBE, 60 ps).

The explored regions in our simulation are plotted in both representations color coded between purple, 0 frames or 0 ps, and yellow, 2000 frames or 60 ps. Stoddart's representation in both poles uses the cartesian coordinates q_x and q_y and Mercator's representation uses polar coordinates θ and ϕ of the Cremer and Pople's sphere.

NOTE: because both representations give no information about the puckering amplitude Q, in the following section, **6.1.2**, it is analyzed.

In **Figures 15** and **16** it can be seen that the northern hemisphere is more explored than the southern hemisphere because it has a higher concentration of points. The southern hemisphere has been explored at the end of the simulation as the points inside it tend to the yellow color. The results indicates that the most stable conformers are located in the northern hemisphere.

Since the behavior of MTD is not entirely diffusive, i.e., the yellow points that are at the end of the simulation are practically not in the northern hemisphere (see **Figure 15**), more picoseconds of simulation are needed to ensure convergence of the FEL. Therefore, the MTD has been extended to 200 ps but still does not have a diffusive behavior (see **Appendix 2**, **Figures 35** and **36**). More simulation time is needed, but it is out of the scope of this project.

NOTE: despite having computed 200 ps of simulation, the posterior treatments are done with the 60 ps simulation to be able to compare with the other functional results, not extended due to limitations of time.

6.1.2. Puckering amplitude evolution using the PBE functional

Since Stoddart's and Mercator's representation give no information about Q, the evolution and its frequency throughout the simulation are shown in **Figures 17** and **18**, respectively.

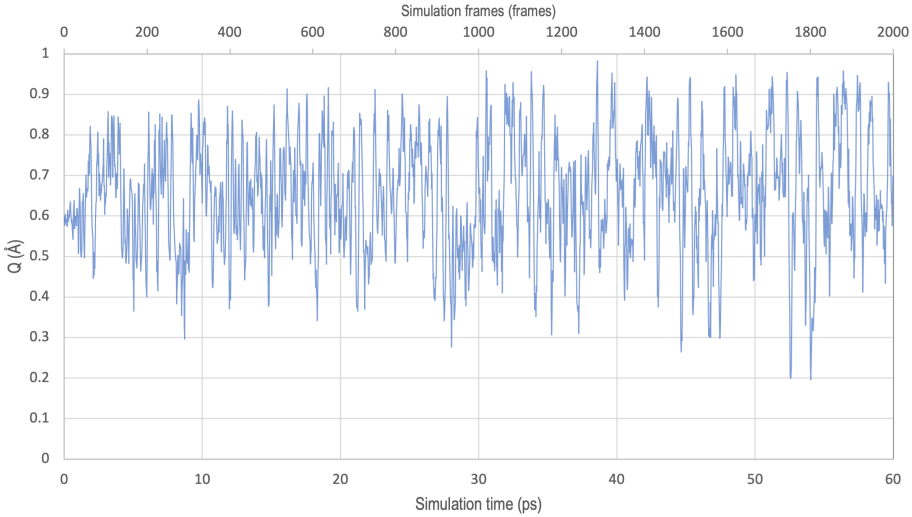

Figure 17. Puckering amplitude throughout the simulation.

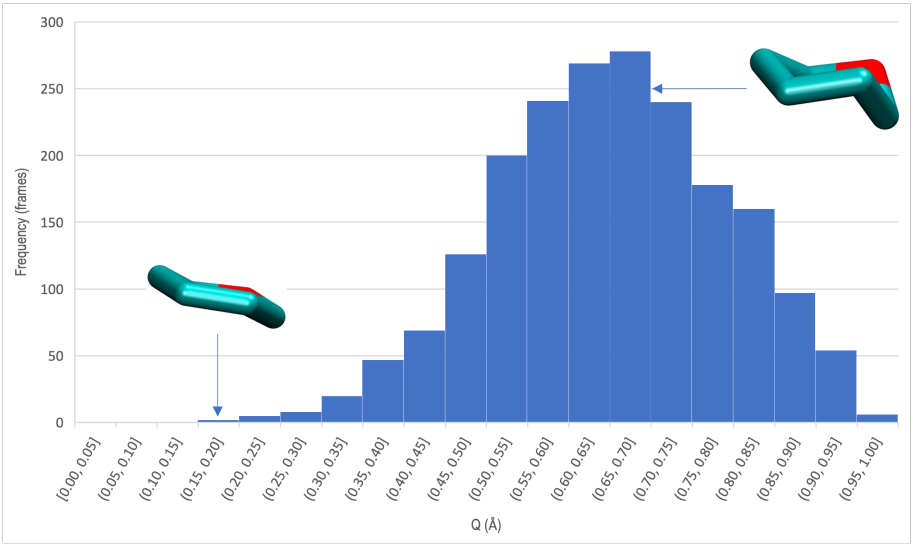

Figure 18. Frequency of puckering amplitude. The $^4C_1$ conformation is represented for two different values of Q.

The puckering amplitude is useful to know how much the conformations are to the center of Cremer and Pople's sphere. It measures how much "puckered" the ring is. If the Q value is 0 Å, the ring is totally planar, a situation very high in energy and never sampled. The other extreme is 1 Å since Q has been normalized to take values between 0 and 1. When Q is close to 1 Å, the glucose ring is more bent.

As can be seen in **Figure 17**, the value of Q is around 0.6 Å throughout the simulation. Still, it is observed that there is a tendency to oscillate between higher and smaller Q ranges as the simulation progresses, probably because as bias potential Gaussians are added, i.e., energy is provided to the system, it can adopt more unstable conformations. Therefore, the value of Q can vary much more by forcing the ring to be flatter or more bent.

In **Figure 18**, it can be confirmed that Q is usually between 0.65 and 0.70 Å. The ring shape has been depicted for Q = 0.15-0.20 Å (more flattened) and 0.65-0.70 Å (more bent).

6.1.3. Comparison using different functionals

The explored conformations of the β-glucose molecule using BLYP and HCTH functionals are shown below using Stoddart's and Mercator's representations. BLYP results are shown in **Figures 19** and **20** and HCTH results are shown in **Figures 21** and **22**.

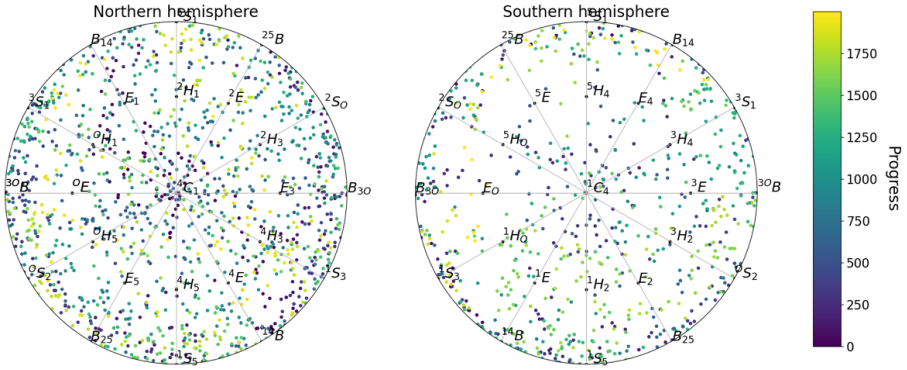

Figure 19. Conformations sampled in the MTD simulation in Stoddart's representation (BLYP, 60 ps).

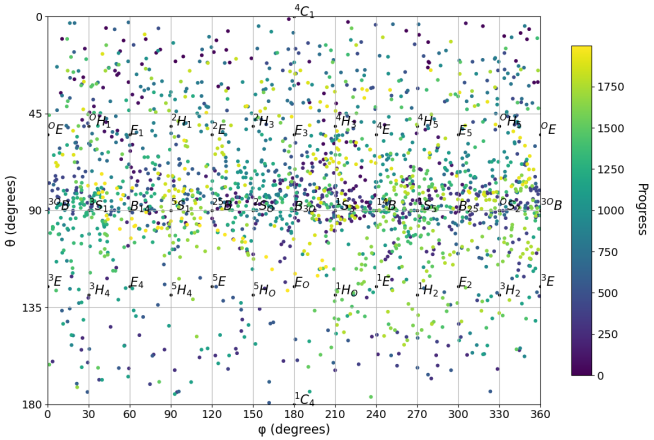

Figure 20. Conformations sampled in the MTD simulation in Mercator's representation (BLYP, 60 ps).

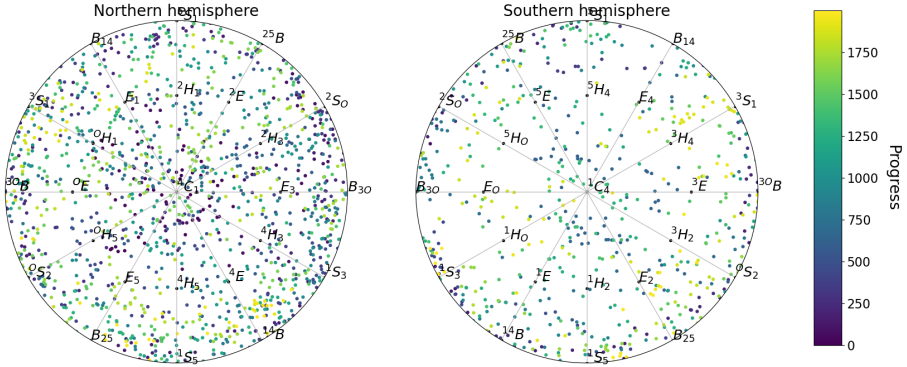

Figure 21. Conformations sampled in the MTD simulation in Stoddart's representation (HCTH, 60 ps).

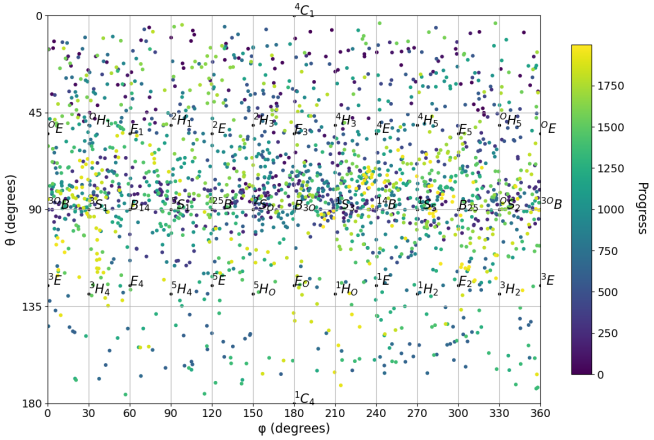

Figure 22. Conformations sampled in the MTD simulation in Mercator's representation (HCTH, 60 ps).

Beforehand, comparing Stoddart's representation of each functional it can be seen that using BLYP, a more diffusive behavior is obtained, i.e., it would seem that convergence would be reached earlier.

The frequency of each glucose conformation throughout the simulation is plotted in **Figure 23** for the three different functionals. The plots are similar for each functional since the most visited conformations are chairs (C), boats (B), and skews (S), while envelopes (E) and half-chairs (H) are the less visited ones. It is not surprising that these results have been obtained because E and H are more energetic (see **6.2**), and the simulation has taken a longer time to visit them.

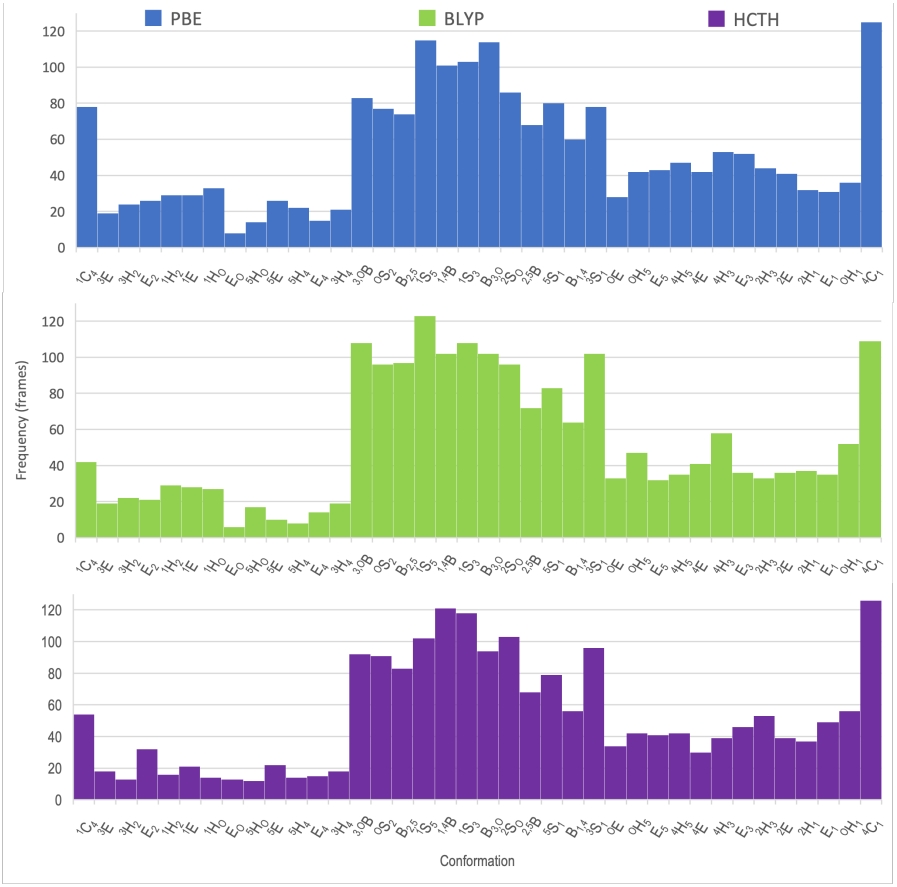

Figure 23. Frequency of the glucose conformations throughout the 2000 frames of the simulation.

6.2. FREE ENERGY LANDSCAPE

The Free Energy Landscape (FEL) informs of the free energy values (ΔG) as a function of some coordinates (in this project, it would be the CVs q_x, q_y and q_z). This is useful to know the free energy barriers of interconversion and the relative energies among the stable ring conformations. A FEL can be built using the MTD method which applies equations (10) and (11), as explained in point **5.2**.

To compute a conformational FEL for 60 ps simulation, 1980 Gaussians have been needed.

6.2.1. Conformational FEL using the PBE functional

Since the conformations of the whole Cremer and Pople's sphere are studied, the conformational FEL would have four dimensions: q_x, q_y, q_z, and ΔG. For visualization purposes, we separate the sphere in two Stoddart's diagram corresponding to the northern and southern hemisphere. To do so, we cut the sphere in a certain q_z value and integrate all the energies from each hemisphere. However, if the sphere is cut by its equator (q_z = 0 Å), some information from it could be lost. Therefore, the northern hemisphere must be cut by q_z values smaller than 0 Å, while the southern hemisphere, with q_z values greater than 0 Å, to ensure not losing any information from the equator.

For the northern hemisphere, at the q_z values chosen were -0.1, -0.2, and -0.3 Å, while for the southern hemisphere, they were 0.1, 0.2, and 0.3 Å. The chosen conformational FELs are the ones with $|q_z|$ = 0.2 Å and are shown in **Figure 24**. The discarded conformational FELs are shown in **Appendix 3**, **Figures 37** and **38**.

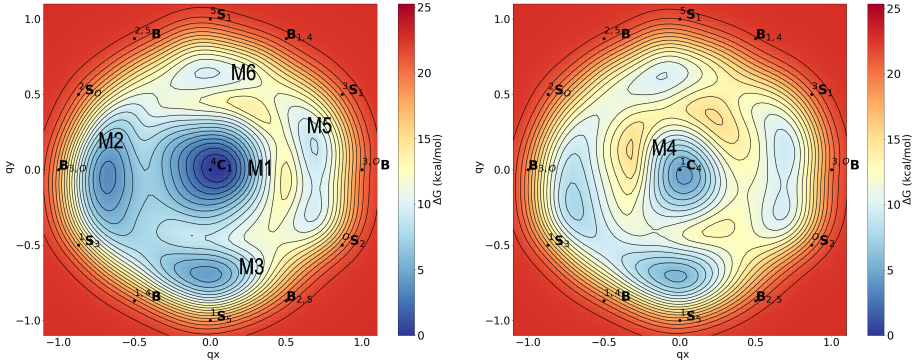

Figure 24. Northern (left) and southern (right) hemispheres conformational FELs that have been obtained cutting at $|q_z|$ = 0.2 Å (using PBE). The minima are labeled according to their relative stability.

The minima with their coordinates, conformations and energies, are resumed in the **Table 2**.

Minima	Hemisphere	Coordinates q_x, q_y (Å)	Conformation	Energies (kcal·mol^{-1})	Relative energies (kcal·mol^{-1})
M1	North	0.05, 0.01	$^4C_1$	-25.2659	0.0000
M2	North	-0.69, -0.05	$B_{3,O}$	-21.6755	3.5904
M3	North	-0.01, -0.69	$^1S_5$	-20.7785	4.4874
M4	South	0.02, -0.02	$^1C_4$	-20.7016	4.5643
M5	North	0.67, 0.17	$^{3,O}B/^3S_1$	-16.3568	8.9091
M6	North	-0.04, 0.64	$^5S_1$	-15.444	9.8219

Table 2. Properties of the FEL (using PBE). The relative energies are with respect to $^4C_1$.

The relative stability of each conformation can be discussed considering two main aspects[27]:

1. electrostatic interactions (e.g., hydrogen bonds) and steric effects.
2. electronic effects (e.g., anomeric effect).

On the one hand, hydrogen bonds (HBs) are electrostatic interactions that stabilize the potential energy of a system, i.e., the more HBs a conformation has, the more stable it will be. On the other hand, the anomeric effect helps to stabilize conformations by hyperconjugation, i.e., interaction between non-bonding electron pair of the oxygen and the anti-bonding orbital of the anomeric carbon, C1.

Figure 25 shows the four most stable conformations with their most important interactions; HBs. $^4C_1$ is the most stable conformation as it can create a stabilizing "clockwise" hydrogen bond pattern between the hydroxyls of C1, C2, C3, C4 and with smaller frequency also with the hydroxyl of C6, thus a total number of 4 hydrogen bond interactions. $B_{3,O}$ is the second most stable conformation due to the loss of the hydrogen bond between the hydroxyls of C1 and C2 while the hydroxyl group of C6 interacts with OH group from C1 instead of C4. $^1S_5$ has still one less hydrogen bond interaction, a hydrogen bond interaction between the hydroxyls of C1 and C3 and another one between the hydroxyls of C3 and C4. Finally, $^1C_4$ is not as stable as $^4C_1$ even though it has the same number of hydrogen bond interactions due to how the exocyclic groups are placed, in an axial orientation; less stable due to steric reasons, not even compensated with the anomeric effect. The study of HBs is shown in **Appendix 3**, **Table 7**.

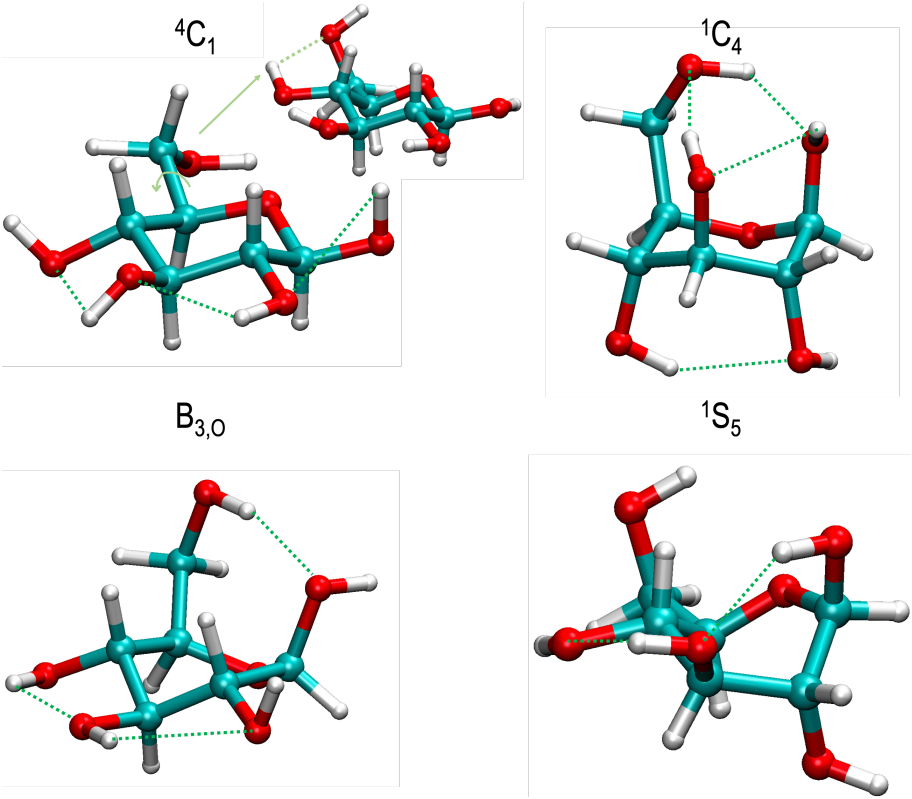

Figure 25. HBs (green) justify relative stability. The HBs considered have been taken smaller than 3.3 Å.

Furthermore, it can be seen in **Figure 24** that envelope and half-chair conformations are found between chair minimum and the other minima, therefore, are more energetic than C, B and S conformations, which justifies that E and H have been visited less during the simulation (see **Figure 23**).

Also, one thing that calls our attention is that not all minimums correspond to canonical conformations, e.g., M5 corresponds to a conformation between $^{3,O}B$ and $^3S_1$. As reported Biarnés et al.[4], this fact is due to hydrogen bonds and can occur because of the ring flexibility. The Stoddart's diagram was made for cyclohexane and the canonical conformations coincide with FEL minima[28], but glucose has exocyclic groups that interact with each other. These interactions, e.g., HBs, are the cause of the deviation of the minima because stabilize intermediate conformations.

Regarding the energy convergence, in **Table 3** and in **Figure 26** are shown the relative energies to the global minimum and its evolution throughout the simulation. There is no clear trend towards convergence, a fact that demonstrates the need to lengthen the simulation, as has already been seen in the previous sections.

Minima	Relative energies (kcal·mol^{-1})						
	12 ps	24 ps	36 ps	42 ps	48 ps	54 ps	60 ps
M1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
M2	5.7234	5.3809	3.8525	5.6024	3.8794	3.6115	3.5904
M3	7.6186	7.1748	4.4289	6.8483	5.3261	5.754	4.4874
M4	5.4842	8.1639	5.0044	7.1579	5.8165	6.7887	4.5643
M5	1.4944	5.0025	4.6492	7.1791	8.2477	10.1716	8.9091
M6	6.4849	8.5359	7.7844	10.6741	10.1624	11.5119	9.8219

Table 3. Minima energy values along the simulation (using PBE).

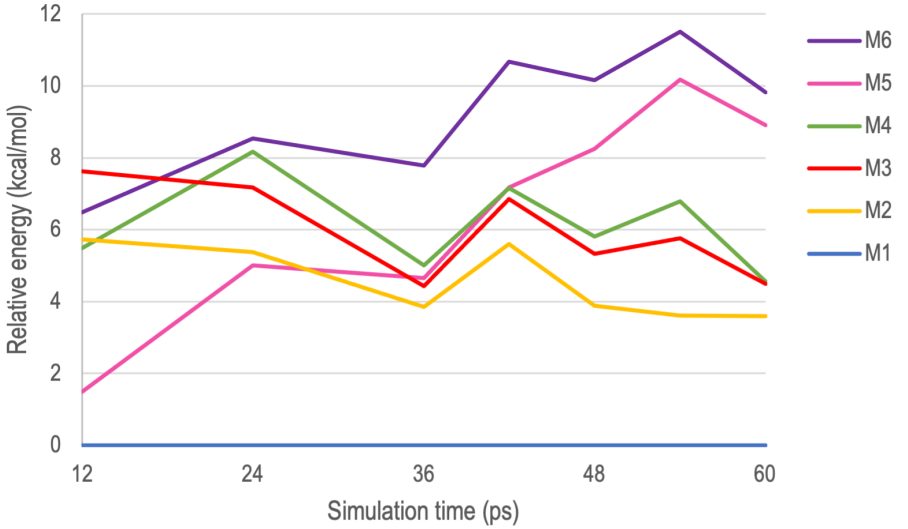

Figure 26. Graphical representation of the energy convergence (using PBE).

The conformational FEL evolution in Stoddard's representation for both hemispheres throughout the simulation has also be done and is shown in **Appendix 3**, **Figures 39** and **40**.

6.2.2. Comparing the conformational FELs using different functionals

To compare the difference between functionals (PBE vs BLYP and HCTH) it has also been decided to cut at $|q_z|$ = 0.2 Å. The FELs obtained are shown in **Figures 27** and **28**.

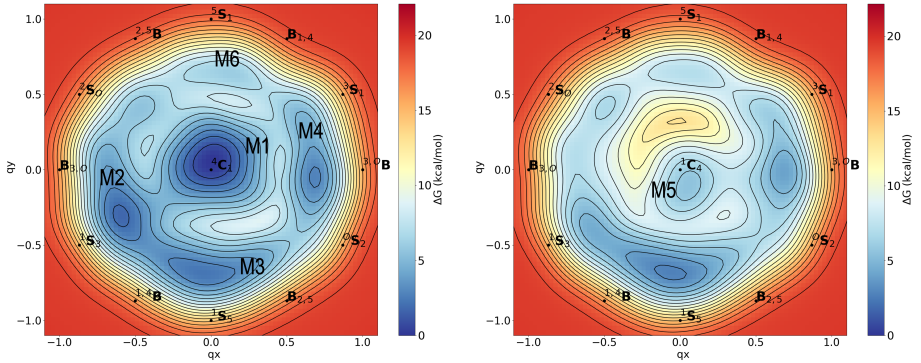

Figure 27. Northern (left) and southern (right) hemispheres FELs that have been obtained cutting at $|q_z|$ = 0.2 Å (using BLYP).

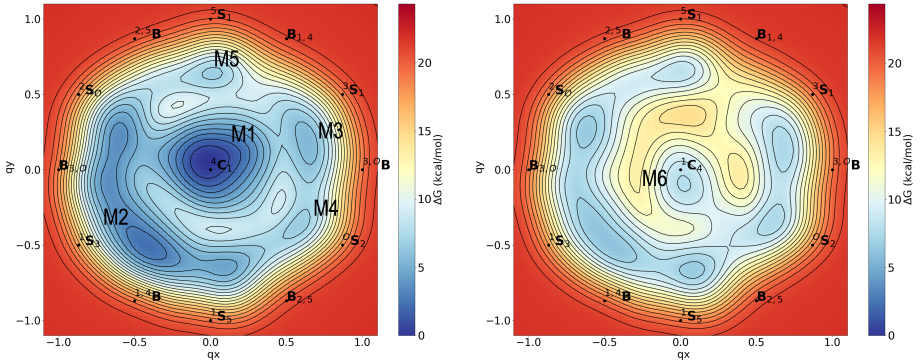

Figure 28. Northern (left) and southern (right) hemispheres FELs that have been obtained cutting at $|q_z|$ = 0.2 Å (using HCTH).

A table has been made to make it easier to compare the results obtained with the different functionals (**Table 4**). This table shows the comparison of the minima with their respective conformation and energy for PBE, BLYP, and HCTH.

NOTE: again, the minima are labeled according to their relative stability (M1, M2, …) instead of labeling as has been done for PBE because new minima appear, and others change their coordinates.

Minima	PBE		BLYP		HCTH	
	Relative energies (kcal/mol)	Conformation	Relative energies (kcal/mol)	Conformation	Relative energies (kcal/mol)	Conformation
M1	0.0000	$^{4}C_{1}$	0.0000	$^{4}C_{1}$	0.0000	$^{4}C_{1}$
M2	3.5904	$B_{3,O}$	1.7341	$^{1}S_{3}$	2.2392	$^{1}S_{3}/^{1,4}B$
M3	4.4874	$^{1}S_{5}$	2.2916	$^{1,4}B/^{1}S_{5}$	5.4927	$^{3,O}B/^{3}S_{1}$
M4	4.5643	$^{1}C_{4}$	2.7742	$^{3,O}B$	6.5724	$^{O}S_{2}$
M5	8.9091	$^{3,O}B/^{3}S_{1}$	5.4838	$^{1}C_{4}$	6.6974	$^{5}S_{1}$
M6	9.8219	$^{5}S_{1}$	6.0395	$^{5}S_{1}$	7.7760	$^{1}C_{4}$

Table 4. Properties of each FEL. The relative energies are with respect to the global minimum (in all cases it is $^{4}C_{1}$).

It can be seen in **Table 4** that the global minimum is $^{4}C_{1}$ for all the functionals, but the other chair conformation, the $^{1}C_{4}$, is more stable for the PBE functional than the other ones. Furthermore, the second minimum (M2) for PBE corresponds to $B_{3,O}$ while for BLYP and HCTH corresponds to $^{1}S_{3}$.

The fact that PBE stabilizes some conformations that the other functionals do not, is due to the lousy description of HBs by BLYP and HCTH because, as it has been said above (see **Figure 25**), $^{1}C_{4}$ and $B_{3,O}$ present a great stabilization by HBs while $^{1}S_{5}$, it is only stabilized by two HBs. Therefore, if a functional with a poorer description of HBs is used, the conformations that are stabilized due to HBs will be less stable, hence with this example we can see the importance of choosing the appropriate functional for a given system.

As in this project we are working with sugars, which present many HBs, the PBE functional is the best option to ensure a good description of the system and, above all, obtain a reliable result.

Another important factor to consider is the computational cost, i.e., how much time the computer spends to do the calculations. **Figure 29** shows the computational cost expressed in seconds per step for each functional used.

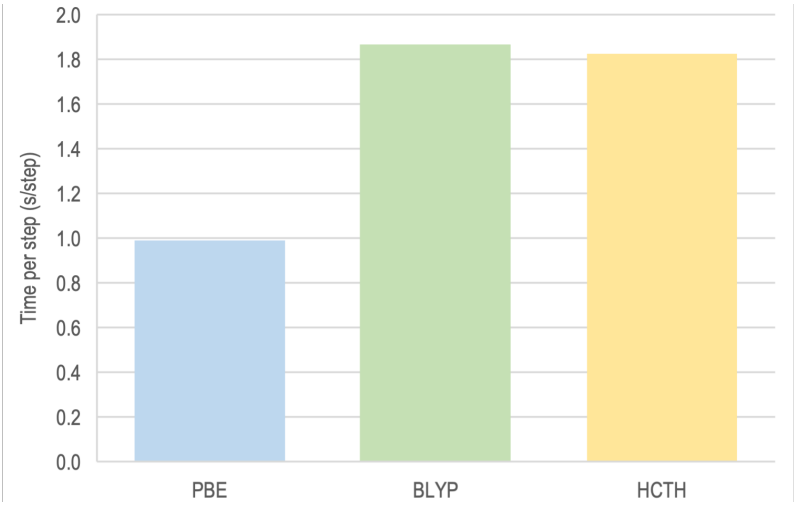

Figure 29. Computational cost for PBE, BLYP and HCTH.

The PBE, in addition to describing HBs much better, also requires a lower computational cost, and therefore for sugars, it has been a good choice in terms of accuracy and computational cost.

6.2.3. Comparison with previous works

In this section, our results in the northern hemisphere are compared with the ones obtained by Biarnés et al.[4]

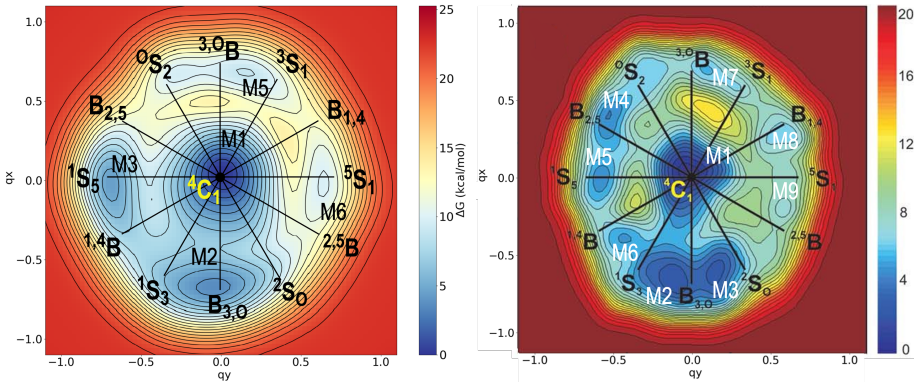

Figure 30. FEL obtained in this project (left) and FEL obtained by Biarnés et al.[4] (right; image taken from reference [28]), both using the PBE functional. The axes have been placed as in Biarnés et al. to compare.

From **Figure 30** can be seen that:

- the global minimum is $^4C_1$ conformation
- local minima and maxima are approximately in the same coordinates
- they obtained more minima

Minima	Our FEL			Biarnés et al.'s FEL[4]		
	Relative energies (kcal/mol)	Conformation	Coordinates q_x, q_y (Å)	Relative energies (kcal/mol)	Conformation	Coordinates q_x, q_y (Å)*
M1	0.0000	$^4C_1$	0.05, 0.01	0.0	$^4C_1$	0.03, -0.03
M2	3.5904	$B_{3,O}$	-0.69, -0.05	2.6	$B_{3,O}$	-0.63, -0.12
M3	4.4874	$^1S_5$	-0.01, -0.69	3.0	$B_{3,O}$ /^2S_O	-0.60, 0.15
M4		($^1C_4$)		5.5	$B_{2,5}$	0.40, -0.53
M5	8.9091	$^{3,O}B/^3S_1$	0.67, 0.17	5.8	$^1S_5$	0.00, -0.58
M6	9.8219	$^5S_1$	-0.04, 0.64	6.3	$^{1,4}B/^1S_3$	-0.38, -0.42
M7	-	-	-	7.2	$^{3,O}B$	0.68, 0.08
M8	-	-	-	7.9	$B_{1,4}$	0.28, 0.57
M9	-	-	-	9.0	$^{2,5}B/^5S_1$	-0.18, 0.58

Table 5. FEL comparison with Biarnés et al.[4]. The colors mark the correspondence between the minima of both results. Approximate values* (the original ones do not show agreement with FEL[4]).

In **Table 5** there is a more exhaustive comparison. M1 corresponds in both cases to $^4C_1$ with the same coordinates. M2 is similar, but Biarnés et al. obtained two minima instead of one, and therefore their M2 is displaced a little bit but corresponds to the same conformation, $B_{3,O}$. Our M3 probably corresponds to their M4 and M5 and the relative energies are similar. Our M5 and their M7 are clearly the same minimum, and our M6 and their M9 too. They have two minima that we have not obtained; their M6 and M8. It could be thought that their M6 could be part of our M2 but M6 has a relative energy of 6.3 kcal/mol while it should be around 3 kcal/mol. Furthermore, their M8 is probably not a part of our M5 because of the difference in relative energy.

These slight differences between both FELs could be due to the improvements of the methods used. From 2007 to 2021, some parameters of the programs have been modified. Moreover, some simulation parameters such as the Gaussian deposition time are not the same.

However, both conformational FELs present the same low and high energy regions and are qualitatively very similar. Most importantly, the interpretation of the results does not alter the conclusions of the previous study of the group.

7. CONCLUSIONS

We can conclude from our initial objectives that:

1. the conformational FEL of the β-glucose has been obtained. The minima of the FEL do not correspond to the canonical conformations, as was previously observed in a study of the northern hemisphere conformations alone, because of the intramolecular interactions. The values of the relative energies of each conformation have been rationalized in terms of HB interactions. Furthermore, the trajectory analysis has been of great help to see the residence time of each conformation and to track the simulation.
2. PBE is the best functional (from the ones studied) to describe β-glucose because it describes HBs much better than BLYP and HCTH. Moreover, the computational cost of PBE is far below the others.
3. the northern hemisphere results obtained by the previous work are qualitatively similar but with slightly quantitative differences. These variations in results may be attributed to some parameters that have been modified and the increase of the computational power, which allows optimizing the error-computational cost ratio. The results for the southern hemisphere of the FEL are new thus no comparison with previous work applies.

8. REFERENCES

[1]. Lehninger, A. L.; Nelson, D. L.; Cox, M. M. Lehninger Principles of Biochemestry (Seventh Edition). *W. H. Freeman and Company*, **2017**.

[2]. Schwarz, J. C. P. Rules for Conformation Nomenclature for Five- and Six-membered Rings in Monosaccharides and their Derivatives. *Journal of the Chemical Society, Chemical Communications*, **1973**,14, 505-508.

[3]. Stoddart, J. F. In: Stereochemistry of Carbohydrates. *Wiley*, **1971**.

[4]. Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. The Conformational Free Energy Landscape of β-D-Glucopyranose. Implications for Substrate Preactivation in β-Glucoside Hydrolases. *Journal of the American Chemical Society*, **2007**, 129, 35, 10686-10693.

[5]. Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T. How Sugars Pucker: Electronic Structure Calculations Map the Kinetic Landscape of Five Biologically Paramount Monosaccharides and Their Implications for Enzymatic Catalysis. *Journal of the American Chemical Society*, **2014**, 136, 3, 1008-1022.

[6]. Taylor, E. J.; Goyal, A.; Guerreiro, C. I. P. D.; Prates, J. A. M.; Money, V. A.; Ferry, N.; Morland, C.; Planas, A.; Macdonald, J. A.; Stick, R. V.; Gilbert, H. J.; Fontes, C. M. G. A.; Davies, G. J. How Family 26 Glycoside Hydrolases Orchestrate Catalysis on Different Polysaccharides. *Journal of Biological Chemistry*, **2005**, 280, 38, 32761-32767.

[7]. Marx, D.; Hutter, J. Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods, *Cambridge University Press*, **2009**.

[8]. Biarnés, X. Mecanismes de Preactivació de Substrat en 1,3-1,4-β-glucanasa. *Memòria per optar al títol de Doctor per la Universitat de Barcelona*, **2007**.

[9]. Rovira, C. Study of Ligand-Protein Interactions by Means of Density Functional Theory and First-Principles Molecular Dynamics. In: Ulrich Nienhaus, G. Protein-Ligand Interactions. *Methods in Molecular Biology*™, **2005**, 305.

[10]. Kohn, W.; Becke, A. D.; Parr, R. G. Density Functional Theory of Electronic Structure. *Journal of Physical Chemistry*, **1996**, 100, 12974-12980.

[11]. Monticelli, L.; Tieleman, P. D. Force Fields for Classical Molecular Dynamics. *Biomolecular Simulations: Methods and Protocols, Methods in Molecular Biology*, **2013**, 924, 197-213.

[12]. Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Physical Review Letters*, **1985**, 55, 2471-2474.

[13]. Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters*, **1984**, 52, 12, 997-1000.

[14]. Buda, F. Density Functional Theory and Car-Parrinello Molecular Dynamics Methods. In: Aartsma, T.J.; Matysik, J. *Biophysical Techniques in Photosynthesis. Advances in Photosynthesis and Respiration*, **2008**, 26.

[15]. Jensen, F. Introduction to Computational Chemistry (Second Edition). *Wiley*, **2007**.

[16]. Lindner, A.; Strauch, D. A Complete Course on Theoretical Physics: From Classical Mechanics to Advanced Quantum Statistics. *Springer Nature Switzerland AG*, **2018**.

[17]. Niklasson, Anders M. N. Extended Born-Oppenheimer Molecular Dynamics. *Physical Review Letters*, **2008**, 100, 12, 123004.

[18]. Alonso, J. L.; Andrade, X.; Echenique, P.; Falceto, F.; Prada-Gracia, D.; Rubio, A. Efficient Formalism for Large-Scale Ab Initio Molecular Dynamics based on Time-Dependent Density Functional Theory. *Physical Review Letters*, **2008**, 101, 096403.
[19]. Nin-Hill, A. Conformational catalytic itineraries of five- and six-membered sugar rings in enzymatic and superacid media. *Memòria per optar al títol de Doctora per la Universitat de Barcelona*, **2020**.
[20]. Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics*, **2020**, 2, 200-212.
[21]. Cremer, D.; Pople, J. A. A General Definition of Ring Puckering Coordinates. *Journal of the American Chemical Society*, **1975**, 97, 6, 1354-1358.
[22]. CPMD, http://www.cpmd.org/, Copyright IBM Corp 1990-2019, Copyright MPI für Festkörperforschung Stuttgart 1997-2001.
[23]. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. Plumed 2: New feathers for an old bird. *Computer Physics Communications*, **2014**, 185, 2, 604–613.
[24]. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, **1996**, 14, 1, 33-38.
[25]. Williams, T.; Kelley, C. Gnuplot: An Interactive Plotting Program, **1986**.
[26]. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, **2007**, 9, 3, 90-95.
[27]. Marianski, M.; Supady, A.; Ingram, T.; Schneider, M.; Baldauf, C. Assessing the Accuracy of Across-the-Scale Methods for Predicting Carbohydrates Conformational Energies for the Examples of Glucose and α-Maltose. *Journal of Chemical Theory and Computation*, **2016**, 12, 6157-6168.
[28]. Iglesias-Fernández, J.; Raich, L.; Ardèvol, A.; Rovira, C. The complete conformational free energy landscape of β-xylose reveals a two-fold catalytic itinerary for β-xylanases. *Chemical science*, **2015**, 6, 1167.
[29]. Ardèvol, A.; Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *Journal of the American Chemical Society*, **2015**, 137, 24, 7528-7547.

9. ACRONYMS

Å	angstrom
a.u.	atomic units
AIMD	*Ab Initio* Molecular Dynamics
B	Boat
BLYP	Becke-Lee-Yang-Parr
BOMD	Born-Oppenheimer Molecular Dynamics
C	Chair
classical MD	classical Molecular Dynamics
CPMD	Car-Parrinello Molecular Dynamics
CV	Collective Variable
DFT	Density Functional Theory
E	Envelope
E_h	Hartree
EMD	Ehrenfest Molecular Dynamics
FEL	Free Energy Landscape
FES	Free Energy Surface
fs	femtoseconds
GGA	Generalized Gradient Approximation
H	Half-chair
HB	Hydrogen Bond
HCTH	Hamprecht-Cohen-Tozer-Handy
IUPAC	International Union of Pure and Applied Chemistry
K	Kelvin

KS	Kohn-Sham
LDA	Local Density Approximation
LSDA	Local Spin-Density Approximation
MD	Molecular Dynamics
MTD	Metadynamics
PBE	Perdew-Burke-Ernzerhof
ps	picosecond
RMSD	Root-Mean-Square Deviation
Ry	Rydberg
S	Skew
TDDFT	Time-Dependent Density Functional Theory
TDSE	Time-Dependent Schrödinger Equation
TISE	Time-Independent Schrödinger Equation
TS	Transition State

APPENDICES

APPENDIX 1: COMPUTATIONAL DETAILS

Atom	x (Å)	y (Å)	z (Å)
C	-1.599000	0.273000	-0.228000
C	-1.032000	-1.069000	0.244000
C	0.406000	-1.210000	-0.265000
C	1.234000	-0.022000	0.232000
C	2.657000	-0.124000	-0.322000
C	-0.694000	1.404000	0.269000
O	-1.186000	2.655000	-0.214000
O	-2.915000	0.448000	0.302000
O	-1.829000	-2.135000	-0.276000
O	0.971000	-2.426000	0.229000
O	0.635000	1.195000	-0.215000
O	3.460000	0.918000	0.235000
H	-1.639000	0.290000	-1.317000
H	-1.038000	-1.105000	1.333000
H	0.407000	-1.223000	-1.355000
H	1.266000	-0.033000	1.322000
H	3.082000	-1.092000	-0.056000
H	2.632000	-0.024000	-1.407000
H	-0.687000	1.412000	1.359000
H	-0.660000	3.418000	0.062000
H	-3.336000	1.279000	0.043000
H	-2.755000	-2.104000	0.001000
H	0.493000	-3.221000	-0.044000
H	4.377000	0.915000	-0.072000

Table 6. Initial atomic coordinates of β-glucose.

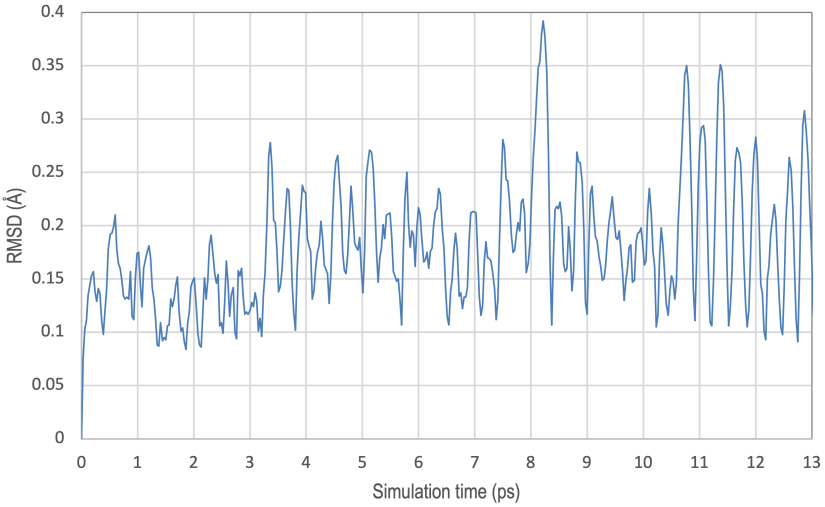

Figure 31. RMSD of the 13 ps equilibration using BLYP functional.

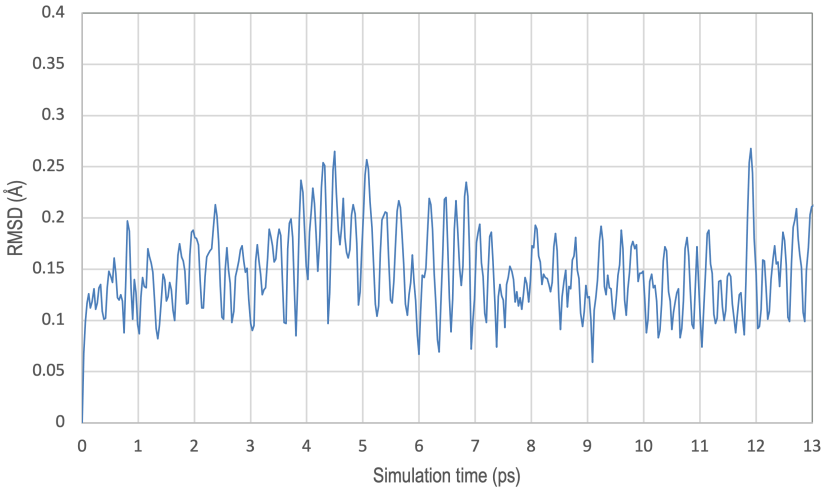

Figure 32. RMSD of the 13 ps equilibration using HCTH functional.

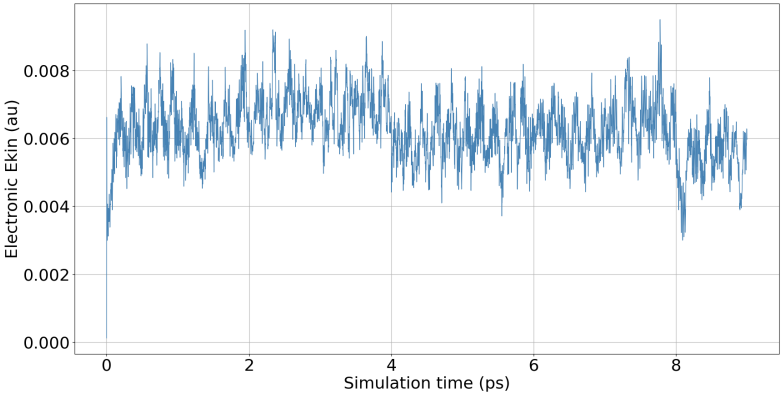

Figure 33. Electronic kinetic energy of the first 9 ps equilibration using BLYP functional.

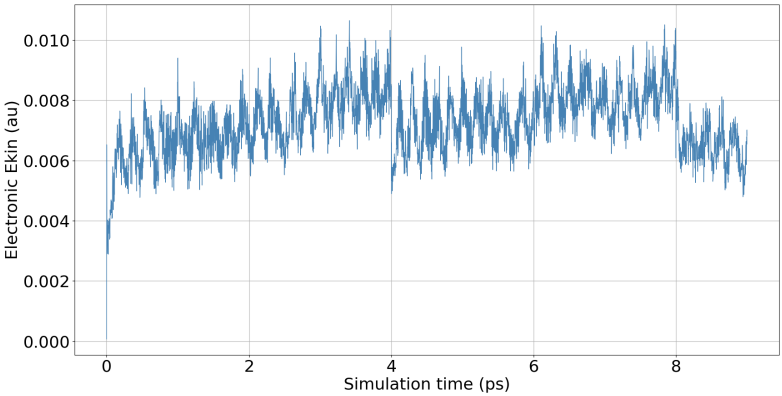

Figure 34. Electronic kinetic energy of the first 9 ps equilibration using HCTH functional.

APPENDIX 2: TRAJECTORY ANALYSIS

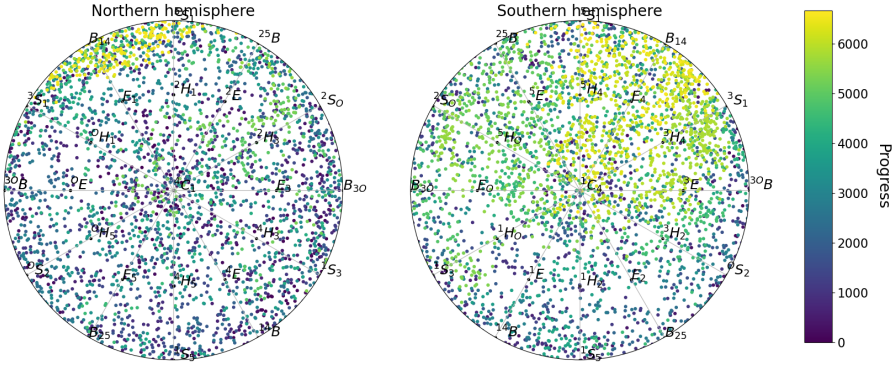

Figure 35. Conformations sampled in the MTD simulation in Stoddart's representation (PBE, 200 ps).

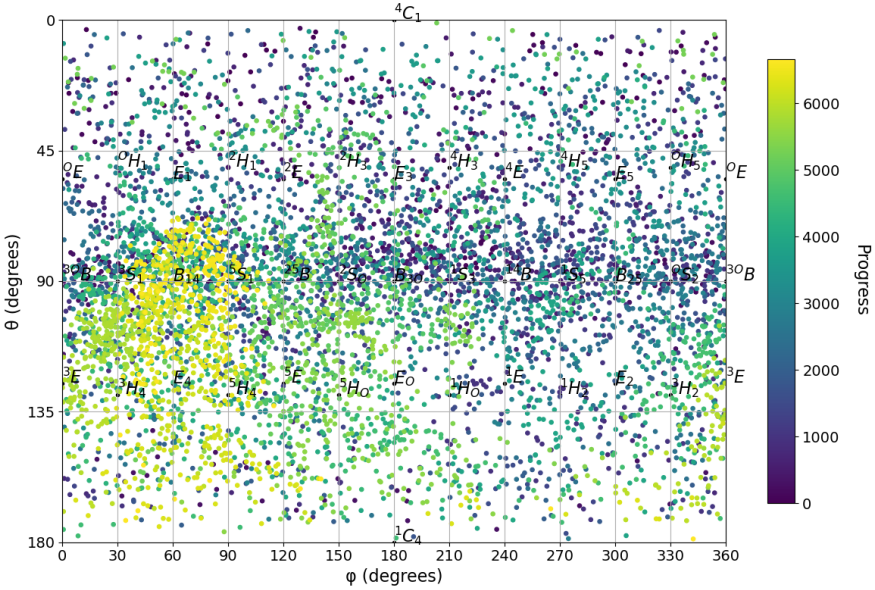

Figure 36. Conformations sampled in the MTD simulation in Mercator's representation (PBE, 200 ps).

APPENDIX 3: FREE ENERGY LANDSCAPES

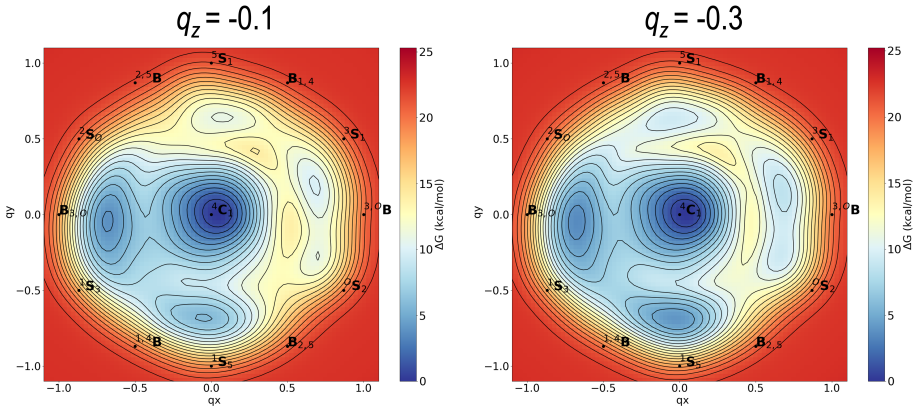

Figure 37. Northern hemisphere FELs that have been obtained cutting at q_z values of -0.1 and -0.3 (using PBE).

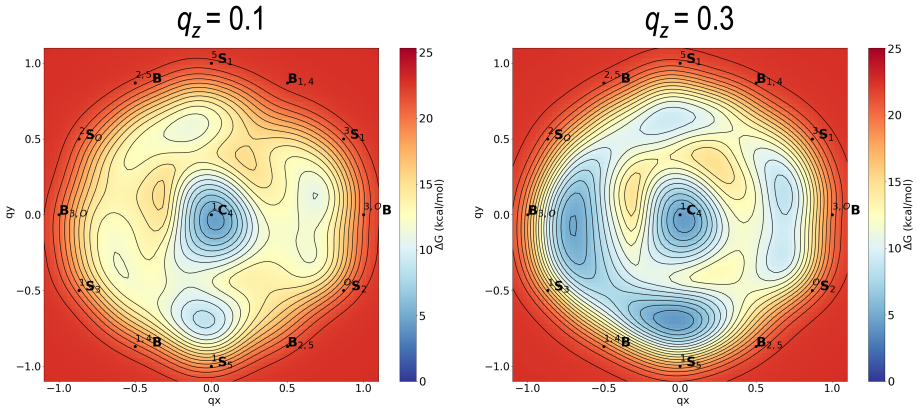

Figure 38. Southern hemisphere FELs that have been obtained cutting at q_z values of 0.1 and 0.3 (using PBE).

Conformation	$^{1}C_{4}$	^{3}E	$^{3}H_{2}$	E_{2}	$^{1}H_{2}$	^{1}E	$^{1}H_{O}$	E_{O}	$^{5}H_{O}$	^{5}E	$^{5}H_{4}$	E_{4}	$^{3}H_{4}$
HO1-HO2	4.0604	4.0168	4.0346	4.0119	3.8818	3.8222	3.9094	4.0225	3.8867	3.9574	3.7974	3.9074	3.8265
HO2-HO1	4.2317	4.232	4.2144	4.2188	4.1561	4.1163	4.1972	4.1408	4.0425	3.9201	4.1067	4.1111	4.1441
HO2-HO3	4.0173	4.1016	3.9405	4.0004	3.9956	4.0693	3.8072	3.868	3.8434	3.6503	3.7851	3.933	4.0303
HO3-HO2	3.9749	4.1711	4.106	4.0461	4.1731	4.1823	3.9539	3.5066	3.7975	3.6881	3.9252	3.9494	4.0378
HO3-HO4	4.0803	3.9666	3.9948	3.8949	3.7882	3.7475	3.6329	3.7252	3.8813	4.0382	4.2011	4.1023	4.0427
HO4-HO3	4.0925	4.0092	4.1517	4.0636	3.9542	3.7637	3.8484	3.6975	3.9254	4.1512	4.1705	4.0563	4.1049
HO4-HO6	4.7812	4.7958	4.3783	4.3763	4.3132	4.4635	4.5025	4.4533	4.767	4.8203	4.9857	4.9257	4.8497
HO6-HO4	4.93	4.9763	4.7539	4.7611	4.5953	4.7767	4.7891	4.7667	4.7942	4.7553	4.752	4.9546	5.2238
HO1-HO3	2.7148	2.835	2.2761	2.3344	2.5065	3.2575	3.7409	4.3428	4.2915	4.2875	3.8275	3.1478	3.1122
HO3-HO1	3.5697	3.6588	3.0698	3.2277	3.0625	3.1194	3.9626	4.4319	4.3323	4.4663	4.7974	4.4053	4.0152
HO2-HO4	2.9337	2.5826	2.6633	2.8206	3.481	3.7411	4.2638	4.042	3.6843	3.8415	3.3976	3.0601	2.8533
HO4-HO2	2.7169	2.9049	2.8355	2.9233	3.5733 3	3.6655	3.9942	4.3311	4.1895	3.7153	3.1231	2.6618	2.6781
HO1-HO6	2.8914	3.5442	3.5943	2.9787	2.8827	2.6715	2.605	2.4136	2.8444	3.649	2.9842	2.8867	3.8755
HO6-HO1	2.8558	3.8659	3.6441	3.169	3.1449	2.6546	2.6801	2.7871	2.6268	3.1615	3.7052	3.4058	3.9702
HO3-HO6	3.0521	3.3845	2.8608	3.3721	3.6152	3.982	4.0788	3.2618	3.4425	3.8647	3.8418	3.5431	3.6027
HO6-HO3	2.8026	2.4688	3.1342	3.235	3.7281	4.2177	4.5389	3.8182	3.4197	3.9119	3.3024	2.3262	2.8743
Mean	3.6066	3.7196	3.6033	3.5896	3.6782	3.7657	3.9066	3.8506	3.8606	3.9924	3.9189	3.7110	3.8276

$^{3,O}B$	$^{O}S_2$	$B_{2,5}$	$^{1}S_5$	$^{1,4}B$	$^{1}S_3$	$B_{3,O}$	$^{2}S_O$	$^{2,5}B$	$^{5}S_1$	$B_{1,4}$	$^{3}S_1$	^{O}E	$^{O}H_5$	E_5
3.6994	4.025	3.9774	3.9455	3.9386	3.9266	3.7682	3.3295	3.0606	2.8554	2.8525	3.302	2.9554	3.1545	3.5419
4.1085	4.2514	4.2205	4.2632	4.2406	4.2031	3.8427	3.645	3.5484	3.4069	3.6113	3.7627	3.6682	3.7567	3.9845
4.0225	3.9633	3.9197	3.9577	3.6715	3.3467	3.0575	2.8276	2.7777	3.1061	3.483	4.0412	3.3463	3.4598	3.4208
4.1023	4.0349	4.1743	4.043	3.8598	3.3573	3.032	3.2012	3.4192	3.697	3.9698	3.9978	3.6937	3.799	3.6885
3.9975	3.8911	3.5346	3.2417	2.8332	2.7145	2.771	3.0775	3.4802	3.8236	3.8966	3.8837	3.506	3.3469	2.909
4.0253	4.0527	3.8468	3.4541	3.3119	3.2992	3.2992	3.698	4.0565	4.1788	4.1689	3.9921	3.6186	3.4767	3.4858
4.6728	4.3713	3.8699	3.6358	3.881	4.3068	4.5493	4.7757	4.8144	4.7773	4.6592	4.9416	4.3961	4.0593	3.5899
4.8453	4.285	3.877	3.7706	4.0725	4.4683	4.6476	4.7996	5.083	5.0254	5.0568	5.0732	4.2198	3.8142	4.241
3.4271	2.7487	2.9944	3.3365	4.1704	4.8537	5.0443	4.9762	4.9024	4.9439	4.7735	4.298	5.0204	5.0534	4.6627
4.2471	3.8079	3.3871	3.8445	4.4168	4.9437	5.2672	5.359	5.3146	5.2903	5.291	4.8396	5.3042	4.8397	4.9459
2.8104	3.3605	4.1418	4.7226	5.0209	5.0622	5.0525	5.0354	4.7466	4.5777	4.1429	3.0206	5.0514	5.0742	4.9362
3.1766	3.2292	4.4154	4.6425	5.0338	5.1939	5.1398	5.2573	5.1367	4.6522	4.2111	3.6843	5.1779	5.1967	5.097
4.8313	4.7947	4.5726	4.2336	3.5378	3.2154	2.6428	3.3025	3.9694	4.1898	4.1499	4.7254	4.4695	4.7852	5.047
4.9118	4.8937	4.5422	3.9267	3.3783	2.8603	3.0841	3.7196	4.0376	4.556	4.891	5.1031	5.1206	5.0083	4.9147
3.7531	4.3579	4.939	5.2848	5.418	5.6588	5.2669	4.9079	4.3247	3.8835	3.2672	3.4026	4.7426	5.215	5.3689
3.2898	4.0397	5.1709	5.3271	5.6865	5.9597	5.507	4.9808	4.6187	3.9188	3.1944	2.6486	5.4629	5.7914	6.0077
3.9951	4.0067	4.0990	4.1019	4.1545	4.2106	4.1233	4.1808	4.2057	4.1802	4.1012	4.0448	4.3596	4.3644	4.3651

$^4H_5$	^4E	$^4H_3$	E_3	$^2H_3$	^2E	$^2H_1$	E_1	$^OH_1$	$^4C_1$
3.5416	3.5163	3.3835	3.2835	2.8695	2.739	2.7047	3.0569	2.8842	3.2264
3.9285	3.9836	3.7912	3.6839	3.5365	3.3183	3.4211	3.1626	3.4649	3.2878
3.4285	3.0669	2.9915	2.9587	2.6361	2.6331	2.8313	2.9457	3.068	3.0884
3.4268	3.3192	3.2838	3.0542	3.4055	3.4577	3.535	3.3328	3.7815	3.0953
3.0281	2.6173	2.5259	2.8077	2.7408	2.7558	3.1819	3.1502	3.4252	2.9245
3.1822	3.3263	3.2402	3.3135	3.5312	3.7194	3.7032	3.7694	3.9069	3.2073
3.5504	3.3	3.6454	4.2716	4.5339	4.4819	4.4528	4.3991	4.3006	4.191
3.6139	3.8906	4.2861	4.3965	4.5651	4.6906	4.6965	4.5443	4.1958	4.2707
4.9167	4.8289	5.07	5.0604	4.9635	4.9611	5.1331	5.0708	5.0412	5.2348
5.2148	5.3766	5.3427	5.5086	5.4516	5.319	5.3375	5.1557	5.231	5.2192
5.1977	5.0113	4.9611	5.0857	5.043	5.0513	4.9342	4.9849	4.9594	5.1202
5.2836	5.2565	5.3688	5.2126	5.4521	5.4377	5.3086	5.1593	5.0229	5.2369
4.8179	4.6735	4.0747	3.5041	3.8856	4.5203	4.5763	5.2165	4.9531	4.5927
5.0134	4.6219	4.4021	4.2543	4.5479	4.6587	4.8969	5.3119	5.4042	5.1984
5.5043	5.3278	5.3516	5.4659	5.4888	5.4684	4.9424	5.2173	4.8785	5.6741
5.8681	5.8191	5.9938	5.7317	5.7486	5.6056	5.5828	5.4923	5.6634	6.1213
4.3448	4.2460	4.2320	4.2246	4.2750	4.3011	4.3274	4.3731	4.3863	4.3556

Table 7. HB interactions for each conformation expressed in Å. The nomenclature used is, e.g., HO1-HO2, which means the distance between the O of the C1 hydroxyl and the H of the C2 hydroxyl.

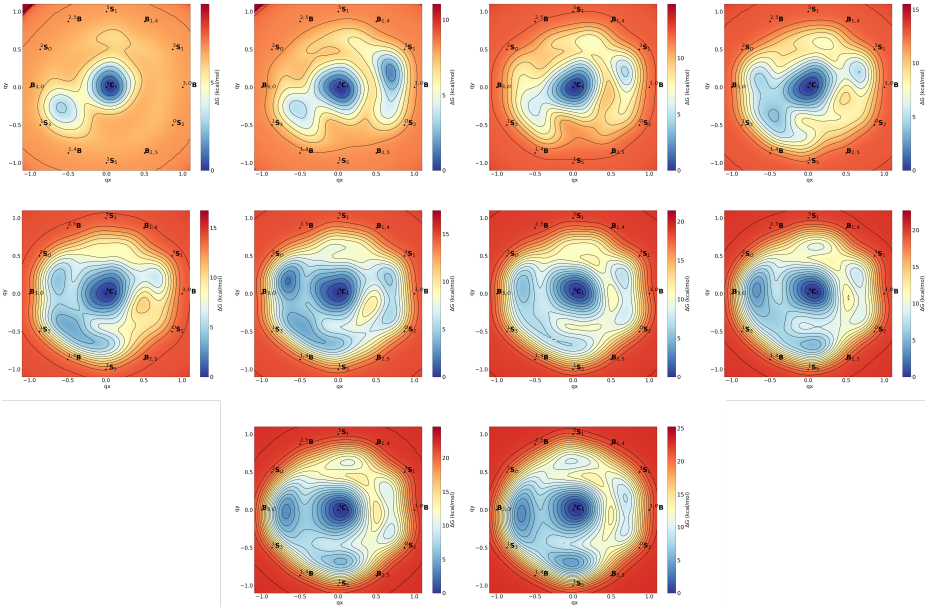

Figure 39. FEL evolution for the northern hemisphere using PBE (every 163 Gaussians).

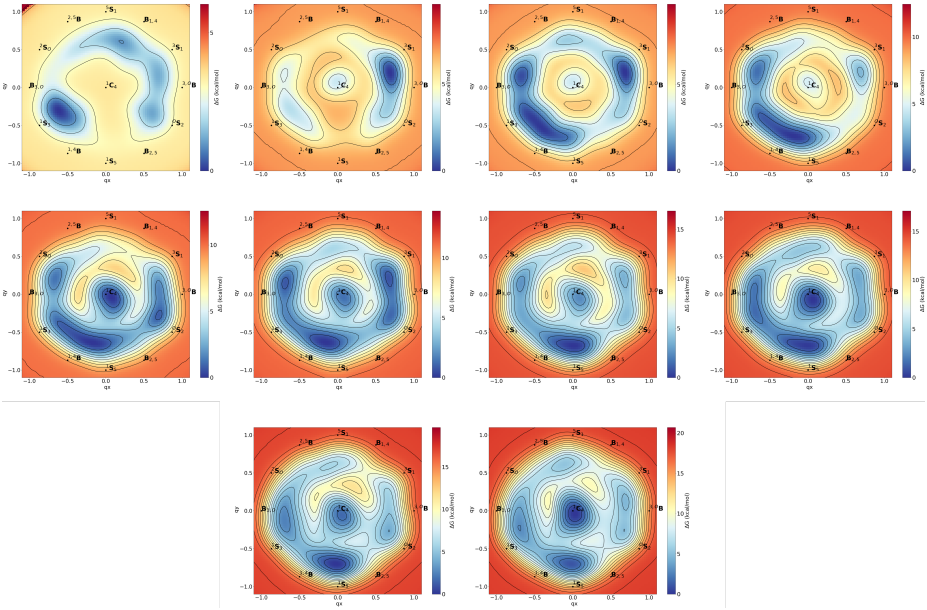

Figure 40. FEL evolution for the southern hemisphere using PBE(every 130 Gaussians).