

Master Erasmus Mundus in Theoretical Chemistry and Computational Modelling

Computational characterization of the photophysics of oxyluciferin within the luciferase environment

Henar Mateo de la Fuente

Acknowledgments

First of all, I would like to express my greatest gratitude to my supervisor, Juanjo, for his valuable feedback, and constant encouragement during the project. I would also like to thank all his awesome group for their support and kindness, making me feel like I belong since day one.

Además, me gustaría agradecer a mi familia por apoyarme siempre en mi etapa académica. Especialmente, a mi hermana por recordarme la importancia de los descansos en la productividad y a mi pareja por estar siempre a mi lado y apoyarme en todo momento.

Finally, I would like to thank Marco Garavelli for hosting me in Bologna, Davide Avagliano for his guidance during my stay there and in general to all his group “Spectroscopy and Computational Chemistry” for all the great moments.

The work has been performed under the Project HPC-EUROPA3 (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme; in particular, the author gratefully acknowledges the support of Marco Garavelli of the Department of Industrial Chemistry “T. Montanari” of the University of Bologna and the computer resources and technical support provided by CINECA. Moreover, the computational resources from the Centro de Computación Científica at the Universidad Autónoma de Madrid (CCC-UAM) are also acknowledged.

Abstract

Luciferin/luciferase bioluminescent systems are important in medicine since they can be used, among many other applications, to detect tumours. Therefore, improving its properties is crucial to increase its applicability, for example inducing a red-shift in the emission energy to improve the tissue penetration. To achieve this, it is important to develop a clear theoretical protocol able to accurately compute the photophysic properties of the system. As a first step for doing so, this work analyses not only the impact of the sampling process on the emission spectrum but also the effect of the employed potential energy model by means of classical and Quantum-Mechanics/Molecular-Mechanics (QM/MM) Molecular Dynamics (MD) simulations. It has been seen that the sampling method to select the snapshots from the dynamics, either random, equidistant or following a Metropolis criterion, did not affect the shape of the spectra and that the results coming from the classical MD are very similar to those obtained from the QM/MM MD simulation. Additionally, it has been observed that the emitting state, of $\pi\pi^*$ character, is mixed with a dark higher-energy state of $n\pi^*$ character along both types of trajectories. Moreover, for some of the sampled geometries these states are swapped and, therefore, the first absorption band and the emission band presented partial $n\pi^*$ character. However, these swapping is larger in the case of the QM/MM trajectories than in the classical ones. Furthermore, the dihedral angle distribution around the C-C single bond of the system has been studied. It has been observed that classical MD samples a narrower range of angles than QM/MM MD. This non-planar conformations are related with lower transition energies and larger charge transfer character, as confirmed by a rigid static scan of the torsion angle. However, these electronic behaviour with the dihedral angles was not observed in the spectra computed from the dynamics, indicating that other degrees of freedom counteract the effect of the dihedral angle.

Resumen

Los sistemas luciferina/luciferasa presentan bioluminiscencia, siendo importantes en medicina ya que pueden utilizarse, entre otras muchas aplicaciones, para detectar tumores. Por ello, la mejora de sus propiedades es crucial para aumentar su aplicabilidad. Por ejemplo, produciendo un desplazamiento al rojo de la energía de emisión para mejorar la penetración en los tejidos. Para conseguirlo, es importante desarrollar un protocolo computacional claro capaz de calcular de manera exacta las propiedades fotofísicas del sistema. Como primer paso para ello, este trabajo analiza no sólo el impacto del proceso de muestreo en el espectro de emisión, sino también el efecto del modelo de energía potencial empleado mediante simulaciones de dinámica molecular clásica y QM/MM. Se ha podido observar que el método de muestreo para seleccionar las geometrías de la dinámica, ya sea de manera aleatoria, esquispaciada o siguiendo el criterio de Metrópolis, no afecta a la forma de los espectros y que los resultados procedentes de la dinámica molecular clásica fueron muy similares a los obtenidos mediante la simulación QM/MM. Además, se ha visto que el estado emisor, de carácter $\pi\pi^*$, está mezclado con un estado oscuro de más alta energía y de carácter $n\pi^*$ a lo largo de ambos tipos de trayectorias. Además, en algunas de las geometrías muestreadas estos estados aparecen intercambiados y por tanto, la primera band de absorción y la de emisión presenta un carácter parcial $n\pi^*$. Sin embargo, este intercambio es mayor en el caso de las trayectoria QM/MM que en el caso de la trayectoria clásica. Por otro lado, se ha analizado la distribución de probabilidad de los ángulos diedro resultantes de la torsión alrededor de enlace simple C-C. Se ha podido observar que la dinámica molecular clásica muestrea un rango de ángulos dihedros más estrecho que la QM/MM MD. Estas conformaciones que rompen la planaridad del sistema están relacionadas con una menor energía de transición y un mayor carácter de transferencia de carga, como ha sido confirmado por medio de un escáner estático rígido del ángulo de torsión. Sin embargo, este comportamiento de las propiedades electrónicas con el ángulo diedro no se ha observado en los espectros calculados a partir de la dinámica, indicando que existen otros grados de libertad que contrarrestan el efecto del ángulo diedro.

Contents

Introduction	1
1. Bioluminescence	1
2. Effect of the Environment and the Sampling in Excited States	4
3. Objectives	8
Computational Methods	9
4. Nuclear Motion Simulation	9
4.1. Classical Molecular Dynamics	10
4.1.1. Interaction Model	10
Seminario Method	12
4.1.2. Integrator	14
Velocity Verlet Algorithm	14
4.1.3. Thermodynamic Ensemble	15
Langevin Thermostat	16
Monte Carlo Barostat	17
4.1.4. Periodic Boundary Conditions	17
Ewald Summation	19
4.1.5. Constrained Dynamics	21
4.2. Hybrid Quantum Mechanics/Molecular Mechanics	22
5. Density Functional Theory	24
5.1. Hohenberg-Kohn Theorems	24
5.2. Kohn-Sham Approach	24
5.3. Time-Dependent DFT	26
5.4. Exchange-Correlation Functionals	27
5.5. Basis Sets	29
5.5.1. Classification of Basis Sets	30
5.5.2. Contracted Basis Sets	30
Pople Basis Sets	31

6-31+G(d)	31
6-311G(2d,p)	31
Theoretical Procedure	32
6. Classical Molecular Dynamics Simulations	32
7. Hybrid QM/MM Molecular Dynamics Simulations	35
8. Potential Energy Surface Rigid Scan	36
Results	37
9. Effect of the Sampling	37
10. Effect of the Torsion around the C-C Single Bond	40
10.1. Static Scan Calculations around the C-C Single Bond	40
10.2. Classical Molecular Dynamics	41
10.3. QM/MM Molecular Dynamics	44
10.4. Quantum Mechanical (QM) vs Force Field (FF) Potentials	46
11. Classical MD vs QM/MM MD Based Spectra	47
11.1. QM/MM Spectra	47
11.2. Comparison of the MD Simulations	48
Conclusions	51
Outlook	53
References	55
Annexes	I
I. Derivations	I
I.I. Hamilton Equations	I

I.II. Velocity Verlet Algorithm	I
I.III. Instantaneous Temperature	II
II. Metropolis Criteria Code	III

List of Abbreviations

ADP Adenosine Diphosphate	MC Monte Carlo
ATP Adenosine Triphosphate	MD Molecular Dynamics
CCD Charge-coupled device	MM Molecular Mechanics
CGTO Contracted Gaussian Type Orbitals	MP2 Møller–Plesset second-order perturbation theory
CT Charge Transfer	PES Potential Energy Surface
DFT Density Functional Theory	PGTO Primitive Gaussian Type Orbitals
DOF Degree of freedom	PPi Phosphate
DZ Double Zeta	QM Quantum Mechanical
f oscillator strength	QM/MM Quantum-Mechanics/Molecular-Mechanics
FF Force Field	RMSD Root Mean Square Deviation
GAFF General Amber Force Field	S₀ electronic ground state
GGA Generalized Gradient Approximation	S₁ first electronically excited state
GTO Gaussian Type Orbitals	STO Slater Type Orbitals
HOMO Highest Occupied Molecular Orbital	TD Time Dependent
LDA Local Density Approximation	TZ Triple Zeta
LSDA Local Spin Density Approximation	ZPE Zero-Point Energy
LUMO Lowest Unoccupied Molecular Orbital	

Introduction

1. Bioluminescence

Nature presents a wide variety of bioluminescent systems resulting from a substrate/enzyme interaction, luciferin/luciferase, leading to a derivative of the substrate in electronically excited state that is able to emit light in the visible region (350-750 nm) during the relaxation process. Nowadays, nine natural luciferins have been documented together with their complementary luciferases giving rise to more than thirty bioluminescent systems. The first one of them to be discovered was D-luciferin, present in some species of insects, as fireflies, and first described in 1957 by Bitler and McElroy.^[1] That same year the luciferin of *cypridina* was also discovered by Shimoura who some years later, in 1968, described *latia* luciferin too. The bacterial luciferin was discovered in 1963 by Cormier; the coelenterazine and the *diplocardia* luciferins in 1976 by Ohtshuka; and the krill and dinoflagellates luciferins in 1989 by Nakamura. Already in the XXI century, *fridericia heliota* (2014) and fungi (2015) luciferins were described by Petushkov and Purtov, respectively, both in collaboration with Yampolsky.^[2]

This work focuses on the effect of different computational factors on the emission spectrum of the D-luciferin/luciferase system, whose emitting specie is oxyluciferin, a chromophore obtained in excited state during the bioluminescence reaction shown in Figure 1. The mechanism starts with (i) the reversible adenylation reaction of D-luciferin with Adenosine Triphosphate (ATP) facilitated by a magnesium cation via a S_N2 nucleophilic displacement.^[3-5] Then, the luciferyl adenilate intermediate can react with oxygen through two reaction paths: (a) the bright path, represented in black and with a 80% yield and (b) the dark reaction path, represented in dark orange and with a 20% of yield. In the latter, dehydroluciferyl adenilate, a very powerful tight-binding inhibitor, and hydrogen peroxide are first produced (ii.b).^[6] In a consecutive step, (ii.b) the inhibitor is able to react with the previously released Phosphate (PPi) producing dehydroluciferin and ATP, which favours the bioluminescent path since it is not as strong inhibitor as dehydroluciferyl adenilate and, thus, can be more easily removed from active site of luciferase. Therefore, the catalysing activity of the protein is recovered.^[7] In the bright reaction path, the increased acidity on the enantiomeric carbon atom (C4), the one bonded to the explicitly represented hydrogen, due to the presence of the Adenosine Diphosphate (ADP) group^[7], allows for the required deprotonation^[8] before (ii.a) the oxidation reaction.

This reaction follows a single-electron transfer mechanism and results in a peroxide^[8], which can (iii.a) further react via internal nucleophilic attack of the α -hydroperoxy group to the electrophilic carbon, releasing the ADP fragment and producing a dioxetanone ring, an energy-rich moiety.^[7] Consecutively, (iv.a) CO_2 is released in an asynchronous two steps process leading to an electronically excited-state carbonyl compound.^[9, 10] Finally, the oxyluciferin keto anion in excited state, that is stabilised by means of an ion pair formation between the thiazole fragment and the residue Lys531 of the protein,^[11] is produce (v.a) via excited state proton transfer.^[10] At this point, the molecule (vi.a) is radiatively relaxed and light is emitted.

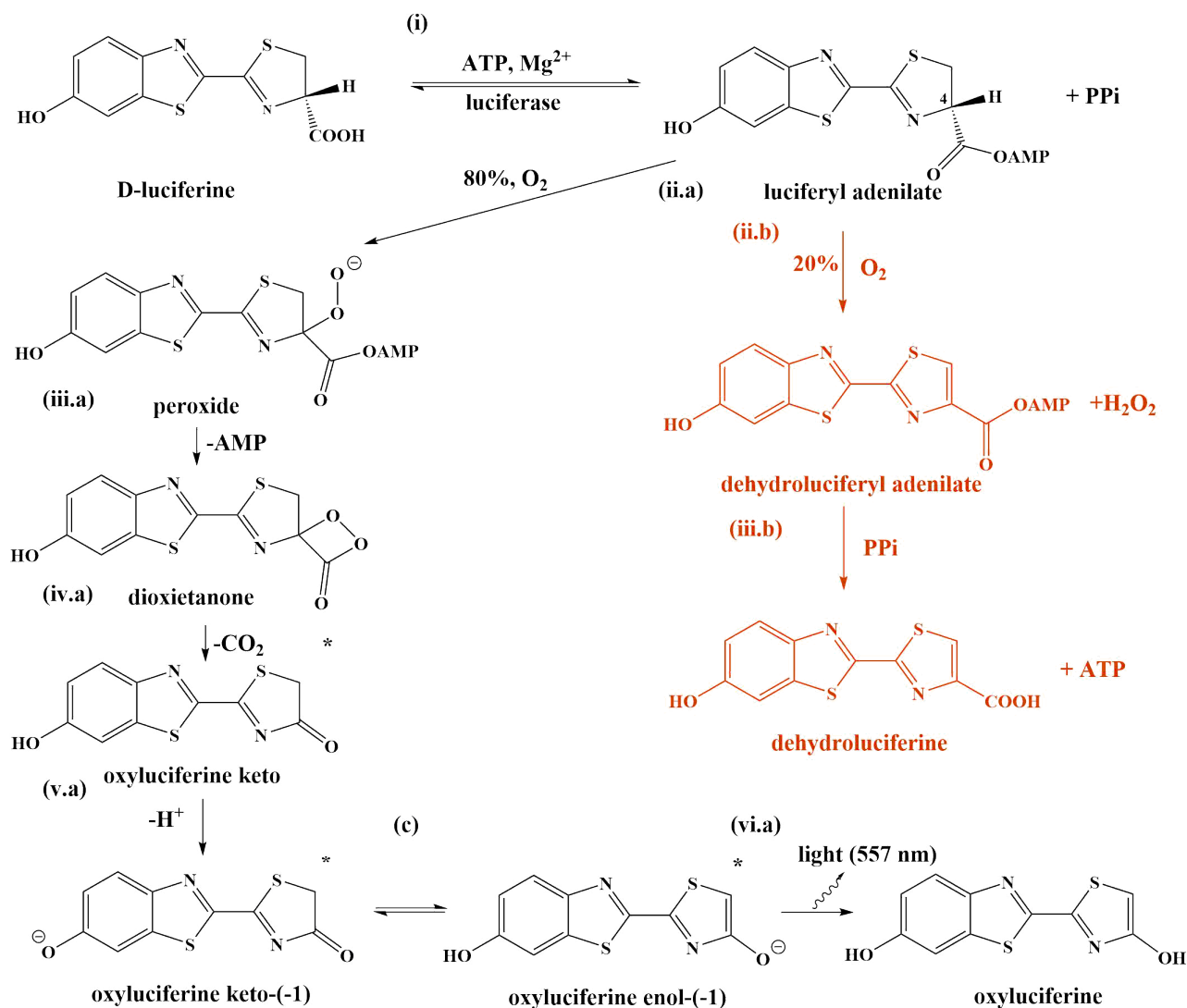


Figure 1: D-luciferin bioluminescence mechanism.

During this reaction, more specifically after the adenilation step, the luciferase protein suffers a conformational change, where a rotation of 140° of the C-terminal domain can be observed, after which a cavity that allows the diffusion of the oxygen molecule to the active site is formed.^[11] ^[12] In this new structure the Gly446 residue lines the cavity, while the His245 further expands

it for the oxygen molecule to access the C4 via side chain rotation.^[8] This conformational change could explain the difference in yield between the two reaction paths since the first step for the bright reaction, the deprotonation, does not require the presence of O₂ and therefore, the intermediate should already be present when oxygen reaches the active site, while the dark reaction path can only take place after the conformational change is completed and oxygen diffuses into the active site. Moreover, although the enol form of the oxyluciferin anion is thermodynamically more stable,^[13] it has been shown that the emitting specie is the keto tautomer because it is the directly produced molecule and the activation energy of the excited state proton transfer reaction that should take place to obtain the enol form is very high.^[11]

D-luciferin bioluminescent processes take places in several different species of animals. However, the luciferase enzymes that they produce present some structural differences that induce a shift in the emission wavelength. Attending tho this, it is possible to distinguish three big groups of natural luciferases.^[2] The most known luciferase is the one present in (*Lampyridae* family), which emits between 540 and 570 nm and has been used as fluorescence marker to study the growth and the metastasis of renal carcinoma in mice *in vivo* and to develop anti-carcinogenic drugs *in vitro*. For this last application, together with the study of infectious illnesses, luciferases produced by click beetle (*Elateridae* family), whose emission lies between 540 and 600 nm, are used. The third family of luciferases is the *Phengodidae* one, to which railroad worms belong. This species is characterise by the expression of two different luciferases in different areas of their bodies, which allows them to emit in two different colours: red (630 nm) and green-yellow (540 nm).

The D-luciferin/luciferase system presents a wide variety of applications additionally to the ones already mentioned. This system has been typically used to determine the intracellular concentration of ATP, which is the molecule that activates the bioluminiscent pathway in natural systems. Knowing the amount of ATP, mainly synthesised in mitochondria, allows both the evaluation of the metabolic potential of a cell and the study of the metabolic behaviour of tumours, directly associated with mitochondria dysfunction.^[14] ^[15] However, the most important application of bioluminescence is found in in bioimaging, a technique that is able to obtain images of life tissues and intracellular processes in a non invasive way by means of the heterogeneous expression of the enzyme within the organism. In this way, only a specific group of cells is able to synthesise the enzyme. This technique presents a high sensitivity to bioluminescence, absence of auto fluorescence and low background noise, which allows its applicability to *in vitro* and *in vivo* studies with luminometers or Charge-coupled device (CCD) cameras

with higher sensitivities than devices based on fluorescent proteins. Bioimaging applications focus on the analysis of reporter genes; tracking of neurodegenerative, tumour and infectious cells; and the study of protein-protein interactions.^[16]

2. Effect of the Environment and the Sampling in Excited States

The study of absorption or emission spectra of a specific chromophore requires to work not only with the electronic ground state of the system but also with its electronically excited states. Moreover, in the case of complex biological systems, like the one under study here, and due to their anisotropic character, the environment has to be explicitly described and conformationally sampled. Both the description of the environment and the sampling of the phase space of the system, space of coordinates and momenta, strongly affect this kind of calculations since they can shift the maxima of the spectra and modify the electronic properties of the transitions.

The electronic transitions can be of different nature, as represented in Figure 2. Local transitions correspond to those in which the initial and final positions of the excited electron are located in the same region of the molecule. In the case of Charge Transfer (CT) transitions, initial and final positions of the excited electron are located in a different fragment of the molecule (intramolecular CT transition) or in different molecules (intermolecular CT transition). In these transitions the dipole moment of the chromophore is strongly affected and, therefore, the presence of a polar environment can have a big influence.^[17] To account for CT transitions, it is mandatory to include all the involved fragments in the quantum mechanical calculation.

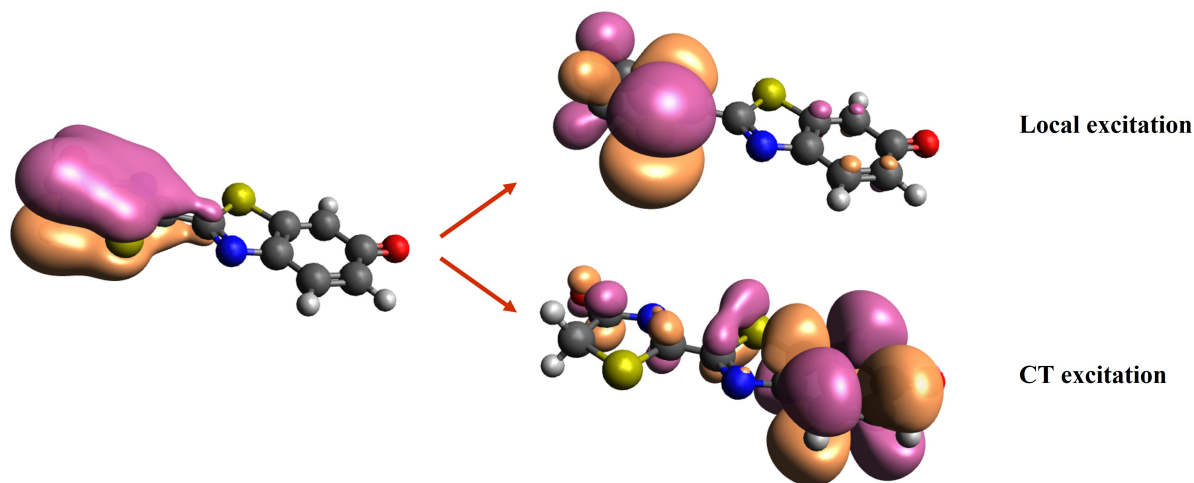


Figure 2: Schematic representation of different possible transitions.

Electrostatic effect and electronic coupling are the main contributions to the energy shifts with regard to the environmental description.^[18] The former is produced due to the charge distribution surrounding the chromophore and results in either, blue- or red-shifting depending if they produce stabilization or destabilization of the energy levels of the system with respect to the vacuum situation, as it is represented in Figure 3(a). An example of this would be the stabilization of n type orbitals, that contain the lone pair of heteroatoms, due to the existence of hydrogen bonding interactions with molecules from the environment. In this cases, $n\pi^*$ transitions experiment a blue-shift when an explicit description of the hydrogen bonding is considered. However, this effect would only be observed in the cases in which the strength of the bonding in the ground state is significantly different to that of the excited state since they would cancel out in the vertical energy calculation.^[17] When electronic coupling, caused by the interaction between the transition dipole momenta of the chromophore and environment, takes place, Frenkel exciton states are generated. They are electronic states for which the excited electron and/or the hole created after the excitation are delocalized not only on the chromophore but also on the parts of the system coupled with it, for example, the solvent or the biological environment. This effect also appears in multichromophoric system and results in the split of degenerated excited states (see Figure 3(b)).^[19, 20] Depending on the symmetry of this interactions, the transition of the lowered or risen states can be allowed or forbidden and, as a consequence, the spectrum can be either blue- or red-shifted. In addition, it can also be split if the transition to more than one of the previously degenerated states is allowed. Electronic coupling becomes relevant when the interaction between transition dipole momenta is strong, thus generating delocalized exciton states. Moreover, local and exciton transitions can mix with CT states leading to new electronic states with different transition energies and properties. In

the case of the electrostatic effect, and within the framework of classical Molecular Dynamics (MD), the best way to account for it would be by means of a polarizable embedding scheme, briefly described in Section 4.2. However, electronic coupling would require to go beyond classical mechanics by applying a quantum treatment of the whole region in which the exciton is localized.

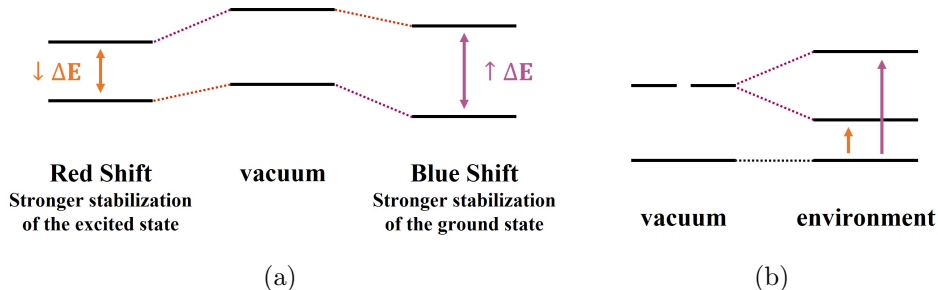


Figure 3: Schematic representation of energy level stabilization or destabilization due to different effects induced by the presence of the environment. (a) Electrostatic effect and (b) electronic coupling.

Small systems, such as diatomic molecules, present a very simple Potential Energy Surface (PES). However, biological and solvent environments, with thousands of atoms, have a huge amount of degrees of freedom and, therefore, a very complex PES, as depicted in Figure 4(a). Thus, the consideration of a unique conformation, as it is usually done for small systems in vacuum or homogeneous solvent, results in a very poor representation of the system. Instead, an ensemble of geometries needs to be sampled from the conformational space to properly represent the Frank-Condon region, around the PES minimum of the electronic state. This sampling can be either quantum, generated from a Wigner distribution, or thermal, obtained from a MD simulation assuming thermal equilibrium at a given temperature. In addition, a combination of both for different parts of the system is also possible, for example, quantum sampling for the chromophore and thermal sampling for the environment.^[17] Each of the sampling strategies presents some benefits and drawbacks. Quantum sampling accounts for the thermal energy and the Zero-Point Energy (ZPE) of each Degree of freedom (DOF). However, its computational feasibility is strongly reduced when the number of atoms is increased since a frequency calculation is required. Moreover, vibrational modes with strong anharmonicity are not properly described because it is often based on the harmonic approximation. In the case of thermal sampling, the ZPE is understimted since each vibrational DOF has an energy of only $k_B T$, but it is able to reproduce low-frequency motions with high anharmonicity and to sample the space of much bigger systems.

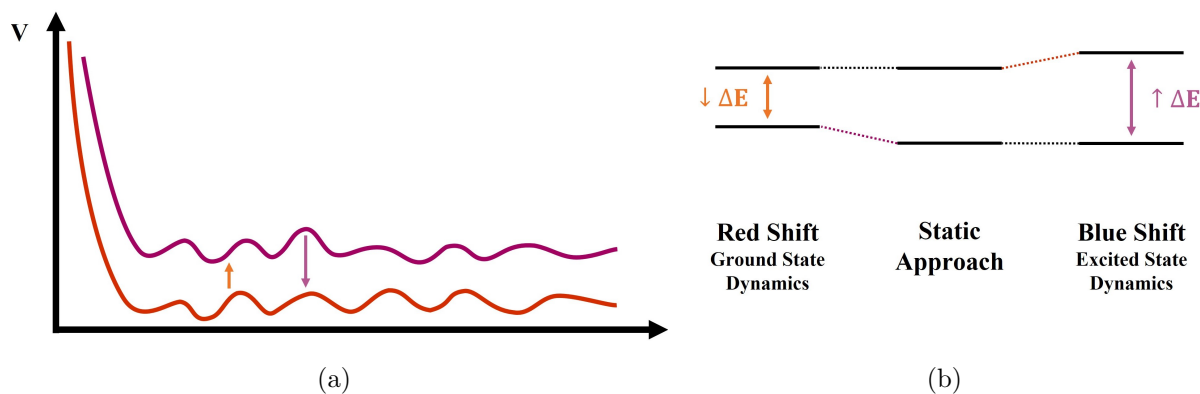


Figure 4: (a) Schematic representation of the PESs of the ground and excited states of a complex system and (b) schematic representation of the sampling effect of the ground and an excited state on the excitation energy with respect to an static approach.

Taking into account the vibrational movement of the system by means of the sampling of the conformational phase can also produce a shift on the excitation energies with respect to static approaches.^[17] During the ground state simulation, not all the obtained geometries correspond to a minimum in the PES and therefore, in average, the ground state of the system is destabilized while the energy of the excited state remains almost unaltered. Hence a red-shift of the energy is usually observed, as represented in Figure 4(b). In the case of the simulation of the excited state, the opposite effect happens: a blue-shift of the vertical transition from the excited state to the ground state due to the averaged destabilization of the excited state. Moreover, CT and delocalization overestimation during static calculations tends to be corrected due to the breaking of the symmetry and variation of intermolecular interactions, respectively, along the simulation.

3. Objectives

The general objective of the project is the computational design of an efficient luciferin/luciferase system with enhanced photophysical properties, including a lower emission energy to increase the tissue penetration and a higher intensity of emission to ease the detection. This work represents the first step of the project, where the effect of different computational factors on the absorption and emission spectra of the oxyluciferin/luciferase system present in fireflies is investigated. The specific goals are the following ones:

1. To analyze the impact of the sampling process on the absorption and emission spectra considering three different approaches for choosing the chromophore/enzyme snapshots from a MD trajectory, on top of which the excited-state computations are performed. These approaches are random selection, equidistant selection, and a Metropolis approach.
2. To analyze the effect of the potential energy model used during the sampling on the spectra by running classical and QM/MM MD simulations to explore the PES of the electronic ground state (S_0) and the first electronically excited state (S_1).
3. To analyze the impact of geometrical features on the excitation energies and electronic properties to obtain a more clear description of the system. In addition, this analysis is performed within both static and dynamic frameworks.

Computational Methods

As previously commented, the main goal of the work is to compute the absorption and emission spectra of the oxyluciferin chromophore in the luciferase biological environment and to analyse the different factors that affect the electronic transitions. For this purpose, MD simulations both classical and a hybrid QM/MM scheme have been used to sample the configurational space. This sampling is based on the ergodic hypothesis that states that the probability of finding the system in some region of the configurational space that presents the same energy over a long period of time is proportional to the volume of this region and therefore that all accessible configurations are equiprobable over a long time period. In addition, electronic structure methods are employed to calculate electronic transitions. This section thoroughly describes these computational methods, together with some other features required to perform them.

4. Nuclear Motion Simulation

Biological systems, such as the one under study here, present many possible stable geometrical configurations due to the big amount of DOF. Hence, to properly describe the properties of the system, it is important to sample properly the configurational space, within a specific ensemble. Since a full Quantum Mechanical (QM) description of the motion would be unfeasible, to obtain it, there exist two classical approaches: stochastic simulations, covered by Monte Carlo (MC) methods, [21] and deterministic simulations, given by MD method. [22] The former, samples the conformational space using trial geometries that are accepted or not by means of a Metropolis algorithm and no time propagation is obtained, [23] whereas the latter obtains a time evolution of the system by integrating Hamilton's equations of motion (see Equation 1), easily derived from Newton's second law and the definition of the kinetic energy (see derivation in Annex I.I.).

$$\frac{d\mathbf{p}_i}{dt} = -\frac{dV}{d\mathbf{r}_i} \quad \frac{dT_i}{d\mathbf{p}_i} = \frac{d\mathbf{r}_i}{dt} \quad (1)$$

During this work MD simulations were performed and, therefore, this methodology will be explained below. Additionally, within this section, the most important approaches needed to evolve the MD simulations will be explained: Force Field (FF) and its reparameterization by means of the Seminario method, [24] the Langevin thermostat [25, 26] and the MC barostat [27]

and QM calculations performed on top of some of the configurations of the trajectory.

4.1. Classical Molecular Dynamics

The Born-Oppenheimer approximation states that the electrons immediately adapt to the movement of the nuclei but that the nuclear position are unaffected by the electronic movement due to the huge difference in the mass of the particles. Under this approach, it is possible to uncouple the electronic and nuclear motions and so to solve independently each of the problems. In the case of classical MD, the latter is described by means of the classical equations of motion, being the nuclei treated as hard spheres connected by strings.[\[28\]](#) This method requires of three different ingredients to be applied: an interaction model to describe the potential energy of the system, an integrator to propagate particle positions and velocities in time, and a selected thermodynamic ensemble to determine the thermodynamic quantities that need to be controlled during the simulation.[\[22\]](#)

4.1.1. Interaction Model

The potential energy of the system can be approximated as sum of analytical and empirical functions. The collection of these functions is called FF. Usually, the functions are divided in bonding and non-bonding terms[\[29\]](#) and each of them tries to describe a certain kind of atomic movement or pair wise interactions. Figure [5](#) represents the interactions associated to each function and their most typical formulations.

$$U^{total} = U^{bonding} + U^{non-bonding} \quad (2)$$

Withing bonding interactions, there are three main types: bond stretching, bond angle bending and dihedral angle interactions. The first one represents the elongation and contraction of the interatomic distance between two covalently bonded atoms. The second one describes the angle variation between three consecutively bonded atoms. They are, respectively, 1-2 and 1-3 interactions. The last one, the 1-4 interactions, can be subdivided in proper dihedrals, which describe the rotation around the BC bond (angle formed by the planes ABC and BCD , see Figure [5](#)), and improper dihedrals that describe the variation of the angle formed by the planes ABC and ACD in Figure [5](#). The latter are used to ensure planarity and, sometimes chirality of some groups. The most common function to describe the bond stretching, the bond

angle bending and the improper dihedral is an harmonic potential, a quadratic function that reproduces oscillations around an equilibrium value (r_0 , ϑ_0 or ξ_0) with a certain frequency, associated to the force constant (k_{AB}^{qs} , k_ϑ or k_ξ). However, this formulation is only able to describe the behaviour of the system close to equilibrium and it is not capable of describing bond breaking and formation. Hence, other functions like the Morse potentials can be used to obtain a better description at regions of the PES far from equilibrium. In the case of the proper dihedral, the description uses a periodic function where k_φ is related to the height of the energy barrier; n to the number of minima in the energy function and φ_0 is the phase factor associated to the position of the minima. This formulation can also be generalized to the description of all the dihedral angles, like it is done in the Amber20 software.[\[30\]](#) One of the problems of splitting the bonding terms in different contributions is that, in real systems, they are usually coupled. To avoid this problem, it is possible to describe crossed terms like in the Urey-Bradley potential, where the variation in energy upon angle bending depends on the distance between atoms in positions 1–3.

Non-bonding terms represent the interactions between atoms separated by more than two bonds or not bonded at all. They are divided in electrostatic and van der Waals interactions. The former comes from the interaction between two-point charges (q), expressed as a Coulombic potential. The latter, typically described by a Lennard-Jones 12-6 potential, comprises a repulsive term based on Pauli’s exclusion principle and an attractive term that arises from dispersion forces between dipoles, either permanent or induced. This expression includes the minimum of the potential well (ϵ) and interatomic distance at which the potential is 0 (σ).

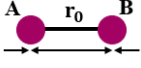
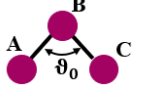
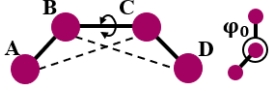
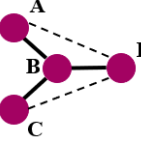
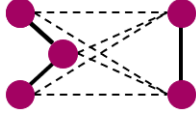
Quadratic streatching Morse streatching		$U_{AB}^{qs}(r_{AB}) = \frac{1}{2}k_{AB}^{qs}(r_{AB} - r_0)^2$ $U_{AB}^{ms}(r_{AB}) = \frac{1}{2}k_{AB}^{ms}(1 - e^{a(r_{AB}-r_0)})^2$
Bond angle bending		$U_{ABC}^b(\vartheta_{ABC}) = \frac{1}{2}k_\vartheta(\vartheta_{ABC} - \vartheta_0)^2$
Proper dihedral		$U_{A-D}^{pd}(\varphi_{A-D}) = \frac{1}{2}k_\varphi(1 + \cos(n\varphi_{A-D} - \varphi_0))$
Improper dihedral		$U_{A-D}^{id}(\xi_{AB}) = \frac{1}{2}k_\xi^{qs}(\xi_{A-D} - \xi_0)^2$
Electrostatic 6-12 van der Waals		$U_{AB}(r_{AB}) = \frac{q_A q_B}{r_{AB}}$ $U_{AB}^{6-12} = 4\varepsilon_{AB} \left(\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 \right)$

Figure 5: Graphic representation of the main kinds of interactions employed in FFs for simplicity together with their main analytical functions. In the upper part of the table the bonding terms and, in the lower, the non-bonding terms.

To apply a FF. it is important to determine the force an equilibrium constants that characterise the interactions. In practice, they are assumed to be transferable, which means that the functions developed on a small set of molecules can be applied to a larger group of them of similar characteristics, similar functional groups. Therefore, each atom of the system can be classified into a specific atom type depending on its surrounding and, for each atom type, these constants are parameterized.

Seminario Method

FFs are usually developed for the S_0 . However, this work aims to reproduce not only the PES of the ground state but also that of the S_1 . To do so, it is possible to parameterize the equilibrium and force constants values of a molecule by means of a QM calculations. The Seminario Method [\[24\]](#) is one of the possible procedures to obtain them.

The equilibrium bond distances and angles can be directly taken from an optimised geometry computed with the preferred software. In the case of the force constant values, they can be obtained from the Hessian tensor $[k]$ of the system:

$$[k] = k_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j} \quad (3)$$

The Hessian, exactly related (up to second order on a Taylor series expansion) to the $3N$ component reaction force ΔF due to a small displacement Δr of the N atoms in a molecular system expressed as Equation 4, represents the intramolecular FF of the system and so, its eigenvalues λ_i are the $3N$ force constants (3 translational, 3 rotational and $3N - 6$ vibrational modes) and its eigenvectors ν_i , the directions of the displacements of the corresponding normal modes.

$$\Delta F = -[k]\Delta r \quad (4)$$

In the case of the bond force constants, the previous equation can be simplified to a pairwise interaction in which the reaction force on an atom A (ΔF_A) is expressed as a function displacement of an atom B (Δr_B) (see Equation 5 where the minus sign is comprised in the tensor). Therefore, the tensor is a 3x3 matrix that has 3 eigenvalues and eigenvectors.

$$\Delta F_A = [k_{AB}] \Delta r_B \quad (5)$$

If one or more eigenvalues or two of three eigenvectors are complex, then A and B are not bonded since the pair wise interaction is unstable. The force constant can have contributions from more than one eigenvalue λ of $[k_{AB}]$ when any of the eigenvectors coincide with the direction from A to B and so, it can be expressed as a projection of each of the eigenvectors (ν^{AB}) in the unitary vector from A to B (u^{AB}):

$$k_{AB} = \sum_{i=1}^3 \lambda_i^{AB} |u^{AB} \cdot \nu_i^{AB}| \quad (6)$$

In the case of of bond angles, the force constant is defined as:

$$\frac{1}{k_\vartheta} = \frac{1}{r_{AB}^2 \sum_{i=1}^3 \lambda_i^{AB} |u^{PA} \cdot \nu_i^{AB}|} + \frac{1}{r_{CB}^2 \sum_{i=1}^3 \lambda_i^{CB} |u^{PC} \cdot \nu_i^{CB}|} \quad (7)$$

where r_{AB} and r_{CB} represent the bond lengths and u^{PA} and u^{PC} , the projection of the unitary vectors in the direction of the bond onto a unitary vector u_N perpendicular to the bonds AB and CB :

$$u^{PA} = u_N \times u^{AB} \quad u^{PC} = u_N \times u^{CB} \quad (8)$$

$$u_N = \frac{u^{CB} \times u^{AB}}{|u^{CB} \times u^{AB}|} \quad (9)$$

For both cases the order in which the atoms are taken does not affect the computation of the force constant since, taking into account the third law of Newton, a pair of forces acting on the two interacting objects are the same in modulus and direction but opposite senses. A similar formulation for the reparametrization of the dihedral angles can also be described, taking into account the projection of the bond unitary vectors on the unitary vectors perpendicular to the ABC and BCD planes. However, dihedral angles are usually not parameterized since bond and angle potentials contribute more to the potential energy and are, in principle, more relevant.

4.1.2. Integrator

The integrator is a key ingredient for the implementation of the MD method since it allows to propagate the positions and velocities of all the atoms of the system in time. It numerically solves the Hamilton equations of motion (Equation [1](#)) and must meet some requirements: (i) accuracy so that the resulting trajectory corresponds to a good approximation of the real trajectory; (ii) stability in the sense that energy is conserved and that small perturbations of the system do not lead to instabilities; (iii) and robustness, that allows the use of large time steps in the propagation.

There exist several types of integrators: the expansion based methods, constructed by expanding the positions and velocities in a Taylor series; the operator splitting methods, based on the phase space description of a classical system and that additionally allow to split the propagator into several time scales; and the multiple time step methods, that allows to decompose the time scales of the system depending on the speed of the motion of the DOF. [\[22\]](#) All the MD simulations performed during this work used the Velocity Verlet algorithm, [\[31\]](#) an expansion based integrator that is explained thereupon.

Velocity Verlet Algorithm

The velocity Verlet algorithm (see derivation in Annex [I.II.](#)) allows to obtain the coordinates (\mathbf{r}_i) and the velocities (\mathbf{v}_i) at a certain time step from the positions and velocities of the previous step and from the acceleration of previous and current time steps:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 \quad (10)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{1}{2}(\mathbf{a}_i(t) + \mathbf{a}_i(t + \Delta t))\Delta t \quad (11)$$

Since the information of the acceleration (\mathbf{a}_i) of the current step is needed, the computation of the velocity is performed in two steps. First, the first two terms of Equation [11](#) are calculated since they only require the known information of the previous time step and then, after the new coordinates are obtained and the acceleration recomputed, the third term is added. Accelerations are calculated by means of Newton's second law:

$$\mathbf{a}_i = \frac{\mathbf{F}_i}{m_i} \quad (12)$$

where the forces acting on each atom (\mathbf{F}_i) are determined from the potential of the system assuming it is conservative (only depends of the positions of the atoms):

$$\mathbf{F}_i(\mathbf{r}_1, \dots, \mathbf{r}_N) = -\frac{\partial U(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \quad (13)$$

At the beginning of the simulation, the coordinates are taken either from experimental results as in the case of proteins or previous theoretical calculations and the velocities from a Maxwell-Boltzmann distribution (see Equation [14](#)) at a certain temperature, supposing that the system is in thermal equilibrium.

$$P(\mathbf{v}_i) = 4\pi \left(\frac{m_i \beta}{2\pi} \right)^{\frac{3}{2}} \mathbf{v}_i^2 e^{-\frac{m_i \mathbf{v}_i^2 \beta}{2}} \quad (14)$$

4.1.3. Thermodynamic Ensemble

A thermodynamic ensemble is the assembly of all possible microstates, all the possible states consistent with the thermodynamic constraints of the system.[\[32\]](#) *A priori*, the selection of one ensemble or another should provide the same results. However, different steps of the MD simulation might benefit from one specific ensemble as they provide a control mechanism to ensure a proper description of the system. For example, during a simulation in which the temperature of the system is increased, it is beneficial to use an ensemble that maintains constant the volume to avoid an abnormal expansion. On the other hand, using an ensemble that fixes temperature and pressure allows a control of the density along the simulation since the volume is allowed to change.

There exist several types of control mechanisms that can be applied to the measurable properties of the system: (i) differential control, in which the thermodynamic quantity is fixed and no fluctuations are allowed; (ii) proportional control, in which some variables are coupled to a thermodynamic property and thus, it is corrected along the simulation by means of a coupling

constant that determines the strength of the fluctuations of the property; (iii) integral control, in which an extended Hamiltonian where variables that represent the effect of an external system are introduced to fix the state to the desired ensemble; and (iv) stochastic control, in which modified equations of motion are used to propagate the values of the variables coupled to the thermodynamic property. [22] Usually, temperature and pressure are the properties that are controlled by the so-called thermostats and barostats. However, it is also possible to introduce a constraint to induce energy conservation, as in the microcanonical ensemble (NVE). Thereupon, the used thermostat and barostat will be explained, although there exist many others.

Langevin Thermostat

The Langevin thermostat [25, 26] is an stochastic thermostat that is used to control that the simulation runs within a canonical ensemble, an ensemble in which the number of particles, the volume and the temperature are constant (NVT). This model is based on collisions between virtual particles and the DOFs of the system. To do so, Gaussian friction random forces ($\mathbf{F}_{\text{rand}}(\mathbf{t})$) with zero mean, taken from Stokes' law, are applied on each atom and therefore, Equation [13] is modified to:

$$\mathbf{F}_i(\mathbf{r}_1, \dots, \mathbf{r}_N) = -\frac{\partial U(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_i} - \gamma m_i \mathbf{v}_i(t) + \mathbf{F}_{\text{rand}}(\mathbf{t}) \quad (15)$$

where γ represents a friction constant parameter and the amplitude of the random force is determined by the second fluctuation dissipation theorem:

$$\langle \mathbf{F}_{\text{rand},i}(\mathbf{t}) \mathbf{F}_{\text{rand},j}(\mathbf{t}') \rangle = 2m_i \gamma k_B T_0 \delta_{ij} \delta(t - t') \quad (16)$$

In this equation k_B represents Boltzmann's constant and T_0 , the objective temperature. The collision frequency, related to γ , should not be too high to avoid a big disturbance of the phase space trajectory and a strong loss of the memory of the particle and thus, a fast decay of dynamic correlation functions.

Additionally, the instantaneous temperature (T) of the system can be calculated along the simulation through the definition of the averaged translational kinetic energy obtained from a Maxwell-Boltzmann distribution [33] (see derivation in Annex [I.III]):

$$T = \frac{\sum_{i=1}^N m_i \cdot \mathbf{v}_i^2}{k_B N_{DOF}} \quad (17)$$

being N the number of atoms in the system; k_B , the Boltzmann's constant and N_{DOF} , the number of DOF.

Monte Carlo Barostat

In the case of requiring an additional control of the pressure (P), like in the isothermal-isobaric ensemble (NTP), the use of a barostat is mandatory. Within this work, the MC[27] barostat has been used. The control of the pressure is obtained stochastically by means of a trial volume change:

$$\Delta V = n_{rand} \Delta V_{max} \quad (18)$$

where n_{rand} represents a random number between -1 and 1 and ΔV_{max} is the volume variation limit chosen so that a typical MC acceptance ratio of about 40–50% is achieved. Then, the boxlengths (l_i) and center of mass coordinates (\mathbf{r}_i) of each molecule are rescaled according to:

$$l_i \longrightarrow l_i \sqrt[3]{\frac{V'}{V}} \quad \mathbf{r}_i \longrightarrow (\mathbf{r}_i - \mathbf{c}_i) \sqrt[3]{\frac{V'}{V}} + \mathbf{c}_i \quad (19)$$

being $V' = V + \Delta V$ and \mathbf{c}_i the coordinates of the center of the periodic box. This new configuration is accepted with the standard MC probability of $e^{-\beta \Delta W}$ if the Metropolis weight function, given by Equation [20] where P_0 represents the target pressure, is larger than 0 and always accepted otherwise.

$$\Delta W = U' - U + P_0 \Delta V - N k_B T \ln \frac{V'}{V} \quad (20)$$

4.1.4. Periodic Boundary Conditions

MD simulations are preformed for a relatively small number of atoms, when compared to the usual amount of them in real systems, which modifies the ratio of atoms on the surface of the system and, therefore, in contact with vacuum. To overcome this problem, periodic boundary conditions are applied to the simulation. Within this model, the computed system[34] is replicated in space along the three dimensions (see a two dimensional scheme in Figure [6]), which allows to account for the long-range non-bonding interactions while the equations of motion are solved only for the atoms in the primary cell.[34] Each of the replicas, called virtual cells, contains exactly the same amount of atoms as the primary cell and are obtained by adding to the coordinates of the system a collection of vectors (\mathbf{n}) expressed in terms of the

axis direction ($\mathbf{L}_x, \mathbf{L}_y, \mathbf{L}_z$), whose length is the size of the box (L). Moreover, for each of the virtual atoms, momenta are the same as in the primary cell and therefore behave in the same way:

$$\mathbf{r}_i \longrightarrow \mathbf{r}_i + \mathbf{n} \quad \mathbf{p}_i \longrightarrow \mathbf{p}_i \quad (21)$$

The collection of \mathbf{n} vectors is given by Equation 22 where a , b and c can have the values of -1, 0 and 1 and n represents the number of times in which the primary cell is replicated in a specific direction.

$$\mathbf{n} \longrightarrow n(a\mathbf{L}_x + b\mathbf{L}_y + c\mathbf{L}_z) \quad (22)$$

Additionally, if one atom leaves the primary box on one side during the simulation, another one with the same momenta enters it on the opposite side to ensure that the number of atoms is conserved. This description of the virtual cells saves a lot of memory in the calculation since they can be very easily computed from the coordinates and momenta of the primary cell.

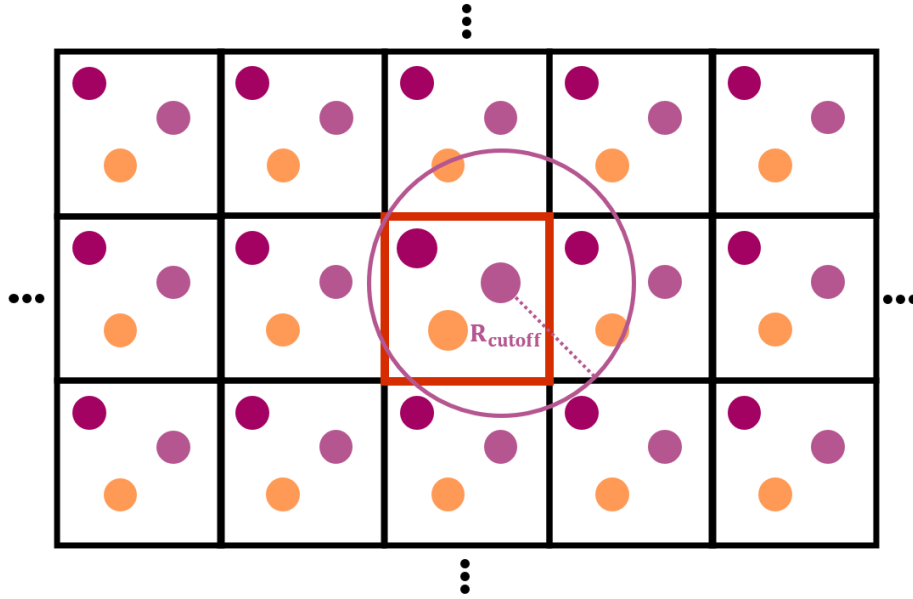


Figure 6: Schematic 2D representation of the replication of the primary cell (dark orange) in virtual cells (black) performed when periodic boundary conditions are applied to the MD simulation.

In order to compute the interactions between the atoms, a cutoff (R_{cutoff}), inside which the non-bonding interactions of each atom are calculated, is used to save computational time. However, this approach presents the problem of the energy continuity with the distance. In the case of van der Waals interactions, since they do not present a strong long-range character and so they decay to 0 very fast as shown in Figure 7, the problem can be solved by multiplying the potential by a switching function ($S(r)$, represented in orange dashed line) that gradually

reduces it to zero over a range of a few angstroms (R):

$$S(r) = \begin{cases} 1 & r_{ij} \leq R_0 \\ \frac{(R_{cutoff}^2 - r_{ij}^2)^2 (R_{cutoff}^2 + 2r_{ij}^2 - 3R_0^2)}{(R_{cutoff}^2 - R_0^2)^3} & R_0 \leq r_{ij} \leq R_{cutoff} \\ 0 & r_{ij} \geq R_{cutoff} \end{cases} \quad (23)$$

where:

$$R_0 = R_{cutoff} - R \quad \text{and} \quad r_{ij} = |\mathbf{r}_i - \mathbf{r}_j| \quad (24)$$

In the case of electrostatic interactions, this approach can not be taken since the interaction at the cutoff distance is not negligible and so, introducing a switching function would lead to a lot of error. In addition, it can happen that, along the simulation time, some atoms that are around R_{cutoff} distance are not taken into account at all the snapshots which would introduce some discontinuities in the PES. A way to solve this problems is by means of the Ewald summation methods. [\[35\]](#)

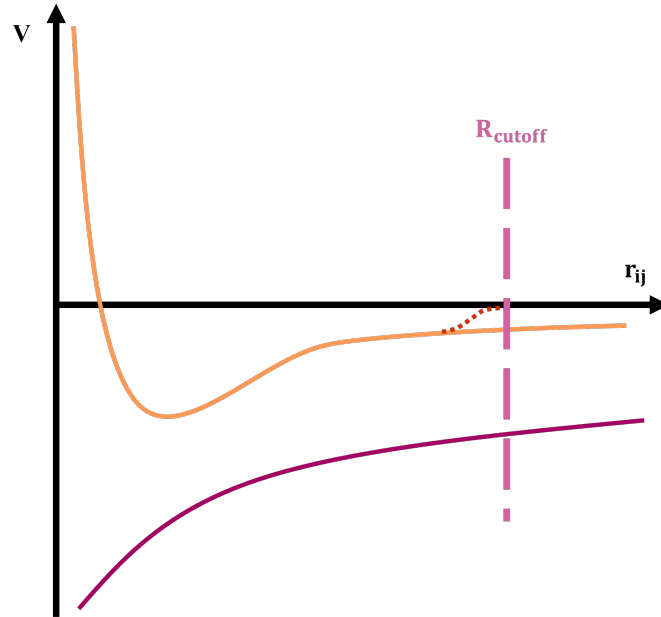


Figure 7: Schematic representation van der Waals (dark orange) and coulombic (dark pink) pair potentials as a function of the interatomic distance.

Ewald Summation

The Ewald summation method [\[35\]](#) is a widespread technique that accelerates the calculation of the N-body force summation of periodic or pseudo-periodic systems in which long-range contributions, such as electrostatic terms, are relevant.

This method consists in the application of a Gaussian charge distribution of opposite sign on top of each of the point charges of the system, as shown in Figure 8, to screen them. This transforms the long-range interactions into short-range interactions that are easier to compute. Then, the added Gaussian distributions are removed in the reciprocal space to compute the real long-range interactions. [22, 36]

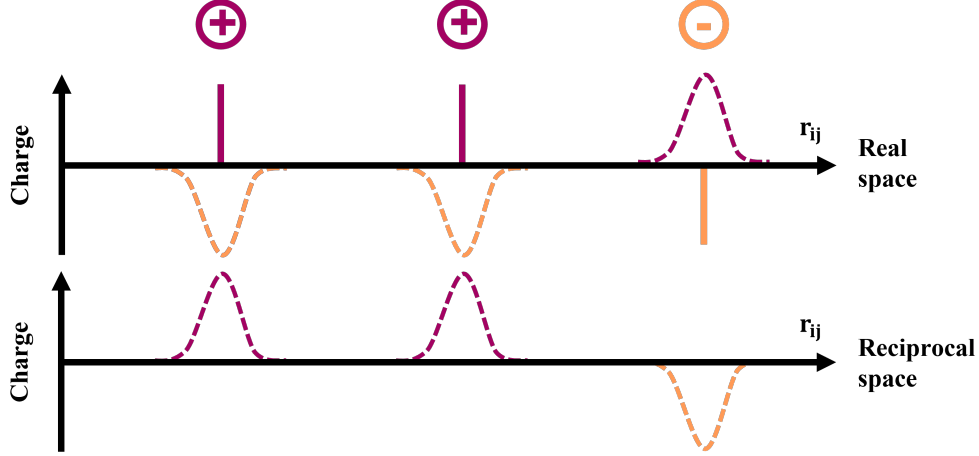


Figure 8: Schematic representation of the Ewald summation.

The electrostatic potential can be then computed as a sum of the short-range screened potential (V_{direct}) that now decays very rapidly to 0 when $\mathbf{n} \rightarrow \infty$ due to the error function ($erfcf$), the potential in the reciprocal space ($V_{reciprocal}$) and a correction term ($V_{correction}$):

$$V = V_{direct} + V_{reciprocal} + V_{correction} \quad (25)$$

that are given by:

$$V_{direct} = \frac{1}{2} \sum_{\mathbf{n}}' \sum_{i,j=1}^N \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|} ercf(\alpha |\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|) \quad (26)$$

$$V_{reciprocal} = \frac{1}{2\pi V} \sum_{i,j=1}^N q_i q_j \sum_{\mathbf{m} \neq 0} \frac{\exp\left(-\left(\frac{\pi \mathbf{m}}{\alpha}\right)^2\right) + 2\pi i \mathbf{m} |\mathbf{r}_i - \mathbf{r}_j|}{\mathbf{m}^2} \quad (27)$$

$$V_{correction} = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \quad (28)$$

where ' represents that each i atom is able to interact with its replicas but not with itself; α is

the width of the Gaussian and \mathbf{m} , a reciprocal-space vector expressed as:

$$\mathbf{m} = m_1 \mathbf{B}_x + m_2 \mathbf{B}_y + m_3 \mathbf{B}_z \quad (29)$$

$$\mathbf{B}_x = 2\pi \frac{\mathbf{L}_y \times \mathbf{L}_z}{\mathbf{L}_x (\mathbf{L}_y \times \mathbf{L}_z)} \quad \mathbf{B}_y = 2\pi \frac{\mathbf{L}_z \times \mathbf{L}_x}{\mathbf{L}_y (\mathbf{L}_z \times \mathbf{L}_x)} \quad \mathbf{B}_z = 2\pi \frac{\mathbf{L}_x \times \mathbf{L}_y}{\mathbf{L}_z (\mathbf{L}_x \times \mathbf{L}_y)} \quad (30)$$

4.1.5. Constrained Dynamics

The restriction of the motion in MD simulations presents several advantages: (i) saving computational time by neglecting several DOF, (ii) increasing of the time step and therefore the simulation time, (iii) enhancing the sampling by applying external forces to keep the system in a certain state and (iv) preventing wrong descriptions due to models that poorly describe certain DOF, like in the case of the O–H bonds in the TIP3P model.[\[37\]](#) However, these constraints need to be applied carefully because they can lead to poor results.[\[22\]](#)

The fix of the bond length can be achieved by introducing Lagrangian multipliers, as shown in Equation [\[31\]](#) in which the first expression at Equation [\[1\]](#) is modified.

$$\frac{d\mathbf{p}_i}{dt} = -\frac{dV}{d\mathbf{r}_i} + g'(\mathbf{r}_i)\lambda \quad (31)$$

Additionally, it is also possible to impose the constraints in an iterative way. First, an unconstrained motion of the system is obtained and then the positions are corrected:

$$\mathbf{r}_i(t + \Delta t) \longrightarrow \mathbf{r}_i(t + \Delta T) + \frac{\Delta t^2}{2m_i} \sum_{\gamma} \mathbf{F}_{\gamma}^c \quad (32)$$

where γ runs over all the constraints and the constrained forces (F_{γ}^c) are expressed as:

$$F_{\gamma}^c = \frac{\mu}{2\Delta t^2} \left(\frac{|\mathbf{d}_0 - \mathbf{d}|}{|\mathbf{d}_0|} \right) \quad (33)$$

being μ the reduced mass; \mathbf{d}_0 , the constrained bond vector and \mathbf{d} , the unconstrained bond vector. This last method is used in the algorithms SHAKE,[\[38\]](#) developed by Ryckaert et al. and in its extension RATTLE,[\[39\]](#) developed by Andersen, in which also momenta are constrained.

4.2. Hybrid Quantum Mechanics/Molecular Mechanics

Classical MD methods allow a fast evaluation of very large systems and account for environmental effects. However, they suppose a loss in the accuracy of the results. To overcome this problem, hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) approaches have been developed. [40–42] In this methodology, the system is usually divided in two regions treated at different levels of theory (see Figure 9(a)). The inner region, treated with a QM method, is composed by the part of the system that is of interest. For example, a chromophore, as in this work, or the reacting site. The rest of the system, the solvent and some other environmental molecules like proteins are considered by means of Molecular Mechanics (MM).

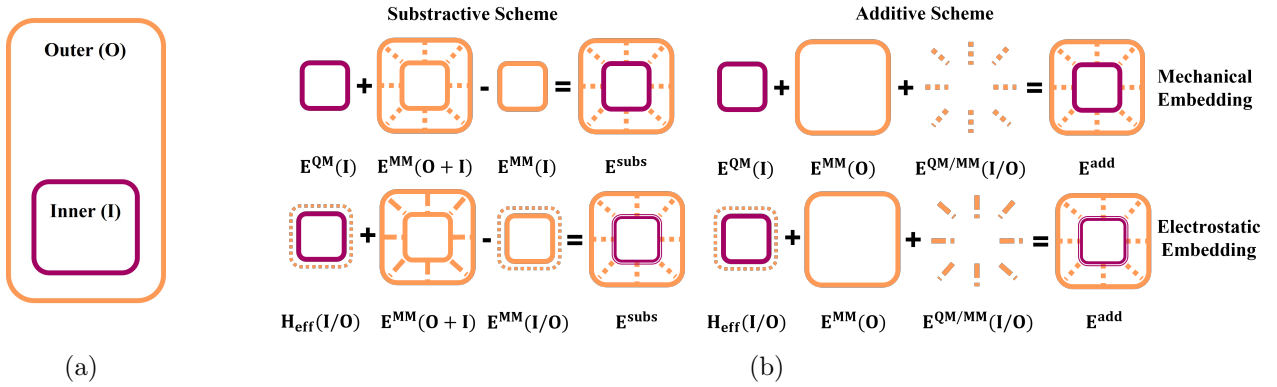


Figure 9: (a) Scheme of a QM/MM subdivision and (b) representation of the QM/MM schemes and embeddings. In light orange represented the classic treatment and in dark pink, the QM one.

There exist two approaches to describe the QM/MM energy of the system: the subtractive and the additive schemes. In the former, the energy (E^{subs}) is computed as:

$$E^{subs} = E^{QM}(I) + E^{MM}(I + O) - E^{MM}(I) \quad (34)$$

where $E^{QM}(I)$ represents the energy of the inner region computed with QM and $E^{MM}(I + O)$ and $E^{MM}(I)$, the energy of the hole system and the energy of the inner region, respectively, computed classically. As shown in the upper part of Figure 9(b), this approach accounts for the interaction energy between the two regions during the calculation of the hole system. Moreover, it computes twice the energy of the inner region (classically and QM) and therefore, it requires a subtractive term to avoid the repetition. This model does not take into account the polarization of the inner layer by the charges of the outer one because the interaction is computed classically

by a FF. However, in the so called electrostatic embedding, it can be considered by means of an effective Hamiltonian that introduces the interaction of the charges of the outer (q^O) region with the electrostatic potential of the inner region (V^I) to the vacuum Hamiltonian ($\hat{H}^0(I)$) so that the time independent Schrödinger equation becomes:

$$\hat{H}^{eff}|\Psi_I\rangle = \left(\hat{H}^0(I) + \hat{H}^0(I/O)\right)|\Psi_I\rangle = E^{QM}(I)|\Psi_I\rangle \quad (35)$$

where the interaction Hamiltonian ($\hat{H}^0(I/O)$) is given by:

$$\hat{H}^0(I/O) = \sum_i^N q_i^O(\mathbf{r}_i) V^I(\mathbf{r}_i) \quad (36)$$

In the additive scheme, the energy of the inner and outer parts, as well as the interactions between them ($E^{QM/MM}(I/O)$), are computed separately (see lower part of Figure 9(b)):

$$E^{add} = E^{QM}(I) + E^{MM}(O) + E^{QM/MM}(I/O) \quad (37)$$

where $E^{QM/MM}(I/O)$ is given by the sum of the classically described bonding energy ($E_{bonded}^{MM}(I/O)$), the van der Waals interactions (E_{vdW}^{MM}) and the electrostatic interactions (E_{elec}):

$$E^{QM/MM}(I/O) = E_{bonded}^{MM}(I/O) + E_{vdW}^{MM} + E_{elec} \quad (38)$$

Depending on the way of computing E_{elec} , the additive schemes can be described as: mechanical embedding, in which it is computed by means of the coulomb expression and no polarization is described; electrostatic embedding, in which the the inner region is polarized by the charges of the outer region by means of an effective Hamiltonian as in Equation 35 and polarizable embedding, in which both the inner and outer parts are polarized in a self consistent way by means of an additional polarization (\hat{H}_{pol}) term in the interaction Hamiltonian:

$$\hat{H}^0(I/O) = \hat{H}_{elec}(I/O) + \hat{H}_{pol}(I/O) = \sum_i^N q_i^O(\mathbf{r}_i) V^I(\mathbf{r}_i) - \sum_a \mu_a^0(\mathbf{r}_a) E^I(\mathbf{r}_a) \quad (39)$$

where $\mu_a^0(\mathbf{r}_a)$ represents the dipoles of the outer atoms located at \mathbf{r}_a and $E^I(\mathbf{r}_a)$, the electric field created by inner atoms at \mathbf{r}_a .

5. Density Functional Theory

Density Functional Theory (DFT) is an QM method that allows to reduce the computational time with respect to those based in the wavefunction since it substitutes the multi-electronic problem with $4n$ variables (three spatial and spin coordinate for each electron) by a one-particle problem with three spatial coordinates independently on the number of electrons of the system (n), since all the properties can be obtained from the electronic density (see Equation 40). This integral defines n , its peaks, the position of the nuclei and their height, v and their respective nuclear charges, among other properties. 34

$$\rho(\mathbf{r}) = n \int_0^\infty |\Psi(\mathbf{r}_1, \dots, \mathbf{r}_n)|^2 d\mathbf{r}_1 \dots d\mathbf{r}_n \quad (40)$$

5.1. Hohenberg-Kohn Theorems

DFT is based on two basic theorems given by Hohenberg and Kohn. The first of them states that the total energy of the system is a functional of the electronic density and that all the properties of the system can be obtained from the ground-state density. The second theorem states that the problem of determining the ground-state energy and density in a given external potential simply requires the minimization of a functional of the three-dimensional density function. 43 However, the shape of the universal functional is yet to be discovered and some approximations have to be used, as will be discussed latter on in Section 5.4.

In practice, the energy, expressed as a functional of the density, can be divided in three terms: the electronic kinetic energy ($T[\rho(\mathbf{r})]$), the electrostatic attraction between nuclei and electrons ($V_{ne}[\rho(\mathbf{r})]$) and the electronic repulsion ($V_{ee}[\rho(\mathbf{r})]$), where the nuclear repulsion energy is just a constant within the Born-Oppenheimer framework.

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + V_{ne}[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] \quad (41)$$

5.2. Kohn-Sham Approach

This approach does not provide an accurate result for the energy mainly because the kinetic energy is not properly described. To solve this problem, Kohn and Sham defined a fictitious system of non-interacting particles but with exactly the same density than the real system. 44

This approach allows to obtain an expression equivalent to the time-independent Schrödinger equation, and thus it needs to be solved self-consistently, in which the Hamiltonian of the non-interacting (ni) system contains an external potential (V_s) to describe the movement of the non interacting electrons and a kinetic term ($T_{ni}[\rho(\mathbf{r})]$).^[28] Of course, this kinetic energy differs from the exact one in the kinetic energy correlation, the missing particle-particle interaction. In addition, it introduces the so called Kohn-Sham orbitals (χ^{KS}) that represent the orbitals of the non-interacting system and that increase the complexity of the system to $3n$ variables. These orbitals can be calculated numerically or variationally as an expansion of basis functions. Hence, Equation ^[41] can be rewritten as:

$$E[\rho(\mathbf{r})] = T_{ni}[\rho(\mathbf{r})] + V_{ne}[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] \quad (42)$$

where $J[\rho(\mathbf{r})]$ represents the electronic repulsion as a functional of the density that can be classically computed from Coulomb's expression (see Equation ^[44]) and $E_{xc}[\rho(\mathbf{r})]$, called exchange-correlation functional, the exchange-correlation energy, expressed as in Equation ^[45]. The first term corresponds to the kinetic energy correlation while the second, that arises from the difference between the classical and QM description of the electronic repulsion due to Pauli's exclusion principle, contains potential correlation as well as exchange energy.^[34] Moreover, $V_{ne}[\rho(\mathbf{r})]$ is given by:

$$V_{ne}[\rho(\mathbf{r})] = - \sum_i^N Z_i \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_i|} d\mathbf{r} \quad (43)$$

$$J[\rho(\mathbf{r})] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (44)$$

$$E_{xc} = (T[\rho(\mathbf{r})] - T_{ni}[\rho(\mathbf{r})]) + (V_{ee}[\rho(\mathbf{r})] - J[\rho(\mathbf{r})]) \quad (45)$$

where Z_i is the charge resulting from the nucleus.^[45] A key point in this procedure is the determination of V_s since it should be generated to equal the density of the fictitious system to that of the real system:

$$V_s(\mathbf{r}) = \sum_i^N \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} + \int \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_2 - \mathbf{r}|} d\mathbf{r}_2 + V_{xc}(\mathbf{r}) \quad (46)$$

with:

$$V_{xc} = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \quad (47)$$

$$\left(-\frac{1}{2} \nabla_i^2 + V_{ni}(\mathbf{r}) \right) \chi_i^{KS} = \epsilon_i \chi_i^{KS} \quad \langle \chi_i^{KS} | \chi_j^{KS} \rangle = \delta_{ij} \quad (48)$$

$$\rho(\mathbf{r}) = \sum_{i=1}^n |\chi_i(\mathbf{r})|^2 \quad (49)$$

Although exact in its formulation, DFT needs to be solved in an approximate way by means of empirically parameterised equations to describe the exchange-correlation contribution that, in practice are responsible in the reduction of the computational time. [46]

5.3. Time-Dependent DFT

The original DFT can only be applied to compute the ground state of the system and its properties. To study the electronically excited states of the system, it is necessary to study the time evolution of the density of the system under the action of an external potential since there exist a one-to-one correspondence between Time Dependent (TD) densities and TD external potentials. [22] This correspondence is given by the Runge-Gross theorem, [47, 48] the TD analogue of the Hohenberg-Kohn theorem, that states that two densities $\rho(\mathbf{r}, t)$ and $\rho'(\mathbf{r}, t)$ evolving from a common initial state $\Psi_0 = \Psi(t_0)$ under the influence of two potentials $V(\mathbf{r}, t)$ and $V'(\mathbf{r}, t)$ are always different provided that the potentials differ by more than a purely time-dependent function:

$$V(\mathbf{r}, t) \neq V'(\mathbf{r}, t) + c(t) \quad (50)$$

As in the original DFT formulation, the correspondence between TD densities and TD potentials is established for an arbitrary particle-particle interaction and thus, a non-interacting system under an external TD potential ($V_{ext}(\mathbf{r}, t)$) that evolves in time as the real one is defined. This approach leads to the so called TD Kohn-Sham equations:

$$i\hbar \frac{\partial \chi_i(\mathbf{r}, t)}{\partial t} = \left(-\frac{\nabla_i^2}{2} + V_{KS}[\rho(\mathbf{r}, t)] \right) \chi_i(\mathbf{r}, t) \quad i = 1, \dots, n \quad (51)$$

where $V_{KS}[\rho(\mathbf{r}, t)]$ can be divided in three contributions: $V_{ext}(\mathbf{r}, t)$, the TD Hartree potential ($V_H[\rho(\mathbf{r}, t)]$) and the exchange-correlation potential (V_{xc}):

$$V_{KS}[\rho(\mathbf{r}, t)] = V_{ext}(\mathbf{r}, t) + V_H[\rho(\mathbf{r}, t)] + V_{xc} \quad (52)$$

with:

$$V_H[\rho(\mathbf{r}, t)] = \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d^3\mathbf{r}' \quad (53)$$

5.4. Exchange-Correlation Functionals

The exchange-correlation functional is a key ingredient of the DFT method since it allows to compute the exchange correlation potential of the system. Even though the existence of an exact and universal functional has been proven, its shape is yet to be discovered. However, it is possible to obtain its properties.^[34] Some of them are:

1. The self-interaction-free energy. For example, the exchange for one-electron systems and the Coulomb energies should exactly cancel each other and the correlation energy has to be zero.
2. Recovery of the uniform electron gas when the density becomes constant.
3. Linear scaling of the electronic coordinates should result in a similar linear scaling of the exchange energy.

$$\begin{aligned}\rho_\lambda(x, y, z) &= \lambda^3 \rho_\lambda(\lambda x, \lambda y, \lambda z) \\ E_x[\rho_\lambda] &= \lambda E_x[\rho]\end{aligned}\tag{54}$$

4. No direct scaling for the correlation energy. However, scaling the electronic coordinates by a factor larger than 1 should increase the magnitude of the correlation.

$$-E_c[\rho_\lambda] > -\lambda E_c[\rho] \quad \text{when } \lambda > 1\tag{55}$$

5. The correlation energy for a finite system approaches a negative constant when the scaling parameter goes to infinity.

In order to satisfy the previous constraints, several approximations to it have been developed. The development is based on fitting some parameters on mathematical expression focusing on fulfilling the properties that the exact functional should have (like the previously mentioned), on the reproduction of some experimental values or on a combination of both.

Figure ^[10] represents the different families of exchange correlation functionals. The most simple of them are the Local Density Approximation (LDA) or its more general version the Local Spin Density Approximation (LSDA) that are based on the uniform electron gas model and only requires the density for computing the exchange correlation energy. A functional belonging to

this rung is the VWN, described by Vosko, Wilk and Nusair.[\[49\]](#) The Generalized Gradient Approximation (GGA) additionally introduces the gradient of the density ($\nabla\rho(\mathbf{r})$) allowing to account for slow variations on the density of the system and Meta-GGAs, apart from the previously mentioned properties, also consider the kinetic energy density given by Equation [\[56\]](#) LYP,[\[50\]](#), [\[51\]](#) constructed by Lee, Yang and Parr, is an example of this two level of the ladder.

$$\tau(\mathbf{r}) = \sum_i |\nabla\chi_i(\mathbf{r})|^2 \quad (56)$$

Upper, in the Jacob's ladder the hybrid or hyper-GGA methods are found. They are based on the Adiabatic Connection Formula[\[52\]](#) that arises from the exact connection between the exchange–correlation energy and the corresponding hole potential connecting the non-interacting reference and the actual system. In this approach the exchange energy is exactly that given by the Hartree–Fock theory for the occupied Kohn-Sham orbitals while a good approximation of the correlation is obtained by means of mixing the Hartree-Fock exchange with LDA. Both functionals used during this work, B3LYP[\[53\]](#), [\[54\]](#) and its long-ranged corrected version CAM-B3LYP,[\[55\]](#) belong to this level of the ladder and are expressed as a linear function of lower rungs in the ladder.

$$E_{xc}^{B3LYP} = aE_x^{HF} + (1 - a)E_x^{LDA} + b\Delta E_x^{B88X} + (1 - c)E_c^{GGA} \quad (57)$$

$$E_{xc}^{CAM-B3LYP} = \alpha E_x^{HF} + (1 - \alpha)E_x^{LDA} + c^{B88}\Delta E_x^{B88} + \beta(E_x^{LR,HF} - E_x^{LR,B88}) + cE_c^{LYP} + (1 - c)E_c^{VWN} \quad (58)$$

with $a = 0.2$, $b = 0.72$, $c = 0.81$, $\alpha = 0.19$, $\beta = 0.46$ and $c^{B88} = 0.81$.

Generalized random phase methods are founded at the top of the Jacob's ladder and they require the knowledge of the full Kohn-Sham orbitals, not only the occupied but also the unoccupied ones, which significantly improves the description of dispersion interactions. The OEP2 functional[\[56\]](#)–[\[58\]](#) belong to this rung and is computationally equivalent to an iterative Møller–Plesset second-order perturbation theory (MP2) calculation.[\[59\]](#)

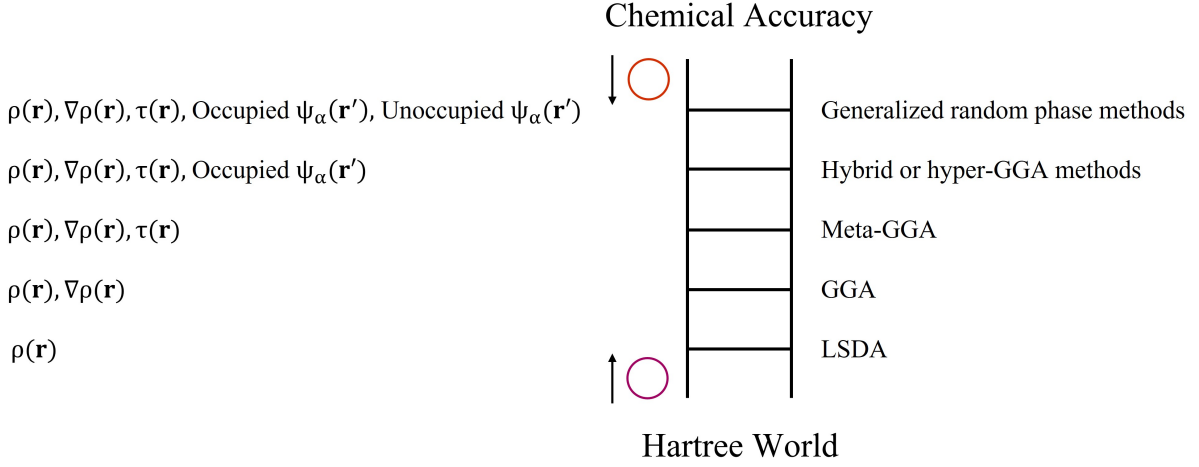


Figure 10: Jacob's ladder of density functional approximations.

5.5. Basis Sets

Not only DFT but also the *ab initio* methods require the description of molecular orbitals. However, the shape of their function is very complex and can not be easily defined. To overcome this problem, they are expanded in a set of known atomic orbitals, called basis set. In the complete basis set limit this approach results in the exact solution for a particular level of theory. Nevertheless, this limit is not reachable since it implies the use of a huge number of functions and an immense computational effort. Hence, this approach only leads to an approximation of the exact molecular orbitals.[\[34\]](#)

There are two main types of basis sets: Slater Type Orbitals (STO) and Gaussian Type Orbitals (GTO) given by:

$$\begin{array}{ll}
 \text{STO} & \chi_{\zeta,n,l,m}(r, \theta, \varphi) = N Y_{l,m}(\theta, \varphi) r^{n-1} e^{-\zeta r} \\
 \text{GTO} & \chi_{\zeta,n,l,m}(r, \theta, \varphi) = N Y_{l,m}(\theta, \varphi) r^{2n-2-l} e^{-\zeta r^2}
 \end{array} \tag{59}$$

where ζ represents the number of functions included to describe an atom; n , l and m , the quantum numbers; r , θ and φ , the polar coordinates; N , the normalization constant and $Y_{l,m}$, a spherical harmonic function. Although GTOs have problems representing the proper behaviour near the nucleus, fall off too rapidly far from the nucleus compared with STO basis sets and need three times more functions to reach the same accuracy, they are preferred in terms of computational efficiency since the required integrals can be more easily calculated.

5.5.1. Classification of Basis Sets

When using either STOs or GTOs, their accuracy will be given by the number of functions used in the expansion. Moreover, there is always a minimum amount of them, referred to as minimum basis set, where only those functions needed to contain all the electrons of the neutral atoms are used. However, the result given by them is rather poor and more functions need to be added to the basis set. This is achieved by doubling, Double Zeta (DZ) type basis; tripling, Triple Zeta (TZ) type basis; or even quadrupling or quintupling the functions of the minimum basis set. Nevertheless, this approach can increase significantly the computational time since it scales with M . Therefore, a further approximation in which only the valence orbitals are duplicated a few times, can be used. This split valence basis can be justified by the fact that core orbitals do not participate in the chemical bonding and are very similar to those of the isolated atoms. [\[34\]](#)

5.5.2. Contracted Basis Sets

Up to this point, it has been assumed that all the coefficients of the molecular orbital expansion need to be optimized to minimize the energy. However, core-electrons account for a large part of the total energy and so this minimization will tend to increase the participation of core orbitals and decrease that of the valence ones, which are actually the most important ones in order to describe the properties of the system. Therefore, it would be needed a bigger amount of functions to achieve an accurate result. However, this approach increases the computational time. For this reason, and taking into account that the variation of core orbitals with respect to the chemical bonding situation is very small, and therefore the coefficients in front of them also change very little, it is a good approximation to describe this core orbitals as a linear combination of Primitive Gaussian Type Orbitals (PGTO), the full set of basis functions, in which the expansion coefficients are fixed. This fixed linear combination of PGTOs to form a smaller basis set is known as a basis set contraction, resulting in Contracted Gaussian Type Orbitals (CGTO), and suppose a reduction of the computational time. [\[34\]](#), [\[60\]](#)

$$\chi(CGTO) = \sum_i^k a_i \chi_i(PGTO) \quad (60)$$

Pople Basis Sets

This work uses the so called Pople $k-nlmG(h, H)$ basis sets, that are split valence basis sets in which k indicates the contraction of the core orbitals, how many PGTOs are used to represent them. The nlm values correspond to both: (i) the degree of splitting of the basis set for s- and p- functions, meaning nl a DZ basis set and nlm a TZ basis set, and (ii) how many PGTOs are used for their representation. Moreover, numbers after the G letter, between parenthesis and separated by a coma, indicate the polarization functions (d- or f- type) added to heavy (h) and hydrogen atoms (H), respectively. In addition, is possible to include diffuse functions, represented by one or two “+” before the G letter.

6-31+G(d)

This basis set [61-67] presents a contraction of six PGTOs for the core orbitals and a double split of the valence orbitals. The inner part of the valence orbitals is a contraction of three PGTOs and the outer part of the valence is represented by one PGTO. Additionally, it contains a d- type function on non-hydrogen atoms to describe polarization and a diffuse function.

6-311G(2d,p)

This basis set [63, 68, 69] presents the same contraction than 6-31+G(d) for the core orbitals but a triple split valence basis where the valence orbitals split into three functions, represented by three, one and one PGTOs, respectively. Moreover, it describes polarization by means of two d-type functions on heavy atoms and one p-type function on hydrogen atoms.

Theoretical Procedure

In this work MD simulations of the complex oxyluciferin/luciferase were carried out. Two different approaches were considered. In the simplest one, the time evolution of the whole system was obtained by means of the classical equations of motion. In this case, the interatomic interactions were computed by means of a FF. In the other approach, some of the conformations, sampled from the classical trajectory, were evolved by means of a hybrid QM/MM MD simulation to obtain a QM equilibration of the chromophore. For each of the approaches, some conformations were sampled and single point QM/MM calculations were performed on top of them to compute the electronic transition properties. This procedure is schematically represented in Figure 11.

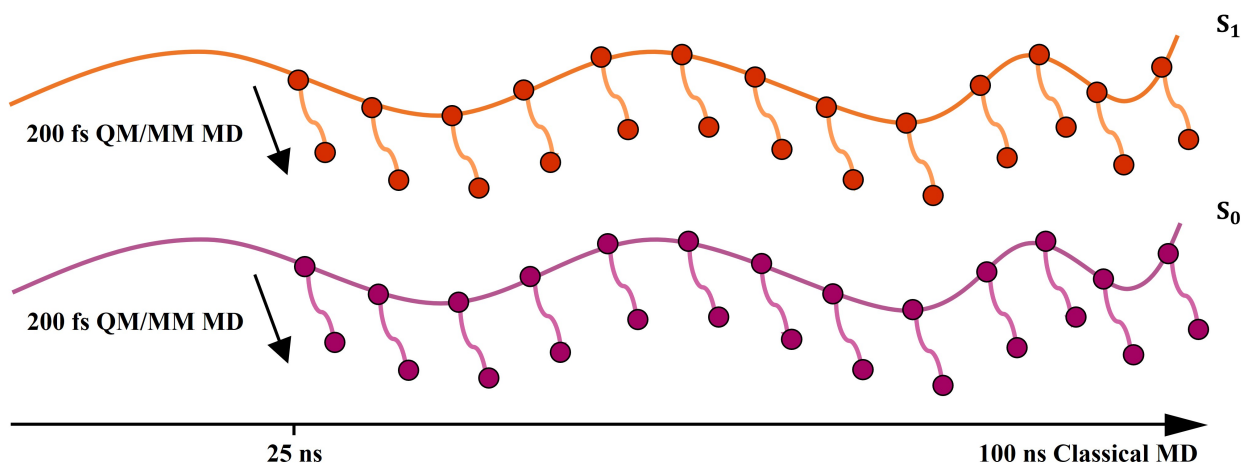


Figure 11: Scheme of the computational procedure followed during this work.

6. Classical Molecular Dynamics Simulations

The system was constructed from the X-ray diffraction structure file of firefly luciferase crosslinked in the second catalytic form (PDB ID: 4G37 [12]) using PyMol. [70] For this, the luciferyl adenilate ligand was substituted by the optimized geometry of the oxyluciferin in both S_0 and S_1 states. This geometry optimization was performed by means of the B3LYP [53, 54]/6-311G(2d,p) [63, 68, 69] level of theory and the Gaussian16 [71] software. In the case of the S_1 optimization, the TD-DFT calculation took into account 10 roots. Then, the system was solvated with the tleap module of AmberTools20 [30] by TIP3P [37] water molecules within a truncated octahedral box, ensuring a solvent shell of at least 12 Å from any solute molecule.

This shape of the cell is more efficient than a sphere since it allows to use periodic boundary conditions while having a similar volume. No ions were added since the cell was already neutral. Moreover, the ff19SB[72] and General Amber Force Field (GAFF)[73] FFs were used to describe, respectively, the luciferase enzyme and the dihedral, improper torsion angles and Lenard-Jones parameters of the chromophore. The parameters corresponding to the description of bonds lengths and bond angles of the S_0 and S_1 states of the oxyluciferin molecule were obtained by means of the Seminario method.[24] The use of this method requires the optimized geometry, the Hessian matrix and the the ESP charges that were all obtained from DFT and TD-DFT calculations, respectively for the S_0 and S_1 states using the Gaussian16[71] software. Specifically, for both of them the B3LYP[53, 54]/6-311G(2d,p)[63, 68, 69] level of theory was employed. This level of theory was selected taking into account previous studies on the system.[74, 75]

After the setup, the system was first minimized for 5000 steps using the steepest descent algorithm followed by another 5000 steps using the conjugate gradient algorithm, the latter being more efficient close to an energy minimum and the former used at the initial stages of the optimization that are far from the minimum. Consecutively, the system was progressively heated from 0 to 300 K for 500 ps in the NVT ensemble, to avoid an abnormal expansion of the system, using a timestep of 2 fs and equilibrated to a constant temperature for another 500 ps in the same ensemble, using a Langevin thermostat[25, 26] with 2.0 ps^{-1} of collision frequency. Afterwards, a production of 100 ns with, again, a 2 fs timestep and within the NPT ensemble was performed. Temperature was again controlled by means of the Langevin thermostat with 1.0 ps^{-1} of collision frequency, while pressure was controlled by means of the MC barostat[27] setting them to 1.0 bar and 303.15 K respectively. This ensemble allows an adjustment of the density during the simulation. The electrostatic interactions were computed using the particle-mesh Ewald[35] method, with a grid spacing of 1.0 \AA , a cutoff for the non-bonded interactions of 9 \AA and a switching distance of 7 \AA . In addition, bonds involving hydrogen atoms were constrained using the SHAKE[38] algorithm.

The first 25 ns (500 steps) of the production run for both MD simulations in the S_0 and the S_1 states were discarded as the equilibration time of the protein structure, as shown in Figure 12 with the Root Mean Square Deviation (RMSD) of the protein. Then, the script `main.qminputs.py` of the MoBioTools[76] was used to extract 200 geometries from the last 75 ns of the MD trajectory, and to generate the QM/MM input files to compute the absorption spectrum (from the S_0 trajectory) and the emission spectrum (from the S_1 trajectory) of the

chromophore. 10 excited states were considered at the B3LYP/6-311G(2d,p) level of theory for all the excited state TD-DFT/MM calculations, that corresponds with an electrostatic embedding scheme since the charges of the environment enter the Hamiltonian. Additionally, in the case of the emission spectrum only the energy of the $S_1 \rightarrow S_0$ electronic transition was considered since most of the systems emit from the S_1 . Moreover, most of the systems emit from the S_1 state since internal conversion between states is usually faster than radiative relaxation. The selection of the geometries was done following three different approaches: randomly, equidistantly and following a Metropolis criteria.[\[23\]](#)

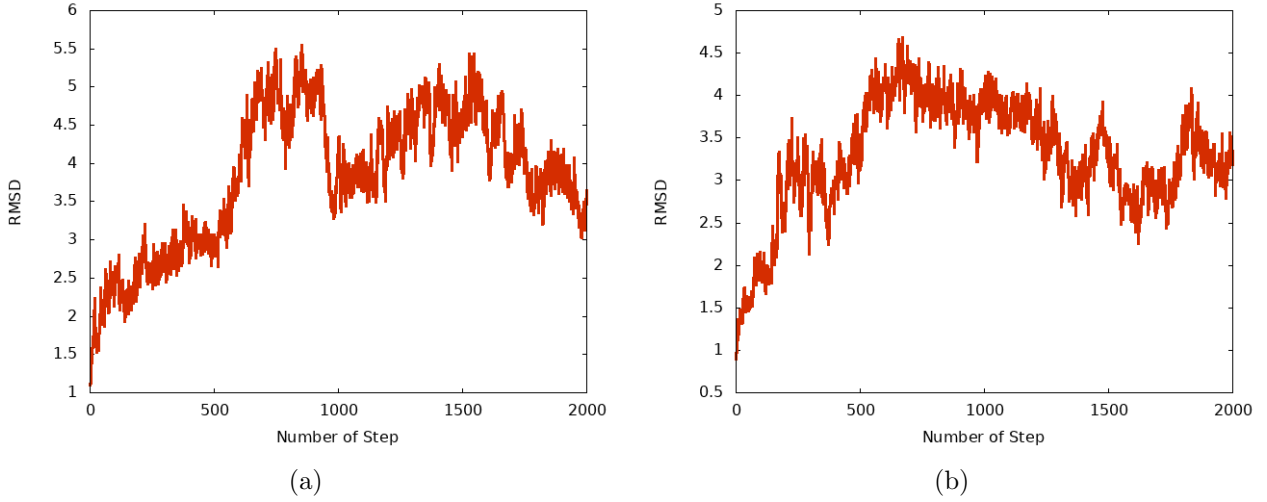


Figure 12: RMSD of the system in the (a) S_0 and (b) S_1 trajectories.

The Metropolis criteria is based on an algorithm of the same name that is able to, stochastically, accept or reject a geometrical configuration taking into account its potential energy. A configuration is accepted with a probability proportional to its Boltzmann factor P , which is given by Equation [\[61\]](#). The potential energy of a particular configuration (U) is computed relative to the lowest-energy snapshot of the dynamics. However, it was found that this lowest-energy snapshot has a much lower energy than all the other configurations, leading to a situation where the probability to accept a configuration by the Metropolis criterion is very low. To circumvent this problem, the potential energy of every snapshot was scaled by dividing by the average potential energy of all the snapshots ($\langle U \rangle$). This reduces the energy differences with respect to the lowest-energy configuration. For this purpose a FORTRAN code, presented in Annex II, has been created.

$$P_i = e^{-\frac{U_i}{k_B T \langle U \rangle}} \quad (61)$$

In addition, the cpptraj[\[77\]](#) module of AmberTools20 was used to extract the dihedral angle

(SCCS) between the tiazol and benzotiazol moieties of the chromophore, marked in Figure 13(b), to perform further analysis of the trajectory. Moreover, TheoDORE[78] software was used to compute the intramolecular CT between the two moieties of the molecule by partial summations over squared transition density matrix elements of molecular fragments.

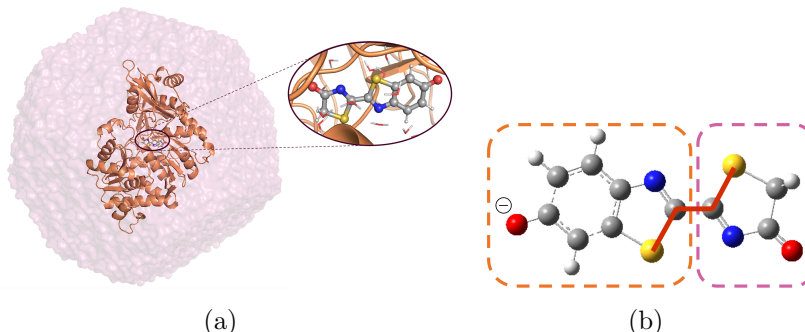


Figure 13: Graphic representation of (a) the whole system and (b) the two moieties of the chromophore. In light orange, the benzotiazol moiety and, in light pink the tiazol one. In dark orange, the SCCS dihedral angle.

7. Hybrid QM/MM Molecular Dynamics Simulations

After both S_0 and S_1 classical MD simulations were performed, 100 equidistant geometries, together with their respective velocities, were extracted from the trajectory. Each of them were processed by means of the `cobramm-droplet.py` script within COBRAMM 2.0,[79] a program package that interfaces several known commercial and academic softwares for molecular modeling and is able to control their simulations and analyse the resulting outputs. In this case, it was used to interface Gaussian16[71], to quantum mechanically compute the electronic part of the system, with Amber20[30] for the classical part.

The system was divided in three different layers depending on level of theory with which they were treated: (i) high layer, composed by the chromophore and treated quantum mechanically; (ii) medium layer, formed by the protein and the water molecules within 4 Å from the chromophore and treated classically; and (iii) low layer, including the rest of the system, the rest of the water molecules, and frozen during the simulation. To do so, a script, that was able to automatically set up each of the geometries, was created. First, it striped all the water molecules from the system. From this part of the system, that only contains the chromophore and the protein, it was possible to get the high and most of the medium layer topology. Then, the protein was removed and the rest of the medium layer, the water molecules within 4 Å from the chro-

mophore, and the low layer were generated. Consecutively, the whole system was analysed to obtain the parameters file and the topology files were merged. Finally, the `cobramm-rattle.py` script is used to apply a constraint in the motion of the hydrogen atoms of the water molecules by means of the RATTLE algorithm.^[39] This not straight forward approach was needed due to the big size of the system since usually the `cobramm-droplet.py` script is able to automatically generate the three different layers for a given structure, if the size is not too large.

The QM region was treated by CAM-B3LYP^[55]/6-31+G(d)^[61–67] surface using ground state DFT in the case of the S_0 and TD-DFT with three roots in the case of the S_1 . The protein and the water molecules of the medium layer were described, respectively, by the previously used ff12SB and TIP3P FFs. The 200 trajectories (100 in the S_0 and 100 in the S_1) were run for 200 fs in the microcanonical ensemble (NVE) without periodic boundary conditions, since it is not installed in COBRAMM 2.0. This is the reason of freezing the external atoms during the QM/MM MD simulations. This time, the cutoff for the non-bonded interactions was 10 Å.

The absorption and emission spectra were calculated by means of a single point calculation on top of the last step of each trajectory at the same level of theory than in the case of the classical trajectory, B3LYP/6-311G(2d,p)/MM with 10 roots. Again, only the $S_1 \rightarrow S_0$ electronic transition was used to compute the emission spectrum. In this case, the preparation of the inputs was done by means of the `cobramm-get-step.py` script.

8. Potential Energy Surface Rigid Scan

During the MD simulations, it was observed a large variation of the SCCS dihedral angle, which breaks the planarity of the molecule. Due to the partial intramolecular CT character of the bright state of the chromophore,^[74] this torsion motion could be relevant in the photophysics and it is analyzed in detail. Therefore, a rigid scan around it was calculated at TD-DFT CAM-B3LYP/6-311G(2d,p) level of theory with 10 roots to determine the shape of the PES of both S_0 and S_1 states as a function of this specific DOF as well as the properties of the electronic transition along the torsion motion. The initial structure of the chromophore was the same as the optimized geometry required for the production of the parameters for the oxyluciferin molecule for the S_1 state with the Seminario method. For this, the Gaussian16 software was used. In addition, the transition density matrix was analysed by TheoDORÉ to determine the intramolecular CT character along the torsion.

Results

As mentioned before, in this work, two types of MD simulations were carried out, classical and QM/MM MD, that allows to investigate the impact of the potential energy model used during the sampling process of the PES of the S_0 and S_1 states. In addition, classical MD trajectory has been used to analyse the effect of the sampling procedure. This section is organized so that the results of sampling methodologies are first discussed and latter on, effect of the potential energy model is studied. Moreover, computed spectra are contrasted to experimental data.

9. Effect of the Sampling

The simulation of electronic spectra from classical MD simulations requires QM single point calculations on top of the geometries of the trajectory. However, and taking into account the long simulation time required to properly sample the conformational space of the system, it is unaffordable to perform those calculations for each of the geometries of the trajectory. Therefore, what it is usually done is to extract some of them as representative of the whole ensemble. Thus, it is crucial to have an adequate criteria for this sampling since the properties under study need to be accurately described. On this behalf, the convergence of the maxima of the S_0 - S_1 electronic transition is studied with respect to the number of sampled geometries from the classical trajectories when three different criteria are applied. For each number of selected geometries (n) of each type of sampling, the electronic excitation energy has been calculated, as shown in Equation [62](#), as an average weighted by the oscillator strength (f) of the transition.

$$\langle E_{transition} \rangle = \frac{\sum_i^n E_{transition,i} f_i}{\sum_i^n f_i} \quad (62)$$

The results of this analysis are presented in Figure [14](#). In the case of the MD in the S_0 surface, all the values of the average excitation energies oscillate between 2.27 and 2.31 eV. However, the oscillations of equidistant sampling (2.27-2.31 eV) are larger than those of random (2.28-2.30 eV) and Metropolis sampling (2.29-2.30 eV). These last two sampling methods present very small variations when more than 100 geometries are considered and converge to 2.30 and 2.29 eV, respectively.

The convergence of the MD simulations on the S_1 surface presents, for the three sampling

methods, stronger oscillations. In the case of random sampling, the values are between 2.14 and 2.17 eV and convergence is reached after 160 geometries are considered, around the value of 2.16 eV. Equidistant sampling oscillates between 2.12 and 2.19 eV and seems to converge to 2.17 eV from 160 geometries on. In the case of the Metropolis approach, the convergence is achieved after 120 geometries at a value of 2.17 eV after oscillating between 2.15 and 2.18 eV. Looking at these results, it can be observed that random and Metropolis sampling are converged. However, more geometries should be considered in the case of the equidistant sampling since it is not completely clear whether it is fully converged or not, although the average excitation energy after 200 geometries is very similar to the value obtained with the random and Metropolis samplings.

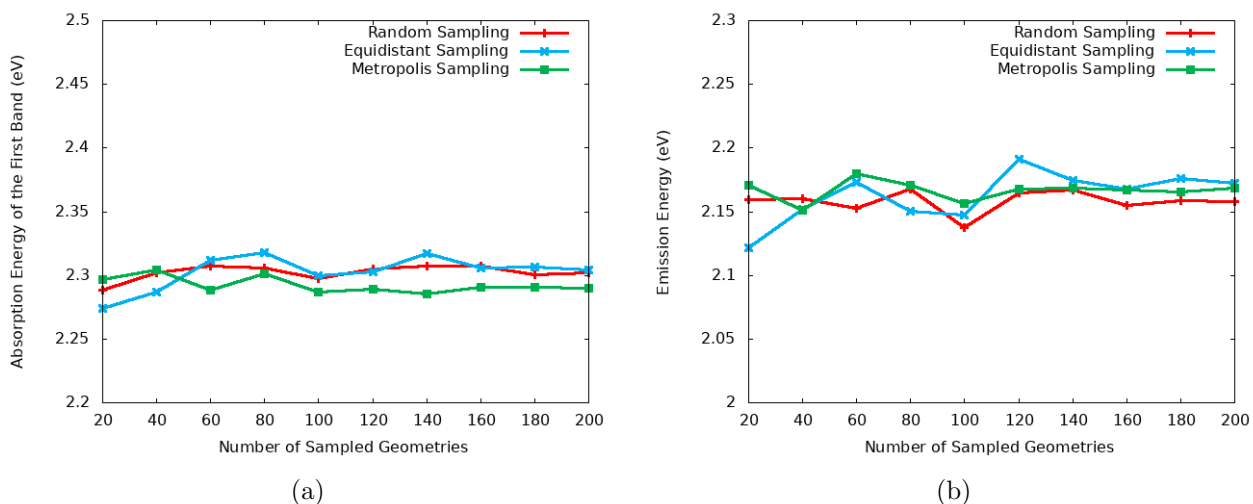


Figure 14: Convergence of the maxima of the S_0 - S_1 transition with respect to the number of sampled geometries for the MD simulation on the (a) the S_0 surface and (b) the S_1 surface.

When these 200 geometries for each of the sampling methods are used to compute the absorption and emission spectra, as shown in the Figures [15\(a\)](#) and [15\(b\)](#), it can be observed that the obtained shapes and intensities are very similar. In the case of the absorption spectra, the two main bands correspond to the $S_0 \rightarrow S_1$ (Highest Occupied Molecular Orbital (HOMO) - Lowest Unoccupied Molecular Orbital (LUMO)) and $S_0 \rightarrow S_3$ (HOMO-2 - LUMO) transitions and peak at 2.34 and 3.08 eV, respectively. In addition, it has been observed that the S_1 presents a small mixing with the S_2 (HOMO-1 - LUMO) transition in a 36% of the analysed geometries, both states contain a significant contribution (more than a 0.2 on the weight of the wavefunction) of the HOMO - LUMO and the HOMO-1 - LUMO transitions. This excitation, that is close in energy to the $S_0 \rightarrow S_1$, is usually dark since it involves a $n\pi^*$ transition (see Figure [15\(c\)](#)) and thus, it presents a low oscillator strength due to a change in the symmetry of the orbitals.

Moreover, in a 2% of the snapshots along the simulation, a state flipping happens between them. In the case of the emission spectra, which peaks at 2.17 eV, also the three sampling methods result in very similar bands (see Figure 15(b)), although the difference is larger than for the absorption band. This was expected since the convergence profile of the emission maximum (see Figure 14) displays larger oscillations than the convergence of the absorption maximum. Additionally, the mixture between states is present for 35% of the snapshots and the state flipping behaviour is present in 2% of the computed geometries.

During the dynamics, an important torsion motion around the C-C single bond that separate both fragments of the chromophore was observed. The effect of this torsion motion is analyzed in the following sections in detail.

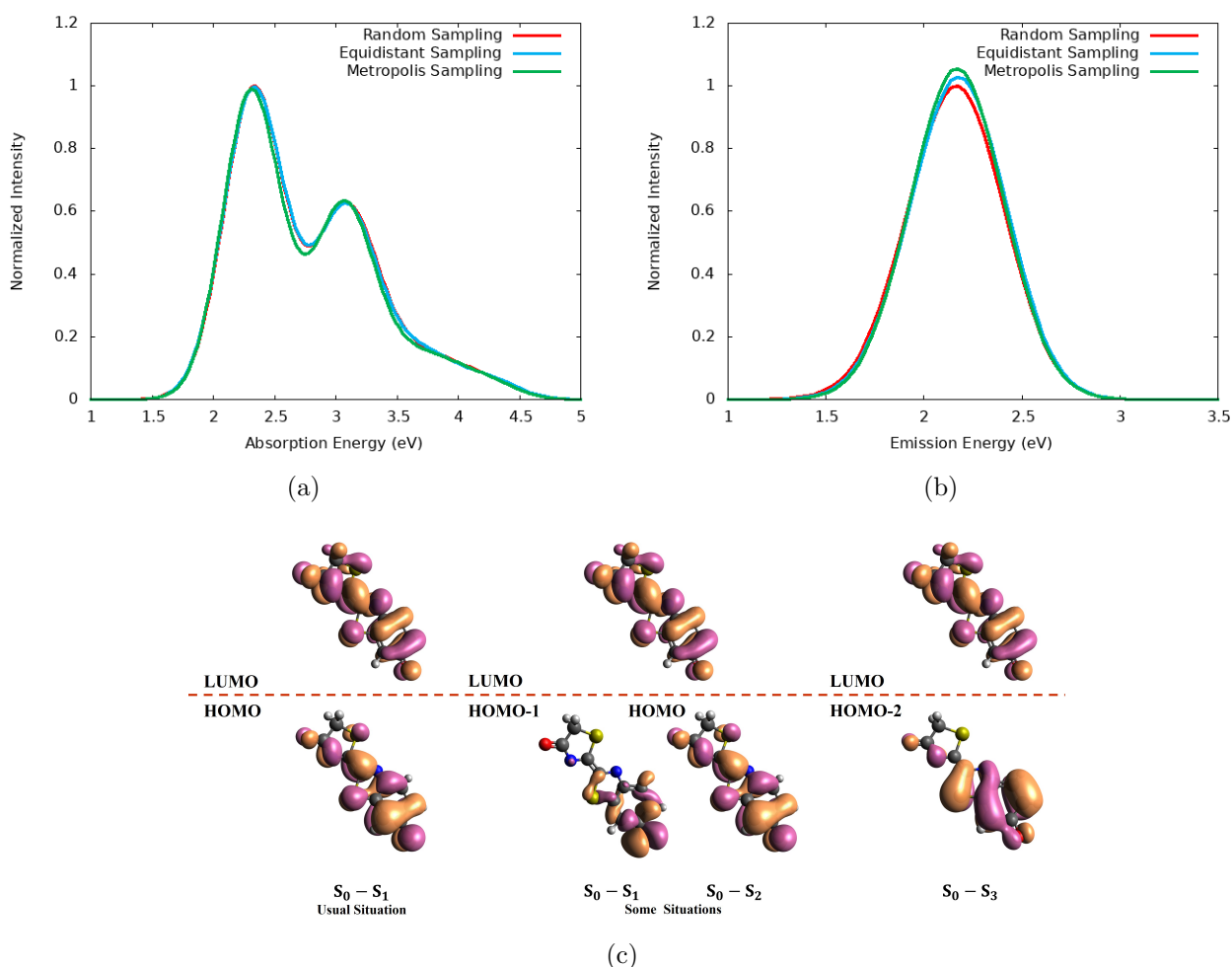


Figure 15: (a) Absorption and (b) emission spectra of the oxyluciferin chromophore within the luciferase enzyme depending on the sampling process and (c) the orbitals involved in the main transitions.

10. Effect of the Torsion around the C-C Single Bond

In addition to the sampling study, this work performs an analysis of the variation the electronic properties with the torsion around the C-C single bond. This specific DOF has been selected because it is the only one in the chromophore with a wide range of motion. Additionally, this choice is supported by the results of the QM/MM simulations, as will be commented later on.

10.1. Static Scan Calculations around the C-C Single Bond

As a first approach, a rigid scan of the SCCS dihedral angle is performed. Figure 16(a) shows the energy of the S_0 (in dark orange) and S_1 (light pink) states as a function of the dihedral angle. The S_0 potential curve presents two minima at 180 and 0° and two maxima at 90 and 270°. On the flip side, the S_1 surface has four minima corresponding to 0, 90, 180 and 270° and four maxima at 40, 135, 225 and 320°, although the region between 320° (-40) and 40 is almost flat. Thus, the maxima of the S_0 state correspond to the minima of the S_1 , resulting in a huge difference in the excitation energy along the scan (from 1.66 to 2.54 eV), as represented in Figure 16(b). However, as the excitation energy decreases, the oscillator strength of the transition decreases, reducing its brightness. This reduction in oscillator strength is due to a change in the symmetry and the overlap of the orbitals involved in the transition, shown in Figure 16(d). This is consistent with the increase in the CT character since the delocalization of both the HOMO and the LUMO orbitals is broken and each orbital is located on a different fragment of the molecule when the planes of the two moieties are perpendicular.

Considering this important variation of the properties when the molecule deviates from planarity (180°), the introduction of sampling in the theoretical model by means of classical or QM/MM MD may potentially modify the electronic properties of the absorption and emission spectra. In addition, apart from the previously mentioned mutation of the protein to induce a modification in the electrostatic interactions and the electronic coupling to induce a red-shift of the system, it would be interesting to modify its sequence of residues to induce an increase in the torsion angle, which will induce a further red-shift. In the following, the impact of the dihedral angles on the electronic structure of the chromophore will be analyzed for ensembles of geometries generated by the classical and QM/MM simulations.

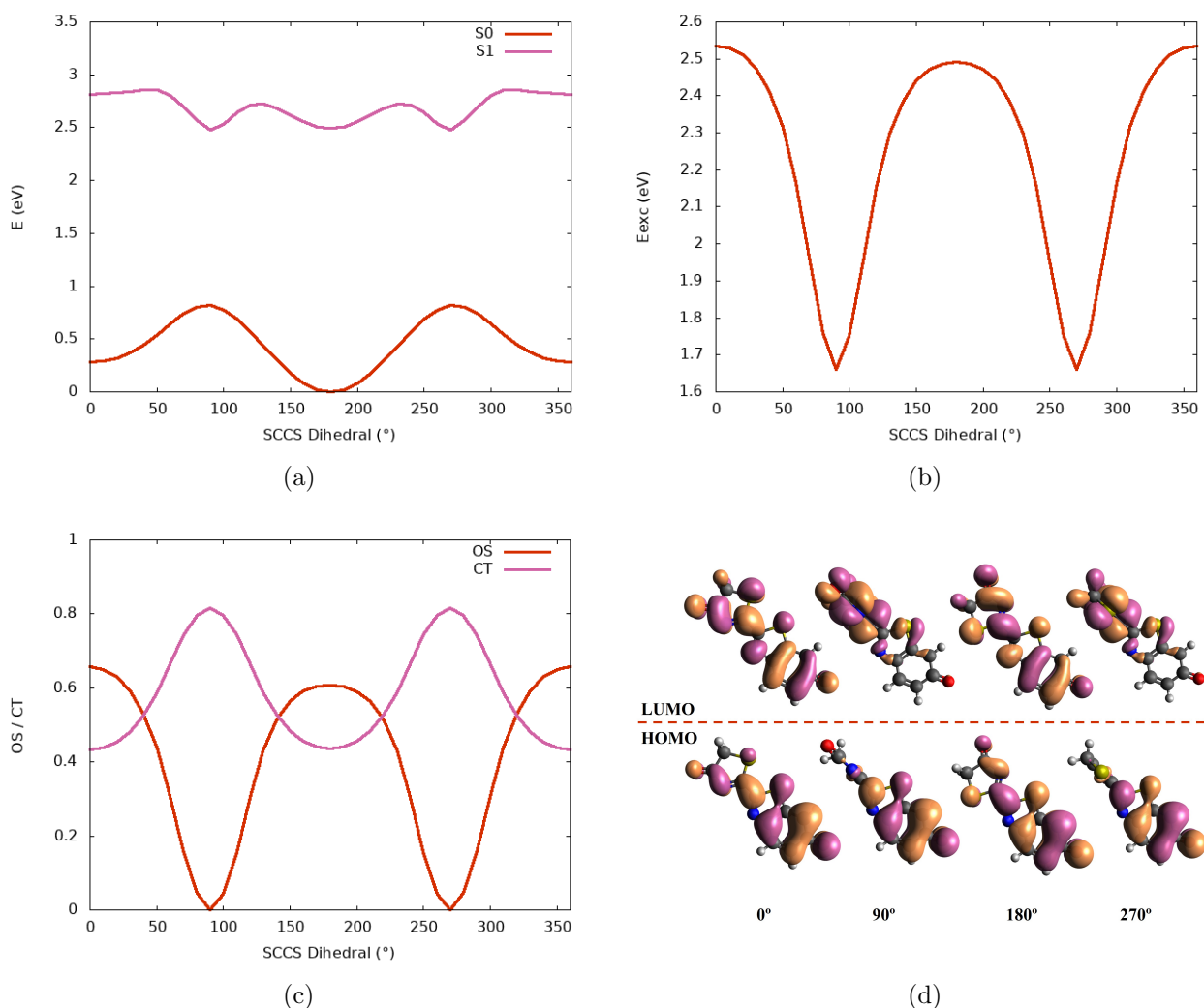


Figure 16: Dependence of some properties of the S_0 - S_1 transition on the SCCS dihedral angle. (a) Energy of the S_0 and S_1 potential energy curve, (b) transition energy, (c) oscillator strength (OS) and charge transfer CT and (d) orbitals involved in the transition.

10.2. Classical Molecular Dynamics

When the SCCS dihedral angle along the last 75 ns of the classical MD trajectories is analyzed, see Figure 17, it can be observed that it mainly oscillates between 150 and 200° in the case of the S_0 surface and between 160 and 210° in the case of the S_1 surface. In both cases, the average is around 180°, as expected. Moreover, it can be observed that the S_0 distribution is wider than the S_1 one, which does not agree with the shape of the QM potential, where the barrier for S_0 is larger than for S_1 . This is maybe a signal that the dihedral potential need to be reparameterized, instead of directly taking the parameters from GAFF.

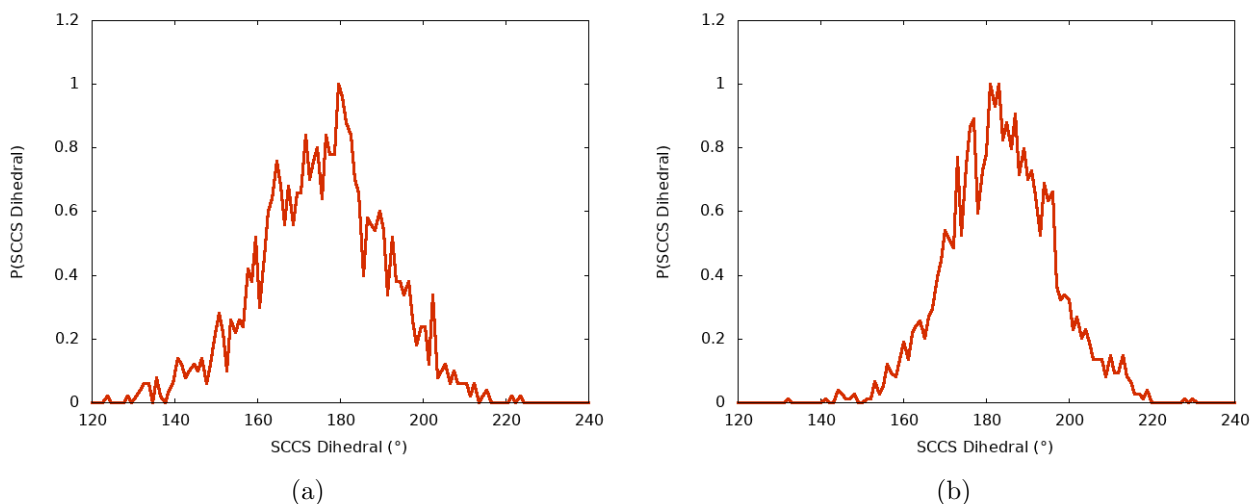


Figure 17: Probability distribution of the SCCS dihedral angle along the classical MD simulations in (a) the S_0 surface and (b) the S_1 surface.

In addition, and taking into account that the spectra is not affected by the sampling method as shown in Section 9, only the geometrical ensemble generated randomly was considered to analyse the electronic properties of the transition depending on the torsion around the C-C single bond along the classical MD simulation. Figure 18 represents the band of the S_0 - S_1 transition, together with its decomposition according to some ranges of the dihedral angle, being the position of the maxima gathered in Table 1. As it can be observed, neither for absorption nor for emission the maxima of the decomposed spectra are significantly shifted. Considering that more important variations were expected after analyzing the static behaviour in the previous section, it is likely that the variation of different coordinates are inducing electronic variations that counteract those coming from the torsion. In addition, the small variations observed in Figure 18 and Table 1 go in the opposite direction to the expected one since, according to the static scan, the excitation energy should decrease when the molecule breaks its planarity, for angles different from 180° . However, in Table 1 one can see that the absorption energy for the angle range 200 - 220° shows a blue-shift when compared with the ranges 160 - 180 and 180 - 200° , that correspond to the situations when the molecule is planar. Moreover, a small blue-shift is also observed in the emission spectrum for the angle range 140 - 160° . This surprising behaviour could be a consequence of two factors: (i) the poor statistics due to a small number number of snapshots that present the above mentioned angles ranges and (ii) the motion of other normal modes that can induce a blue-shift that cancels out the red-shift induced by the torsion motion.

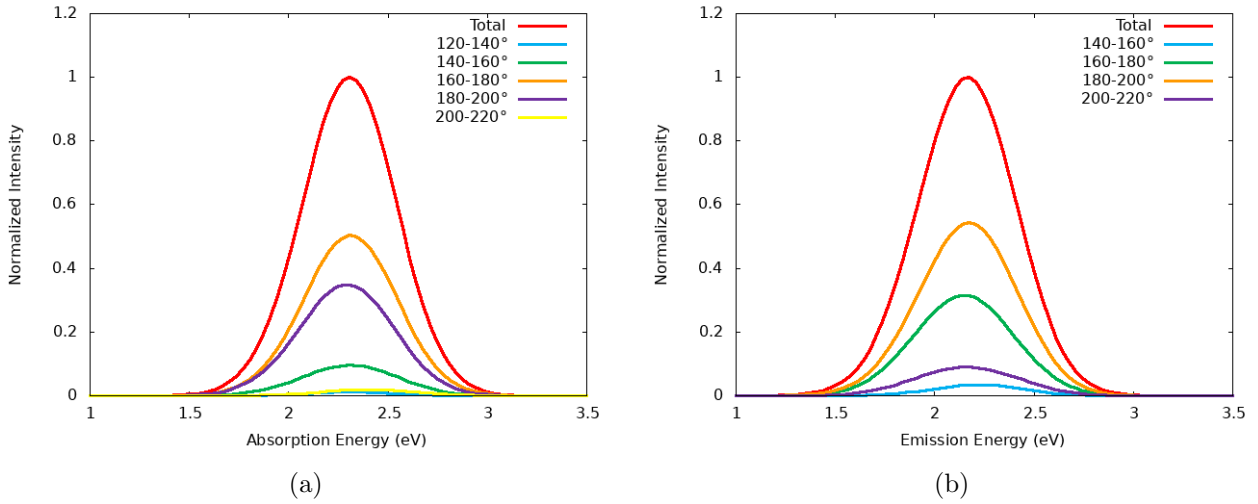


Figure 18: Contributions to the S_0 - S_1 band depending on the SCCS dihedral angle for the classical MD simulation on (a) the S_0 surface and (b) the S_1 surface.

Angle range ($^\circ$)	S_0 MD		S_1 MD	
	Absorption Peak (eV)	CT	Emission Peak (eV)	CT
Total	2.31	0.586	2.17	0.589
120-140	2.31 (1.5%)	0.569	-	-
140-160	2.31 (10%)	0.582	2.21 (4%)	0.587
160-180	2.31 (48.5%)	0.588	2.14 (32%)	0.588
180-200	2.31 (38%)	0.588	2.17 (54%)	0.590
200-220	2.41 (2%)	0.582	2.17 (10%)	0.592

Table 1: Maxima and CT character of the contributions to the S_0 - S_1 band depending on the SCCS dihedral angle. The percentage of conformations in that angle range are shown in parenthesis.

When the CT character of the transitions is analysed, it can be observed that the average value for the specified dihedral angle ranges does not differ a lot from the average value of the whole trajectory, neither for the MD simulation on the S_0 surface nor on the S_1 one. Furthermore, Figure 19 shows that, for each studied dihedral angle range, the width of the CT probability distribution is spread over the same range of angles as the total distribution. Hence, at classical MD level, no correlation can be found between the CT and the SCCS dihedral angle, likely due to the effect of other degrees of freedom on the electronic properties, as mentioned above.

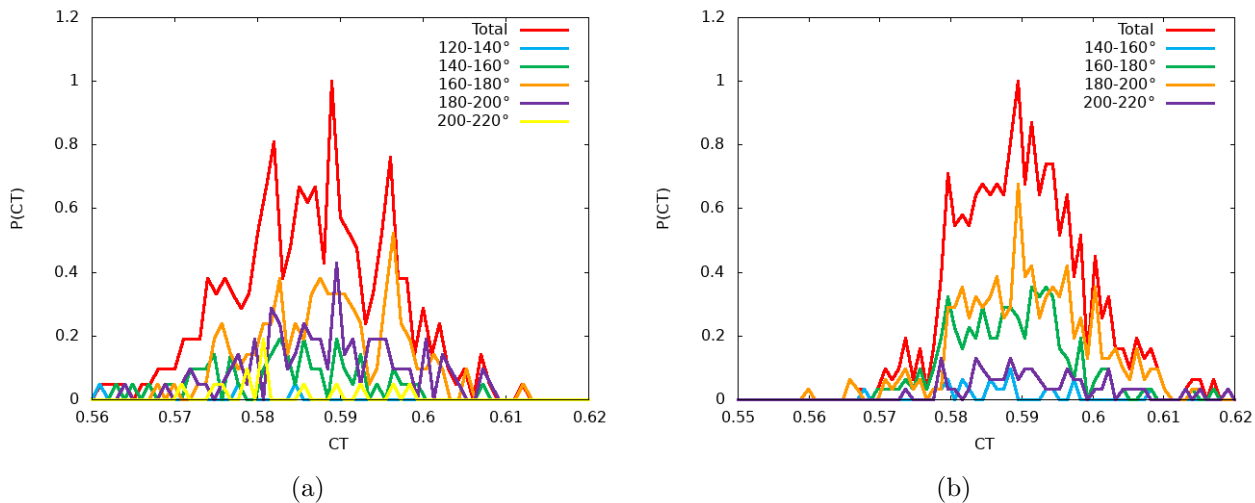


Figure 19: Probability distribution of the CT depending on the SCCS dihedral angle for the MD simulation on (a) the S_0 surface and (b) the S_1 surface.

10.3. QM/MM Molecular Dynamics

The probability distribution for the SCCS dihedral angle was also computed from the QM/MM simulations by considering the last 100 fs from each of the QM/MM trajectories. Figure 20 shows that the simulation on the QM/MM S_0 surface is able to sample the range between 140 and 220° while that on the QM/MM S_1 surface presents a distribution between 130 and 230°. For both surfaces, these dihedral angle ranges are larger than those obtained from the classical MD simulation. Moreover, in contrast to the classical result, this behaviour agrees with the static potential energy scan where the barrier is larger for the S_0 than for the S_1 and, therefore, the angle distribution is narrower for the ground state.

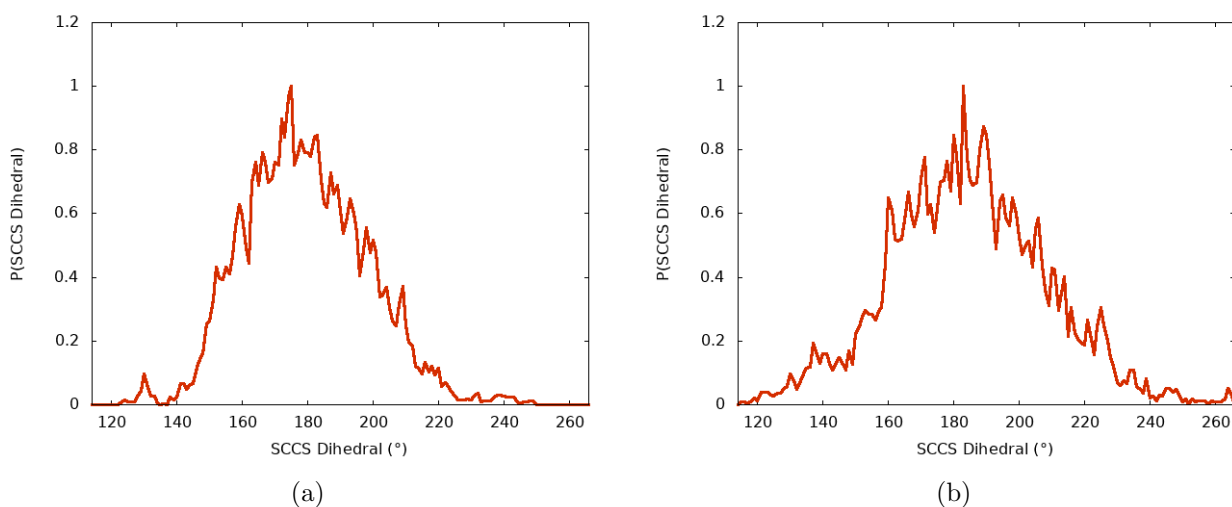


Figure 20: Probability distribution of the SCCS dihedral angle along the hybrid QM/MM MD simulations on (a) the S_0 surface and (b) the S_1 surface.

In addition, the decomposition of the S_0 - S_1 band in ranges of dihedral angles (see Figure 21) shows that, for both absorption and emission spectra, the contributions corresponding to the dihedral angles far from the equilibrium value are shifted with respect to those contributions around it. In the case of the absorption spectra, the red-shift is of 0.27 eV for the range of 200-230° dihedral angle. Moreover, the range between 110 and 140° presents a blue-shift of 0.12 eV with respect to the maxima, which is a unexpected result considering the static behaviour. Further studies are needed to determine whether this result is realistic or not since the sampling of this region is not statistically meaningful or if it corresponds to a contribution of the variation of other coordinates. On the flip side, it is observed that the range between 200 and 230°, which corresponds to a significant amount of conformations, presents a red-shift of 0.22 eV with respect to the range 170-200 (planarity). In contrast to the result from the classical MD trajectory, this behaviour agrees with the static scan since it shows a red-shift when the planarity of the chromophore is broken. The data corresponding to the maxima of the different contributions is gathered in Table 2

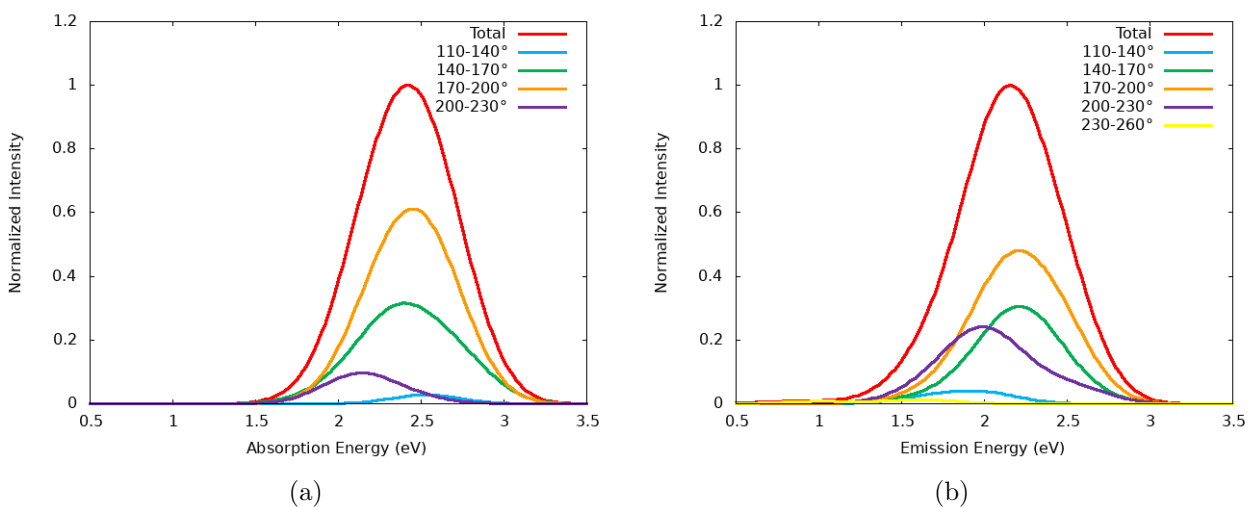


Figure 21: Contributions to the S_0 - S_1 band depending on the SCCS dihedral angle for the QM/MM MD simulation on (a) the S_0 surface and (b) the S_1 surface.

Angle range ($^{\circ}$)	S ₀ MD	S ₁ MD
	Absorption Peak (eV)	Emission Peak (eV)
Total	2.41	2.16
110-140	2.53 (1%)	1.92 (4%)
140-170	2.39 (35%)	2.21 (23%)
170-200	2.45 (55%)	2.21 (47%)
200-230	2.14 (9%)	1.99 (23%)
230-260	-	1.62 (3%)

Table 2: Maxima of the contributions to the S₀-S₁ band depending on the SCCS dihedral angle. The percentage of conformations in that angle range are shown in parenthesis.

10.4. QM vs FF Potentials

To determine how the torsion of the angle contributes to the total potential energy at the two studied levels of theory, classical and CAM-B3LYP/6-311G(2d,p), and therefore how it affects to the sampling of that DOF, the potential energy of the S₀ and S₁ surfaces is represented, together with the analytical expression used by Amber20:

$$V = \frac{PK}{IDIVF} (1 + \cos(PN \cdot \phi - \sigma)) \quad (63)$$

where PK represents half of the height of the barrier, $IDIVF$ is a dividing factor and PN and ϕ correspond to the periodicity of the function and the dihedral angle, respectively. This parameters, for this specific torsion, were considered to be 4.0 kcal/mol, 4, 2.0 and 180.0, correspondingly. Figure 22 shows that the classical barrier (0.043 eV) is much lower than the barriers of the S₀ (0.821 eV) and S₁ (0.234 eV) states. Thus, in contrast to the results of the probability distributions of the dihedral angles, the classical simulation should be able to sample a wider range of angles than the QM/MM due to the lower energy barrier.

The FF is compared with S₀ and S₁ PESs since it has not been reparameterized and therefore is the same for both simulations. Nevertheless, the maxima of the analytical function do not coincide with those of the S₁ surface.

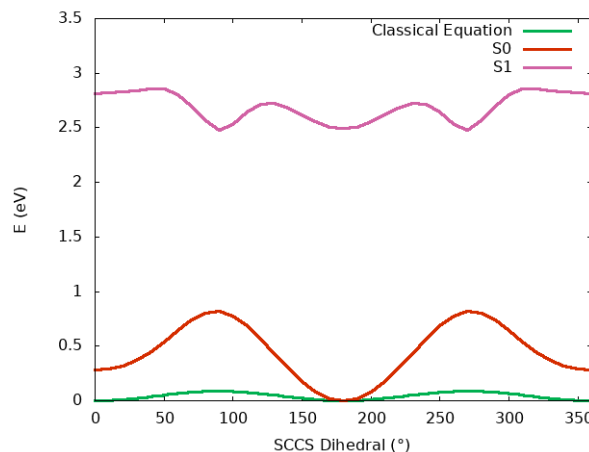


Figure 22: Comparison of the CAM-B3LYP PES around the SCCS torsion with the classical expression of the FF.

11. Classical MD vs QM/MM MD Based Spectra

Within this section, the QM/MM MD spectra will be first discussed. Then, they will be compared with those obtained from the classical MD trajectories.

11.1. QM/MM Spectra

The absorption spectrum, represented in Figure 23(a), presents again two bands, corresponding to the $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_3$ transitions, peaking at 2.49 and 3.31 eV, respectively. However, in this case, the intensity of the $S_0 \rightarrow S_3$ band presents a lower intensity than that of the spectrum obtained from the classical MD. With regards to the mixing of the S_0 with the S_2 state and the swapping between them, it has been observed that 38% of the studied conformations present a significant contribution of both the HOMO - LUMO and HOMO-1 - LUMO transitions, which corresponds to the situation in which both states are mixed. Moreover, the HOMO - LUMO bright transition only contributes to the first band in a 80% of the conformations which means that, for 20% of the conformations a swapping between the states has been produced and therefore, the only contribution to the $S_0 \rightarrow S_1$ band corresponds to the HOMO-1 - LUMO dark transition. In the case of the emission spectrum, whose peak is at 2.16 eV, the mixing between states appears in 29% of the studied geometries, while the switching in 15% of the cases. The S_2 is considered to be mixed with the S_1 , when the HOMO-1 - LUMO transition (the main contribution of S_2) contributes to the S_1 more than 0.2 to the total wavefunction.

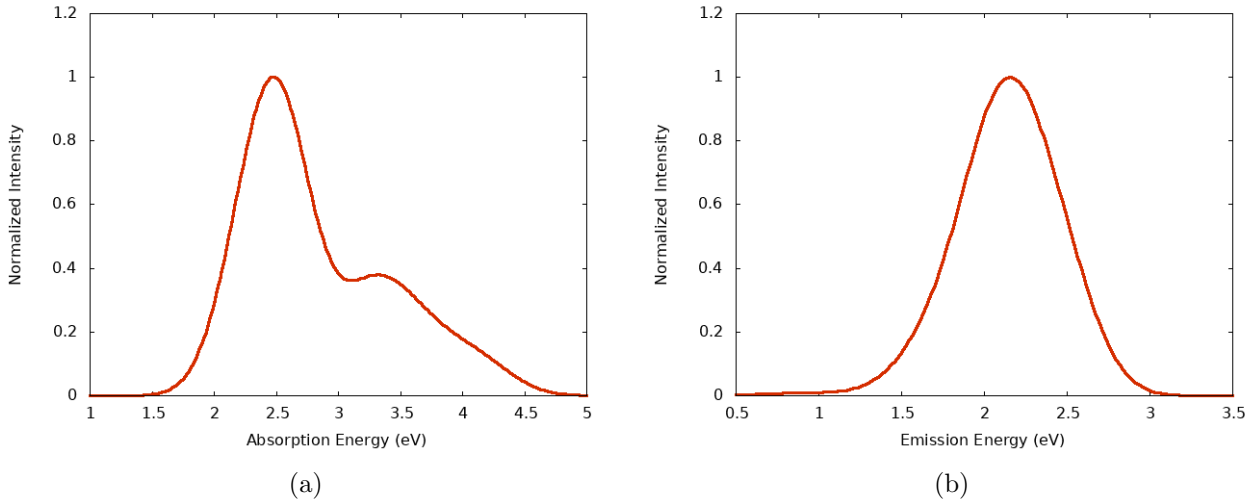


Figure 23: (a) Absorption and (b) emission spectra of the oxyluciferin chromophore within the luciferase enzyme obtained from the hybrid QM/MM simulations.

11.2. Comparison of the MD Simulations

This section analyses the influence of the level of theory used to computed the PES (a force field or QM/MM) on the studied electronic properties. As it can be observed in Table 3, the absorption spectra computed for the QM/MM trajectory presents a blue-shift (0.15 eV in the case of the $S_0 \rightarrow S_1$ transition and 0.23 eV for the $S_0 \rightarrow S_3$) with respect to that of the classical trajectory. This means that the S_0 of the system is more stabilized in the QM/MM PES than in the classical one. However, it also exists the possibility that S_1 - S_2 flipping, negligible in the classical trajectory (2%) but significantly larger in the case of the QM/MM one (15%), is produced due to a larger destabilization of the S_1 state and therefore, the transition energy increases. In the case of the emission, the value of the peak is very similar in both classical and QM/MM spectra. However, this is fortuitous because classically the dihedral potential is the same for S_0 and S_1 . Therefore, there is no reason to expect for a better agreement for S_1 than for S_0 . Furthermore, the S_1 - S_2 mixing values are very similar for both classical and QM/MM trajectories.

	S ₀ MD		S ₁ MD		
	Absorption Peaks (eV)	S ₁ -S ₂ M/F	Emission Peak (eV)	Emission Peak (nm)	S ₁ -S ₂ M/F
Classical MD	2.34/3.08	36/2%	2.17	565	35/2%
QM/MM MD	2.49/3.31	38/20%	2.16	579	29/15%

Table 3: Comparison of the transition energies and the S_1 - S_2 mixing (M) and flipping (F) between the MD of the two sampled PESs.

Additionally, the computed emission spectra are compared with the experimental one, [80] that peaks at 552 nm, as shown in Figure 24. For this, the spectra, shown in Figure 24 have been represented. It can be observed that the spectra calculated from the classical geometries results in a more similar band than that computed from the QM/MM trajectory. This similarity is present not only on the position of maxima but also on the shape of the band since the QM/MM band presents an shoulder at higher energies and an abnormal tail. This feature of the QM/MM spectrum, that was not observed in Figure 23(b) due to the eV and nm scales not being directly proportional, can be due to an unrealistic torsion of the SCCS dihedral along the QM/MM MD simulations, or to the motion of a different internal coordinate not analyzed here. Other reasons could be a wrongly described interaction between some residues of the protein and the chromophore or an over-stabilization of the CT states due to the use of electrostatic embedding that does not allow the polarization of the environment due to the charges of the chromophore.

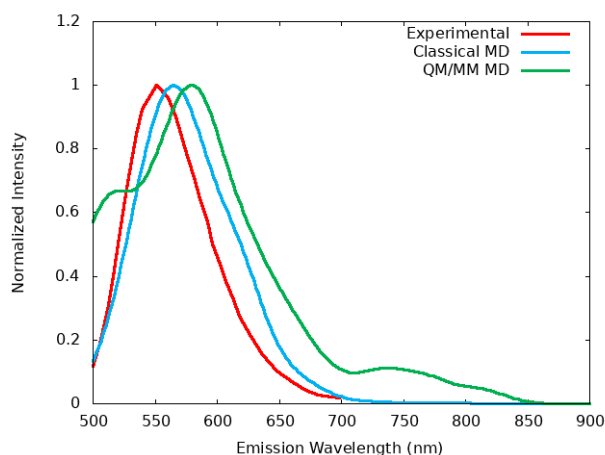


Figure 24: Comparison of the experimental emission spectrum with the calculated with the two different potential energy models.

In addition, Figure 21(b) is represented in the nm scale to determine if the odd shape of the QM/MM emission spectrum is related with the wider torsion distribution of the simulation. However, as it can be observed in Figure 25, all the represented contributions present an apparently random peak distribution. Thus, no correlation can be directly found between the shape and the torsion, apart from the previously commented position of the absolute maxima. Moreover, it can be observed in the enlarged part that the infrared region of the tail presents contributions from all the analysed dihedral angle ranges. Therefore, it is possible that other features, either geometric, electrostatic or electronic, are participating in the red-shifted tail. Further analysis are necessary to clarify this issue.

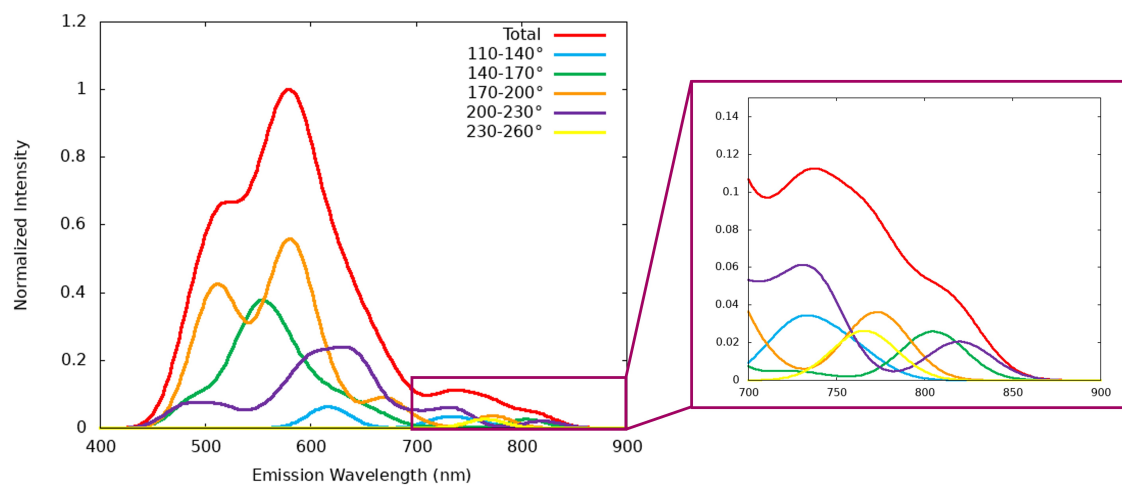


Figure 25: Emission spectrum calculated from the sampled geometries of the QM/MM surface, as well as its contributions to the S_0 - S_1 band depending on the SCCS dihedral angle.

Conclusions

During this work, classical and QM/MM simulations were carried out to analyse the effect of the potential energy model employed in the dynamics on the absorption and emission spectra and geometrical and electronic features. In addition, the convergence of the spectra with respect to the number of geometries is analyzed for three different geometrical selection approaches: random, equidistant and Metropolis selection. On the light of these analyses, the following conclusions have been obtained:

1. The three studied sampling methods result in a similar convergence of the maxima of the S_0 - S_1 band. However, random and Metropolis samplings perform better than equidistant sampling since the latter presents stronger oscillations. Moreover, the convergence of the absorption spectrum is better than that of the emission spectrum, for the three sampling methods.
2. The absorption and emission spectra computed from the classical MD geometries selected by the three different methods are virtually the same although the differences in the emission spectrum are slightly larger due to, likely, the worse convergence of the sampling.
3. The absorption spectra calculated from the classical and the QM/MM trajectories on the S_0 PESs present a very similar shape, involving the peaks corresponding to the $S_0 \rightarrow S_1$ and the $S_0 \rightarrow S_3$ transitions. However, the relative intensity of the latter is lower in the case of the QM/MM trajectory. Moreover, the spectrum from the QM/MM trajectory is blue-shifted with respect to that from the classical trajectory.
4. In the case of the emission, it was observed that the maxima of both computed spectra peak at a very similar wavelength than that of the experimental result. However, the shape of the emission band from the QM/MM trajectory is very different from the experimental one. The computed band presents a high energy shoulder and a low energy tail that do not appear in the experimental spectrum. To determine whether the SCCS dihedral is related with these two features of the spectra, the spectrum was divided into contributions from different ranges of this angle. However, no correlation could be found since all the considered ranges contribute to the tail of the spectra and the three closest to equilibrium, to the shoulder.

5. A further analysis of the S_0 - S_1 transition showed that, along both classical and QM/MM simulations, the S_1 state is mixed with S_2 leading to contributions of the dark $n\pi^*$ transition to the lowest-energy band. Moreover, the QM/MM trajectory also showed a non-negligible swapping between these states.
6. The static QM analysis of the electronic transition properties resulting from the torsion around the C—C bond showed that they drastically change with the dihedral angle. When the planarity of the molecule is broken, the electronic excitation energy and the oscillator strength decrease and the CT character increases due to a localization of the electronic density on different regions of the molecule.
7. The probability distribution of the dihedral angle in the classical MD simulation of the S_0 PES is larger than that of the S_1 , which is not in agreement with the shape of the QM potential since the barrier for the S_0 is larger than for the S_1 . This fact might indicate the need of the reparametrization of the force field to describe better the dihedral potential.
8. When the S_0 - S_1 band is decomposed according to different ranges of the dihedral angle, it is observed that the absorption and emission peaks corresponding to the bands computed from the classical trajectory presented a blue-shift when the planarity of the chromophore is broken. This result does not agree with the expected behaviour from the static scan since the excitation energy should decrease when the dihedral angle is not at the equilibrium value. A possible explanation for this is the poor statistics in the regions far from the 180° dihedral angle or the counteracting effect of other normal modes of the system. In the case of the spectra calculated from the QM/MM simulation, the expected red-shift behaviour is observed.
9. It was observed that the QM/MM simulation was able to sample a wider range of the SCCS dihedral angle compared with the classical simulation. On this behalf, the classical expression of the FF for this angle has been compared with the CAM-B3LYP rigid scan of the torsion. This analysis leads to a contradictory result since the barrier of the classical PES is lower than that calculated QM. However, it is important to realize that the angle distribution from the classical simulation does not only depend on the dihedral potential but also on other potential contributions that are not being considered in the analysis.

Outlook

On the light of this results, it is clear that a lot of work still needs to be done to elucidate the factors that affect the emission processes of the oxyluciferin/luciferase system.

It would be interesting to further analyse the sampled geometries on both classical and QM/MM trajectories to determine the reason of the blue-shift when one goes from the classical to the QM/MM approach. Moreover, the interactions with the environment should also be analysed.

In addition, the reason why the classical MD does not reproduce the expected behaviour of either the width of the dihedral angle probability distribution nor the energy shift of the different contributions to the S_0 - S_1 band depending on the SCCS angle need to be investigated. To do so, more geometries of the classical trajectories will be considered in order to obtain a good statistic for those contributions corresponding to the regions far from equilibrium. Furthermore, the dihedral angle will be reparameterized to determine whether a wrong description of torsion is responsible of the obtained behaviour or not

The reason behind the bad description of the shape of the emission band in the case of the QM/MM trajectory should also be clarified. For this purpose, a different DFT functional or a better QM method, instead of DFT, could be used to elucidate whether the wide dihedral angle distribution obtained from QM/MM MD is realistic or not. Moreover, this also should correct the possible wrong description of other internal coordinates. The high-energy shoulder could appear due to an overstabilization of the CT states as a result of the electrostatic embedding description of the environment. To check if this is the case, some of the geometries contributing to that excitation energy will be recalculated without the environmental description. These results should enlighten whether it is a geometric problem or a wrong description of the chromophore/protein interactions. If this were the situation, some residues could be included in the QM region. In this case, the B3LYP functional should not be used since it is not able to properly describe long range interactions. Therefore, a long range corrected version, such as CAM-B3LYP, should be used instead. Other solution would be the use of a polarizable embedding description of the environment.

In addition, further studies will focus on analysing the chromophore/protein binding energy using free-energy method, such as the Poisson-Boltzmann (MM-PBSA) or generalized Born (MM-GBSA), to determine which residues contribute more to the binding process. With the aim of inducing a red-shift in the emission spectrum, these residues will be mutated.

Finally, and as was previously commented, it would be interesting to mutate some protein residues in order to induce a larger torsion of the SCCS dihedral angle since that modification by itself should induce a red shift.

Bibliography

- (1) Bitler, B.; McElroy, W. D. *Arch. Biochem. Biophysics*. **1957**, *72*, 358–368.
- (2) Fleiss, A.; Sarkisyan, K. S. *Curr. Genet.* **2019**, *65*, 877–882.
- (3) Dukhovich, A.; Sillero, A.; Günther Sillero, M. A. *FEBS Lett.* **1996**, *395*, 188–190.
- (4) Branchini, B. R.; Southworth, T. L.; Murtiashaw, M. H., et al. *Biochemistry* **2004**, *43*, 7255–7262.
- (5) Branchini, B. R.; Magyar, R. A.; Murtiashaw, M. H., et al. *Biochemistry* **2001**, *40*, 2410–2418.
- (6) Fraga, H.; Fernandes, D.; Novotny, J., et al. *ChemBioChem* **2006**, *7*, 929–935.
- (7) Marques, S. M.; Esteves Da Silva, J. C. Firefly bioluminescence: A mechanistic approach of luciferase catalyzed reactions, 2009.
- (8) Branchini, B. R.; Behney, C. E.; Southworth, T. L., et al. *J. Am. Chem. Soc.* **2015**, *137*, 7592–7595.
- (9) Yue, L.; Liu, Y. J.; Fang, W. H. *J. Am. Chem. Soc.* **2012**, *134*, 11632–11639.
- (10) Pinto Da Silva, L.; Santos, A. J. M.; Esteves Da Silva, J. C. *J. Phys. Chem. A* **2013**, *117*, 94–100.
- (11) Song, C. I.; Rhee, Y. M. *J. Am. Chem. Soc.* **2011**, *133*, 12040–12049.
- (12) Sundlov, J. A.; Fontaine, D. M.; Southworth, T. L., et al. *Biochemistry* **2012**, *51*, 6493–6495.
- (13) Navizet, I.; Liu, Y.-J.; Ferré, N., et al. *ChemPhysChem* **2011**, *12*, 3064–3076.
- (14) Senyilmaz, D.; Teleman, A. A. *F1000Prime Rep.* **2015**, *7*, 1–13.
- (15) Chen, X.; Qian, Y.; Wu, S. *Free Radic. Biol. Med.* **2015**, *79*, 253–263.
- (16) Kaskova, Z. M.; Tsarkova, A. S.; Yampolsky, I. V. *Chem. Soc. Rev.* **2016**, *45*, 6048–6077.
- (17) Nogueira, J. J.; González, L. *Annu. Rev. Phys. Chem.* **2018**, *69*, 473–497.
- (18) Nogueira, J. J.; Roßbach, S.; Ochsenfeld, C., et al. *J. Chem. Theory Comput.* **2018**, *14*, 4298–4308.
- (19) Kasha, M.; Rawls, H. R.; El-Bayoumi, M. A. *Pure Appl. Chem.* **1965**, *11*, 371–392.
- (20) Kasha, M. In *Spectrosc. Excit. State*; Springer US: 1976, pp 337–363.

- (21) Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341.
- (22) Grotendorst, J.; Marx, D.; Muramatsu, A., *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms - Lecture Notes*; NIC-Directors: 2002; Vol. 10.
- (23) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N., et al. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (24) Seminario, J. M. *Int. J. Quantum Chem.* **1996**, *60*, 1271–1277.
- (25) Sun, X.; Lin, T.; Gezelter, J. D. *J. Chem. Phys.* **2008**, *128*, 1352.
- (26) Davidchack, R. L.; Handel, R.; Tretyakov, M. V. *J. Chem. Phys.* **2009**, *130*, 234101.
- (27) Åqvist, J.; Wennerström, P.; Nervall, M., et al. *Chem. Phys. Lett.* **2004**, *384*, 288–294.
- (28) Borondo Rodríguez, F.; Merchant Bonete, M. M.; Sánchez Marcos, E., et al., *Química teórica y computacional*; Andrés, J., Beltrán, J., Eds.; Publicacions de la Universitat Jaume I: 2013.
- (29) Monticelli, L.; Tieleman, D. P. In *Biomol. Simulations Methods Protoc.* Springer: 2013, pp 197–213.
- (30) Case, D.; Aktulga, H.; Belfon, K., et al. **2020**.
- (31) Verlet, L. *Phys. Rev.* **1967**, *159*, 98.
- (32) Chandler, D., *Introduction to Modern Statistical Mechanics*; Oxford University Press, Inc.: 1987, p 288.
- (33) Frenkel, D.; Smit, B., *Understanding molecular simulation : from algorithms to applications*; Academic Press: 1996; Vol. 50.
- (34) Jensen, F., *Introduction to Computational Chemistry*, Second; John Wiley & Sons Ltd, Ed., 2007.
- (35) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (36) Frenkel, D.; Smit, B., *Understanding molecular simulation : from algorithms to applications*; Academic Press: 1996; Vol. 50.
- (37) Jorgensen, W.; Chandrasekhar, J.; Madura, J., et al. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (38) Miyamoto, S.; Kollman, P. *J. Comput. Chem.* **1992**, *13*, cited By 4794, 952–962.
- (39) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24–34.
- (40) Senn, H. M.; Thiel, W. *Angew. Chemie - Int. Ed.* **2009**, *48*, 1198–1229.

- (41) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.
- (42) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700–733.
- (43) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.
- (44) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.
- (45) Hofmann, M.; Schaefer III, H. F. In *Encycl. Phys. Sci. Technol.* Meyers, R. A., Ed.; Academic Press: 2003, pp 487–506.
- (46) Cramer, C. J., *Essentials of Computational Chemistry, Theories and Models*, Second; John Wiley & Sons Ltd, Ed., 2004, pp 249–300.
- (47) Runge, E.; Gross, E. K. *Phys. Rev. Lett.* **1984**, *52*, 997.
- (48) Gross, E. K. U.; Dobson, J. F.; Petersilka, M. In *Density Functional Theory II: Relativistic and Time Dependent Extensions*, Nalewajski, R. F., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1996, pp 81–172.
- (49) Vosko, S.; Wilk, L.; Nusair, M. *Can. J. Phys* **1980**, *58*, 1200.
- (50) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (51) Miehlich, B.; Savin, A.; Stoll, H., et al. *Chem. Phys. Lett.* **1989**, *157*, 200–206.
- (52) Harris, J. *Phys. Rev. A* **1984**, *29*, 1648.
- (53) Becke, A. D. *J. Chem. Phys* **1993**, *98*, 5648–5652.
- (54) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F., et al. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (55) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (56) Talman, J. D.; Shadwick, W. F. *Phys. Rev. A* **1976**, *14*, 36.
- (57) Bonetti, A. F.; Engel, E.; Schmid, R., et al. *Phys. Rev. Lett.* **2001**, *86*, 2241.
- (58) Bartlett, R. J.; Lotrich, V. F.; Schweigert, I. V. *J. Chem. Phys.* **2005**, *123*, 062205.
- (59) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (60) Szabo, A.; Ostlund, N. L., *Modern Quantum Chemistry: Introduction to advanced electronic structure theory*; Dover Publications, Inc.: Mineola, New York, 1996.
- (61) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W., et al. *J. Comput. Chem.* **1983**, *4*, 294–301.
- (62) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724–728.

- (63) Francl, M. M.; Pietro, W. J.; Hehre, W. J., et al. *J. Chem. Phys.* **1982**, *77*, 3654–3665.
- (64) Gordon, M. S.; Binkley, J. S.; Pople, J. A., et al. *J. Am. Chem. Soc.* **1982**, *104*, 2797–2803.
- (65) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (66) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (67) Spitznagel, G. W.; Clark, T.; Schleyer, P. v. R., et al. *J. Comput. Chem.* **1987**, *8*, 1109–1116.
- (68) Krishnan, R.; Binkley, J. S.; Seeger, R., et al. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (69) McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639–5648.
- (70) Schrödinger, L.; DeLano, W. PyMOL, version 2.4.0, 2020.
- (71) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B., et al. **2016**, Gaussian Inc. Wallingford CT.
- (72) Tian, C.; Kasavajhala, K.; Belfon, K. A., et al. *J. Chem. Theory Comput.* **2019**, *16*, 528–552.
- (73) Wang, J.; Wolf, R. M.; Caldwell, J. W., et al. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (74) García-Iriepa, C.; Zemmouche, M.; Ponce-Vargas, M., et al. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4613–4623.
- (75) García-Iriepa, C.; Gosset, P.; Berraud-Pache, R., et al. *J. Chem. Theory Comput.* **2018**, *14*, 2117–2126.
- (76) Cárdenas, G.; Lucia-Tamudo, J.; Mateo-delaFuente, H., et al. *ChemRxiv* **2022**, DOI: [10.26434/chemrxiv-2022-cnnp2](https://doi.org/10.26434/chemrxiv-2022-cnnp2).
- (77) Roe, D. R.; Cheatham III, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.
- (78) Plasser, F. *J. Chem. Phys.* **2020**, *152*, 084108.
- (79) Weingart, O.; Nenov, A.; Altoè, P., et al. *J. Mol. Model.* **2018**, *24*, 1–30.
- (80) Mofford, D. M.; Reddy, G. R.; Miller, S. C. *J. Am. Chem. Soc.* **2014**, *136*, 13277–13282.

Annexes

I. Derivations

I.I. Hamilton Equations

$$\frac{d\mathbf{p}_i}{dt} = -\frac{dV}{d\mathbf{r}_i} \quad \frac{dT_i}{d\mathbf{p}_i} = \frac{d\mathbf{r}_i}{dt} \quad (\text{I})$$

Being Newton's Second Law:

$$\mathbf{F}_i = \frac{d\mathbf{p}_i}{dt} \quad (\text{II})$$

and knowing that, for conservative systems:

$$\mathbf{F}_i = -\frac{dV}{d\mathbf{r}_i} \quad (\text{III})$$

Then, the first Hamilton equation of Equation [I](#) is simply obtained by equaling Equations [II](#) and [III](#). In the case of the second Hamilton equation, it is obtained by deriving the expression of the kinetic energy with respect to the momentum:

$$\frac{dT_i}{d\mathbf{p}_i} = \frac{d}{d\mathbf{p}_i} \left(\frac{\mathbf{p}_i^2}{2m_i} \right) = \frac{2\mathbf{p}_i}{2m_i} = \frac{m_i \mathbf{v}_i}{m_i} = \mathbf{v}_i = \frac{d\mathbf{r}_i}{dt} \quad (\text{IV})$$

I.II. Velocity Verlet Algorithm

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 \quad \mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{1}{2}(\mathbf{a}_i(t) + \mathbf{a}_i(t + \Delta t))\Delta t \quad (\text{V})$$

As commented before, the velocity Verlet algorithm is an expansion based method. Therefore, its equations are directly obtained by performing a Taylor expansion around $\mathbf{r}_i(t)$ and $\mathbf{v}_i(t)$ for coordinates and velocities, respectively.

In the case of the coordinates:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \frac{d\mathbf{r}_i(t)}{dt} + \frac{\Delta t^2}{2!} \frac{d^2\mathbf{r}_i(t)}{dt^2} + \frac{\Delta t^3}{3!} \frac{d^3\mathbf{r}_i(t)}{dt^3} + \dots \quad (\text{VI})$$

Hence, truncating up to second order and identifying the derivatives of the first and second order terms in Equation [VI](#) with the velocity and the acceleration, respectively, the equation of the propagation of the coordinates is directly obtained.

Doing the same for the velocities:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \Delta t \frac{d\mathbf{v}_i(t)}{dt} + \frac{\Delta t^2}{2!} \frac{d^2\mathbf{v}_i(t)}{dt^2} + \frac{\Delta t^3}{3!} \frac{d^3\mathbf{v}_i(t)}{dt^3} + \dots \quad (\text{VII})$$

However, in this case it is not possible to identify the second order term of the expansion. An expression for it can be obtained by doing another Taylor expansion around $\frac{d\mathbf{v}_i(t)}{dt}$:

$$\frac{d\mathbf{v}_i(t + \Delta t)}{dt} = \frac{d\mathbf{v}_i(t)}{dt} + \Delta t \frac{d^2\mathbf{v}_i(t)}{dt^2} + \frac{\Delta t^2}{2!} \frac{d^3\mathbf{v}_i(t)}{dt^3} + \dots \quad (\text{VIII})$$

Truncating Equation [VIII](#) up to second term, solving it for the first order term and substituting it in Equation [VII](#), truncated to the second order term; then, it is obtained that:

$$\begin{aligned} \mathbf{v}_i(t + \Delta t) &= \mathbf{v}_i(t) + \Delta t \frac{d\mathbf{v}_i(t)}{dt} + \frac{\Delta t}{2} \frac{d\mathbf{v}_i(t + \Delta t)}{dt} - \frac{\Delta t}{2} \frac{d\mathbf{v}_i(t)}{dt} \\ &= \mathbf{v}_i(t) + \frac{\Delta t}{2} \frac{d\mathbf{v}_i(t + \Delta t)}{dt} + \frac{\Delta t}{2} \frac{d\mathbf{v}_i(t)}{dt} \\ &= \mathbf{v}_i(t) + \frac{\Delta t}{2} \left(\frac{d\mathbf{v}_i(t + \Delta t)}{dt} + \frac{d\mathbf{v}_i(t)}{dt} \right) \end{aligned} \quad (\text{IX})$$

In this equation, the terms in the parenthesis can be easily identified as the acceleration of the step and the acceleration of the previous step, respectively; thus obtaining the equation for the propagation of the velocities of Equation [V](#).

I.III. Instantaneous Temperature

The Maxwell-Boltzmann distribution of molecular speeds, that it agrees with the equipartition result, and so can be used to determine the average kinetic energy of a particle in a gas is given by:

$$P(\mathbf{v}_i) = 4\pi \left(\frac{m_i\beta}{2\pi} \right)^{\frac{3}{2}} \mathbf{v}_i^2 e^{-\frac{m_i\mathbf{v}_i^2\beta}{2}} \quad (\text{X})$$

where $\beta = (k_B T)^{-1}$. Therefore, the average kinetic energy of a particle can be expressed as:

$$\langle K_i \rangle = \left\langle \frac{1}{2} m_i \mathbf{v}_i^2 \right\rangle = \int_0^\infty \frac{1}{2} m_i \mathbf{v}_i^2 P(\mathbf{v}_i) d\mathbf{v}_i \quad (\text{XI})$$

Substituting Equation [X](#) into Equation [XI](#):

$$\langle K_i \rangle = \frac{1}{2} m_i 4\pi \left(\frac{m_i\beta}{2\pi} \right)^{\frac{3}{2}} \int_0^\infty \mathbf{v}_i^4 e^{-\frac{m_i\mathbf{v}_i^2\beta}{2}} d\mathbf{v}_i \quad (\text{XII})$$

Knowing that the result of the integral is given by:

$$\int_0^\infty \mathbf{v}_i^4 e^{-\frac{m_i \mathbf{v}_i^2 \beta}{2}} d\mathbf{v}_i = \frac{3}{2} \left(\frac{k_B T}{m_i} \right)^2 \left(\frac{2\pi k_B T}{m_i} \right)^{\frac{1}{2}} \quad (\text{XIII})$$

Then:

$$\langle K_i \rangle = \frac{1}{2} m_i 4\pi \left(\frac{m_i \beta}{2\pi} \right)^{\frac{3}{2}} \frac{3}{2} \left(\frac{k_B T}{m_i} \right)^2 \left(\frac{2\pi k_B T}{m_i} \right)^{\frac{1}{2}} = \frac{3}{2} k_B T \quad (\text{XIV})$$

On the other hand, the kinetic energy is usually expressed as:

$$\langle K \rangle = \sum_{i=1}^N \frac{1}{2} m_i \cdot \mathbf{v}_i^2 \quad (\text{XV})$$

Hence, combining Equations XIV for all the atoms of the system and XV:

$$\langle K \rangle = \sum_{i=1}^N \frac{3}{2} k_B T = \sum_{i=1}^N \frac{1}{2} m_i \cdot \mathbf{v}_i^2 \quad (\text{XVI})$$

and solving for T:

$$T = \frac{\sum_{i=1}^N m_i \cdot \mathbf{v}_i^2}{3Nk_B} = \frac{\sum_{i=1}^N m_i \cdot \mathbf{v}_i^2}{k_B N_{DOF}} \quad (\text{XVII})$$

II. Metropolis Criteria Code

```

1 program metropolis
2 implicit none
3
4 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
5 ! This code determines whether a given frame of a MD trajectory satisfies
6 ! the Metropolis criteria form the value of the total potential energy of
7 ! the system.
8 !
9 ! It takes a file containing 4 columns: the frame, the total energy, the
10 ! kinetic energy and the potential energy and, if the specific frame
11 ! satisfies the Metropolis criteria, it prints the number of the frame.
12 !
13 ! Additionally, it requires the average of the potential energy of the
14 ! system in and the average temperature of the simulation. These two
15 ! parameters need to be specified before compilation.
16 !

```

```

17 ! Units:
18 !     Energies: Kcal/mol
19 !     Temperature: K
20 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
21
22 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
23 ! Variable declaration
24 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
25
26 real*8, parameter :: kb=0.001987204259d0 ! Kcal/mol*K
27
28 real*8           :: ener(2000)           ! Potential energy of the system
29 real*8           :: enerav=-192460.0929 ! Average potential energy
30 real*8           :: e , k                ! Total energy, kinetic energy
31 real*8           :: temp=303.17          ! Temperature
32 integer          :: fram(2000)           ! Frame of the MD
33
34 real*8           :: rand                 ! Random number
35 real*8           :: expo                 ! Result of the exponent
36
37 integer          :: i                    ! Indexes
38 character(len=100) :: string_lec        ! To read input data
39
40 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
41
42 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
43 ! Read file
44 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
45
46 open (unit = 7, file = "eners_din_cutted2000_relE.dat")
47 read(7,*)
48 do i=1, 2000
49     read(7,*) fram(i), e, k, ener(i)
50 enddo
51 close(7)
52
53 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
54
55 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
56 ! Program

```

[illegible]

Code 1: Metropolis Criteria Code.