

pH-Dependent Absorption Spectrum of Oxyluciferin Analogues in the Presence of Adenosine Monophosphate

Nuno Manuel de Almeida Barbosa, Madjid Zemmouche, Pauline Gosset, Cristina García-iriepa, Vincent Ledentu, Isabelle Navizet, Pascal Didier, Nicolas Ferré

▶ **To cite this version:**

Nuno Manuel de Almeida Barbosa, Madjid Zemmouche, Pauline Gosset, Cristina García-iriepa, Vincent Ledentu, et al.. pH-Dependent Absorption Spectrum of Oxyluciferin Analogues in the Presence of Adenosine Monophosphate. ChemPhotoChem, 2019, 10.1002/cptc.201900150 . hal-02409332

HAL Id: hal-02409332
https://amu.hal.science/hal-02409332

Submitted on 13 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

pH-dependent absorption spectrum of oxyluciferin analogues in the presence of adenosine monophosphate

Dr. Nuno Manuel de Almeida Barbosa[*†] Madjid Zemmouche[‡]
Pauline Gosset[§] Dr. Cristina García-Iriepa[‡¶]
Dr. Vincent Ledentu[†] Prof. Isabelle Navizet[‡]
Prof. Pascal Didier[§] Prof. Nicolas Ferré[*†]

Abstract

The photophysical properties of oxyluciferin, the light emitter responsible for firefly bioluminescence, are pH-dependent. One of the potential proton acceptor/donor is adenosine monophosphate (AMP). We have studied three oxyluciferin synthetic analogues with or without AMP, in water, in the pH=5 to 11 range, using both experimental steady-state absorption spectroscopy or the recently developed computational protocol that uses constant pH molecular dynamics and then hybrid QM/MM calculations (CpHMD-then-QM/MM). The latter features a systematic investigation of all the protonation microstates using molecular dynamics simulations coupled to thousands hybrid QM/MM vertical excitation energies. Our results demonstrate that AMP does not significantly modify the visible light absorption of the analogues, whatever the pH value. We also show that CpHMD-then-QM/MM is capable to qualitatively reproduce the pH-dependent absorption spectrum of the analogues, despite the employed low QM level of theory.

[*]corresponding author
[†]Aix-Marseille Univ, CNRS, ICR, Marseille, France
[‡]Université Paris-Est, Laboratoire Modélisation et Simulation Multi Échelle, MSME UMR 8208 CNRS, UPEM, 5 bd Descartes, 77454 Marne-la-Vallée, France
[§]Université de Strasbourg, CNRS, Laboratoire de Bioimagerie et Pathologies UMR 7021, 67034 Strasbourg, France
[¶]Departmento de Química Analítica, Química Física e Ingeniería Química, Universidad de Alcalá, Ctra. Madrid-Barcelona Km 33,600, 28871 Alcalá de Henares, Spain

1 Introduction

The molecular origin of the firefly bioluminescence lies in subtle microscopic interactions between the light emitter, called oxyluciferin, and its biological host, the luciferase enzyme catalyzing the oxyluciferin formation reaction (Figure 1)[1, 2]. When completed, this multi-step chemical reaction leads to oxyluciferin in an electronic excited state, which eventually releases energy by emitting visible light.

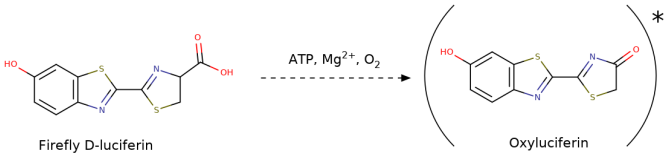


Figure 1: Bioluminescence reaction leading to the formation of the oxyluciferin light emitter in the excited state.

Excited oxyluciferin can exist in various forms whose relative contributions in the light-emission process have been debated over the years [3–6]. Most of these oxyluciferin forms can be interconverted by acid-base reactions or keto-enol tautomerization (Figure 2, left). Accordingly, oxyluciferin photochemical properties have been found to be pH-dependent [7]. Also, pH can modify the oxyluciferin environment, e.g. the protonation state of its luciferase host.

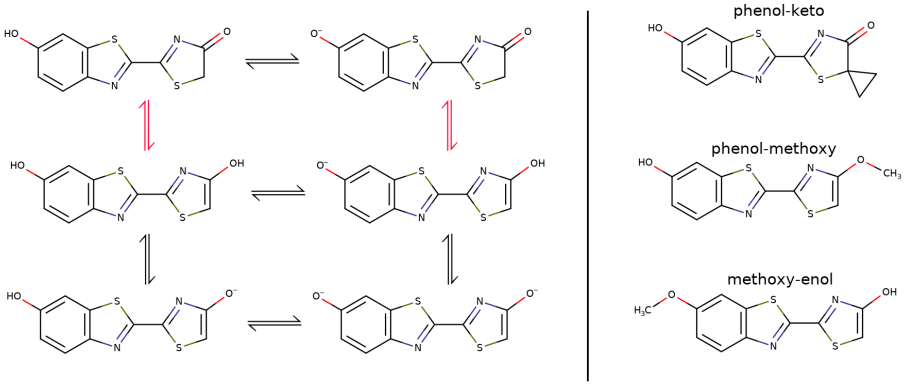


Figure 2: Natural (left) and synthetic (right) oxyluciferin structures. The natural ones are related by either keto-enol tautomerization (red arrows) or acid-base (black arrows) reactions. The synthetic analogues can only undergo acid-base reactions.

In luciferase, the time course of bioluminescence is found to be pH-dependent and explained by the same pH tendency as the rate of adenosine triphosphate

binding at the protein active site [8]. According to fragment molecular orbital calculations, the pH-induced adenosine monophosphate (AMP) change of protonation state is responsible for the polarization of a single oxyluciferin form (the deprotonated phenolate-keto one) in luciferase [9]. Other calculations have proposed a reactive route involving a proton transfer between the excited oxyluciferin and AMP[10]. The presence of specific interactions between the oxyluciferin anion (as well as a doubly-methylated analogue) and a water molecule is also known to tune the absorption spectrum by 50 nm [11]. This effect has also been found in computational studies by Kim and Rhee [12]. Very recently, the pH-sensitivity of luciferase photo-properties has been also explained by the presence of metal ions, weakening specific electrostatic interactions between oxyluciferin and titrating amino acid residues [13]. Finally, some of us have very recently studied the interplay between the protonation states of oxyluciferin+AMP complex and the size of its cavity in luciferase [14].

The extraction of the contribution due to each oxyluciferin form to the absorption or emission spectra is often tedious[15, 16]. For that reason, oxyluciferin analogues have been proposed and synthesized for inhibiting certain forms [6, 17–19]. In this work, we are interested in a particular set of analogues (Figure 2, right), studied both experimentally and theoretically [20–22]. In a nutshell, such analogues are meant to avoid any keto-enol tautomerization. Each analogue can be deprotonated at a single position. As a matter of fact, a good candidate for accepting or donating protons is AMP (chemical formula: $C_{10}H_{14}N_5O_7P$) which is present in the luciferase pocket. AMP first pKa being 3.8, it is assumed that it is deprotonated at pH values between 5 and 11. However, its second pKa value (6.2) is close from the oxyluciferin ones (7.1 and 9.0) as well as the analogue ones (phenol-keto: 7.8, phenol-methoxy: 8.7, methoxy-enol: 7.3). Accordingly, in the molecular complex formed between AMP and the oxyluciferin analogue, monodeprotonated and doubly-deprotonated AMP (hereafter denoted AMP for simplicity) could coexist with the neutral analogue and its deprotonated form. The existence of all these different possible protonation microstates is probably responsible for the pH-dependent properties of oxyluciferin and its analogues. The most basic demonstration is found in the variation of the absorption spectra of the bare analogues in water with the pH [20]. In the current work, AMP is introduced in the molecular models and its importance is evaluated using both experimental and theoretical absorption spectra of the analogues. Because of the absence of luciferase in the molecular model, the present work does not deal with the possible bioluminescence of the oxyluciferin analogues.

In the crudest approach, the simulation of an absorption spectrum is based on the vertical excitation energy computed at the chromophore ground state geometry. Among the numerous accurate methods rooted in quantum mechanics (QM), we may cite the three most popular ones: the Time-Dependent Kohn-Sham Density Functional Theory (hereafter simply denoted TDDFT [22, 23]), the second-order Perturbation Theory applied to a reference wavefunction commonly obtained from a multiconfigurational wavefunction [24] and the Equation-Of-Motion approach applied to a Coupled-Cluster wavefunction [25]. The latter two methods are usually restricted to small-to-medium size molecules because

of their high computational costs. Because of its favorable scaling, TDDFT can be applied to larger chromophores and/or take into account vibronic couplings [26], sometimes responsible of large maximum absorption wavelength (λ_{max}) shifts (with respect to the vertical transition) and the homogeneous broadening of the absorption peaks. On the other hand, when the chromophore is embedded in a complex environment (solvent, macromolecule, ...), the influence of its surroundings is responsible for the inhomogeneous broadening of the absorption bands. The latter effect is usually taken into account by considering a statistically meaningful distribution of the system configurations by means of molecular dynamics simulations (MD). Hundreds or thousands snapshots being extracted from the MD trajectory, their vertical excitation energies are calculated with TDDFT or a more qualitative QM level of theory, and then averaged [27]. In the case of pH-dependent absorption spectra, some of the authors have very recently proposed a new protocol, combining Constant-pH MD simulations and hybrid quantum mechanics/molecular mechanics (QM/MM) vertical energy calculations (CpHMD-then-QM/MM) in which the chromophore interacts with its environment through empirical atomistic potentials, i.e. molecular mechanics (MM) [28, 29].

In summary, we report in this article both experimental and computed absorption spectra of three oxyluciferin analogues in interaction with AMP and water at different pH values. Besides rationalizing their microscopic origin, we evaluate the effect of AMP on the oxyluciferin analogue spectroscopic signatures and validate the selected modeling protocol. We emphasize that the present work introduces the possibility to evaluate and analyze the pH-dependent absorption spectrum of a titrating chromophore in a titrating environment.

2 Results and Discussion

As mentioned above, two main effects can contribute to the computed absorption spectrum λ_{max} and its lineshape: (i) the homogeneous broadening which will be evaluated thanks to vibrationally-resolved TDDFT calculations using a single structure for each electronic state; (ii) the inhomogeneous broadening that will be obtained using the CpHMD-then-QM/MM protocol, using a semi-empirical QM level of theory.

2.1 Oxyluciferin analogues in water

In this section, we compare the experimental and computed spectra obtained for each analogue in water, at pH values between 5 and 11. These spectra are reported in Figures 3, 4 and 5. Because these experimental spectra are completely equivalent to the ones already available in the litterature [20, 22], we will just briefly comment them and focus on the computed ones.

Between 300 nm and 600 nm, the phenol-keto experimental spectra (Figure 3A) present a single large but structured absorption band at acidic pH with λ_{max} = 389 nm. At neutral and basic pH, two bands appear: an intense one

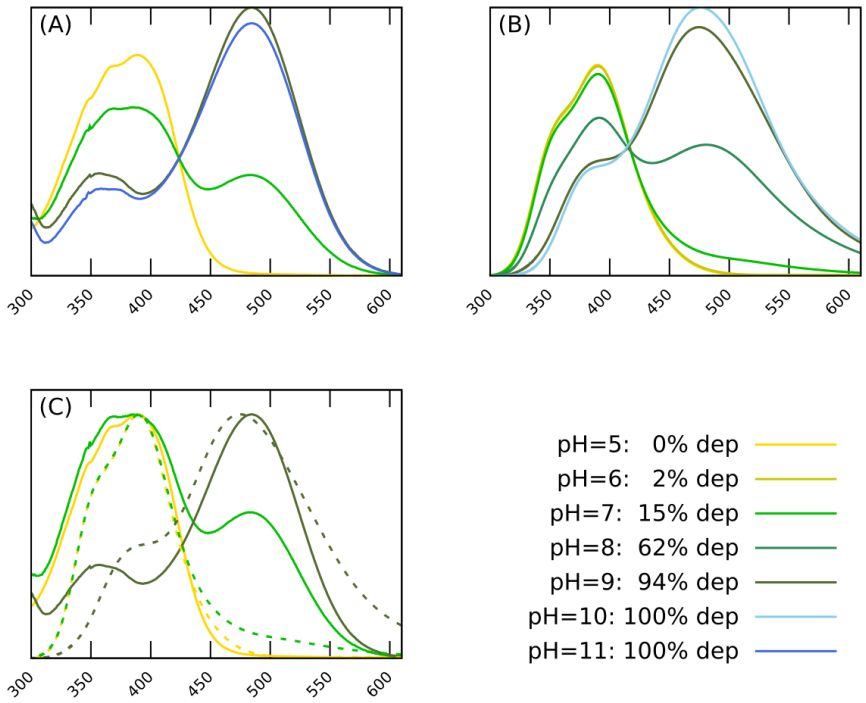


Figure 3: Absorption spectra of phenol-keto analogue (pKa=7.8) between pH=5 and pH=11. (A) Experimental unnormalized spectra. (B) Computed unnormalized spectra. (C) Normalized experimental (plain lines) and computed (dashed lines) spectra. Wavelengths are given in nm, intensities are given in arbitrary units. The deprotonated phenol-keto populations are calculated out of the CpHMD trajectories and are indicated as "% dep".

at λ_{max} = 480 nm and a second one, weaker, around λ_{max} = 360 nm. These features are correctly reproduced with the computational protocol (Figure 3B). An isobestic point is found at about 420 nm. After normalization of the two sets of spectra, their comparison in Figure 3C shows a large difference at pH=7: the absorption band corresponding to the deprotonated form is present in the experimental spectrum, while it barely appears in the computed one. Two possible explanations can be given: (i) the oscillator strengths using the semi-empirical (PM7) QM level of theory are underestimated in the case of the deprotonated form or (ii) the computed deprotonated population (15%) does not compare with the experimental one. Indeed the refined measurement of the solution pH is 7.4, explaining why the experimental spectrum is intermediate between the computed pH=7 and pH=8 ones. The experimental protonated–deprotonated phenol-keto energy gap (90 nm) is correctly reproduced using the CphMD-then-

QM/MM protocol (85 nm).

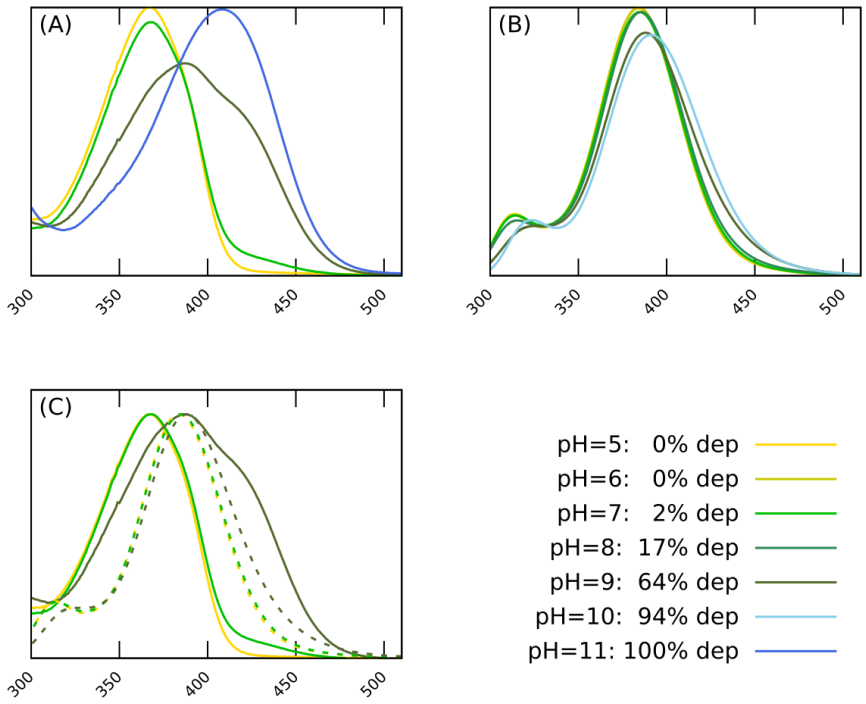


Figure 4: Absorption spectra of phenol-methoxy analogue (pKa=8.7) between pH=5 and pH=11. (A) Experimental unnormalized spectra. (B) Computed unnormalized spectra. (C) Normalized experimental (plain lines) and computed (dashed lines) spectra. Wavelengths are given in nm, intensities are given in arbitrary units. The deprotonated phenol-methoxy populations are calculated out of the CpHMD trajectories and are indicated as "% dep".

Regarding the phenol-methoxy analogue (Figure 4), the experimental spectra can be analyzed as the superposition of two bands, one for the protonated form (λ_{max} = 367 nm) and one for the deprotonated form (λ_{max} = 407 nm). Owing to this analogue large pKa value (8.7), the spectrum at pH=9 contains an almost equal contribution of both forms, resulting in a broad band between 320 and 500 nm. With respect to these experimental spectra, the computed protonated form λ_{max} is red-shifted by 17 nm, whereas the deprotonated form λ_{max} is blue-shifted by 16 nm. Consequently, the protonated–deprotonated phenol-methoxy energy gap is largely underestimated (7 nm instead of 40 nm) using the CpHMD-then-QM/MM approach. The most probable explanation is rooted in the QM semi-empirical level of theory which may be not accurate enough (see the discussion below).

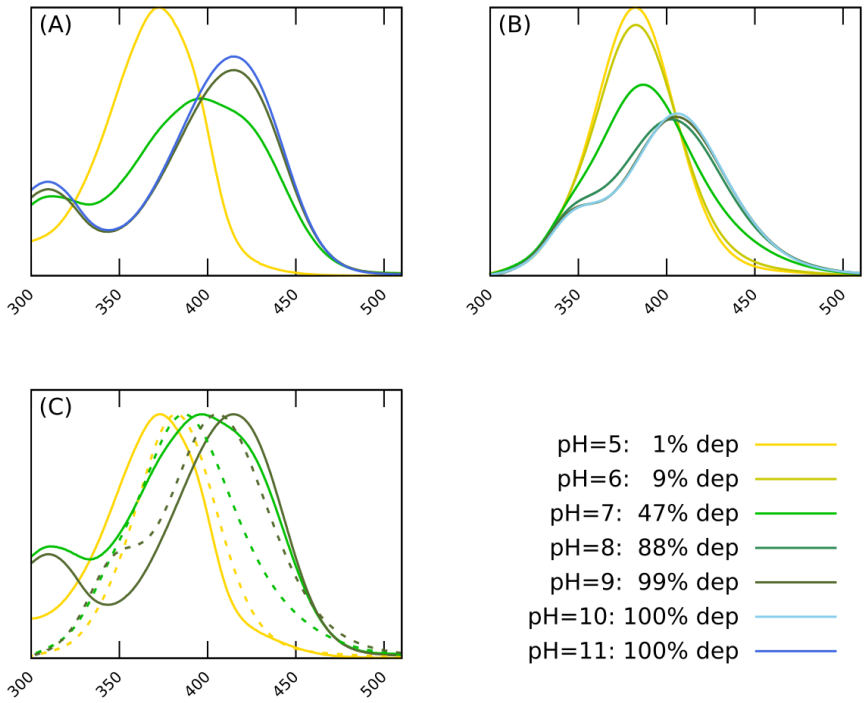


Figure 5: Absorption spectra of methoxy-enol analogue (pKa=7.3) between pH=5 and pH=11. (A) Experimental unnormalized spectra. (B) Computed unnormalized spectra. (C) Normalized experimental (plain lines) and computed (dashed lines) spectra. Wavelengths are given in nm, intensities are given in arbitrary units. The deprotonated methoxy-enol populations are calculated out of the CpHMD trajectories and are indicated as "% dep".

In the case of the methoxy-enol analogue (Figure 5), the protonated form can be identified at pH=5 with an experimental λ_{max} = 371 nm while the deprotonated form spectroscopic signature is found at λ_{max} = 414 nm. The computed spectra show similar features: 382 nm for the depronated form (11 nm red-shifted with respect to experiment) and 406 nm for the depronated form (8 nm blue-shifted with respect to experiment). Accordingly the computed protonated–deprotonated energy gap (24 nm) is only half the experimental one (43 nm), hence showing the same trend already found with the phenol-methoxy analogue. Noticeably, in Figure 5C, the computed spectra at pH=9 and pH=10 feature a shoulder around 340 nm, whereas a more intense band starts in the corresponding experimental spectra.

The PM7 semi-empirical level of theory is well adapted to the calculation of tens of thousands of vertical excitation energies. It has already been vali-

dated on a different chromophore, the retinal in a rhodopsin protein[29]. Owing to the experimental-computational spectrum differences highlighted above, we have investigated the possibility to run time-dependent density-functional theory (TDDFT) calculations on a reduced set (1000 instead of 10000) of snapshots for each oxyluciferin analogue, at three pH values: acidic (5), close to the analogue pKa (7 or 8), and basic (10). Because of the TDDFT computational cost, the small 6-31G(d) basis set (together with the B3LYP exchange-correlation functional) has been selected. Aware of the possible charge transfer character of some excited states, we also evaluated a range-separated functional (Supplementary Information). However, its unbalanced treatment of protonated and deprotonated species leads us to prefer the B3LYP one. The 3 sets of spectra are compared in Figure 6. The resulting spectra are less smooth than the PM7 ones,

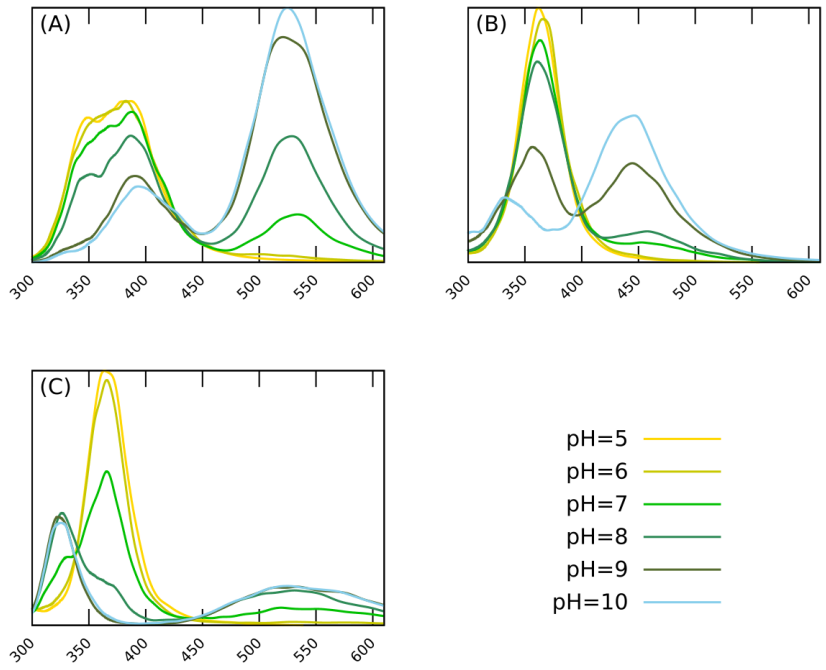


Figure 6: TDDFT absorption spectra of the three analogues (A: phenol-keto, B: phenol-methoxy, C: methoxy-enol). Wavelengths are given in nm, intensities are given in arbitrary units.

because of the limited sampling. The TDDFT spectrum of the phenol-keto analogue shows a too large energy gap between the protonated and deprotonated forms, due to the red-shift of the band that appears now around 550 nm. Nevertheless, it confirms that the large band between 300 and 450 nm is due to the protonated $S_0 \rightarrow S_1$ and the deprotonated $S_0 \rightarrow S_2$ transitions, depending

on the pH value [22]. At variance with the PM7 spectra, the phenol-methoxy TDDFT spectra show a significant 100 nm gap, however too large with respect to the experimental one. Similar to the phenol-keto spectrum, the deprotonated contribution shows a second excitation energy close to the protonated first excitation [22]. In the case of the methoxy-enol, the TDDFT spectra are clearly not accurate for the deprotonated form. Indeed, the first transition appears as a very large band between 450 and 700 nm while the corresponding experimental band can be found between 350 and 500 nm.

Despite these TDDFT inaccuracies, we have also investigated the effect of possible couplings between the analogue molecular vibrations and their excitation energies using TDDFT with a larger basis set. Ideally, we would need to apply a vibronic+ensemble approach as the one proposed recently by Zuehlsdorff and Isborn [30]. However, its computational cost would be prohibitive when multiple protonation states have to be considered. Instead we have used the crude approximation that intramolecular vibronic effects could be evaluated from a single gas phase calculation. Because the selected vibronic model [31] is based on the extraction of both the electronic ground state and excited state normal modes, this approach has been applied using a single geometry for each protonation state of each analogue, except for the protonated phenol-keto analogue for which the Franck-Condon approximation cannot be applied (the excited state geometry is too much distorted with respect to the ground state one, see Figure 7).

For all analogues, the maximum of the vibrationally-resolved spectra can be associated with the corresponding 0–0' excitation energy. Interestingly, the absorption of the deprotonated form of the methoxy-enol analogue is strongly red-shifted with respect to the other analogues (Table 1). Accordingly, the deprotonation of the methoxy-enol analogue is reflected in a 270 nm red-shift, much larger than the one in water, either experimentally observed or calculated. This shift is reduced to 135 nm in the case of the phenol-methoxy analogue, still much larger than the one in water. Assuming that the (protonated to deproto-

analogue	protonation	vertical	0–0' (max)
phenol-keto	prot	375	
	deprot	494	541 (539)
phenol-methoxy	prot	364	400 (398)
	deprot	497	535 (535)
methoxy-enol	prot	371	406 (405)
	deprot	606	674 (674)

Table 1: Vertical excitation energy, 0–0' excitation energy and maximum absorption energy (in nm) for each oxyluciferin analogue.

nated) differential effect of the vibronic coupling in vacuum can be transferred (added) to the CpHMD-then-QM/MM spectra (Figures 4 and 5), the total protonated to deprotonated shift of calculated λ_{max} increases to 9 nm (instead of 7 nm, phenol-methoxy) and 57 nm (instead of 24 nm, methoxy-enol). The latter

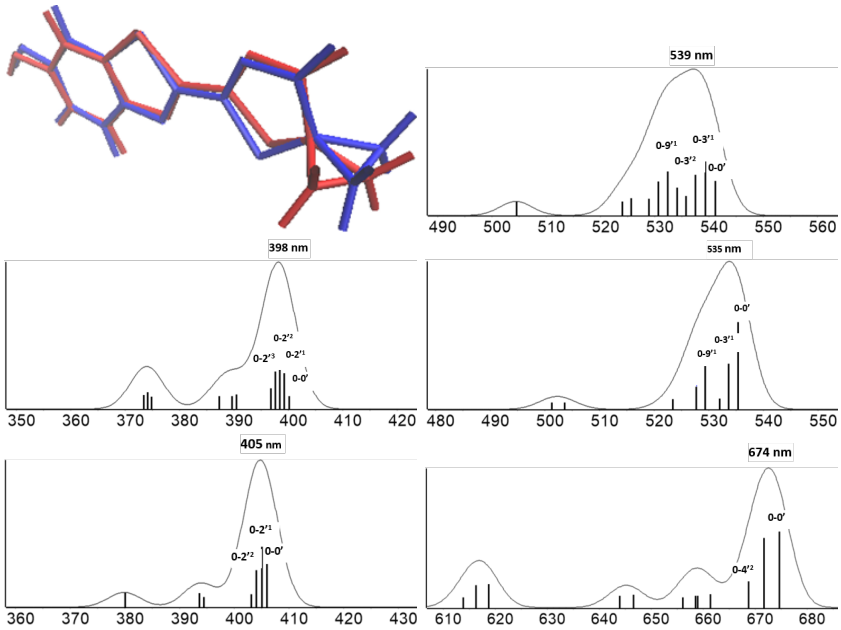


Figure 7: Vibrationally resolved TDDFT normalized absorption spectra of the three analogues (top: phenol-keto, middle: phenol-methoxy, bottom: methoxy-enol; left: protonated, right: deprotonated) in vacuum. Wavelengths are given in nm. The vibronic transitions are depicted with sticks. In the case of the protonated phenol-keto analogue, the superimposition of the electronic ground and first excited states minimum energy structures is provided and explains why the spectrum cannot be calculated.

value is in better agreement with the experimental 43 nm reported above. Qualitatively speaking, these results suggest that both pH and vibronic effects would deserve attention in some particular situations, like the methoxy-enol analogue. The evaluation of the vibronic contribution in water would require much more involved models, able to evaluate the intramolecular and intermolecular vibronic couplings in each and every populated protonation microstate.

Because of the sampling limitation and of the need to use a small-yet-computationally inexpensive basis set (at variance with the one recommended by some of us [22]), we keep PM7 as the QM level in the following, unless explicitly mentioned.

Among all the possible structural factors that could be connected to the analogue absorption spectrum pH-dependency, we have considered an intramolecular one (the torsion angle around the S–C–C–S central bond) and an intermolecular one (the radial distribution function (RDF) of the analogue alcohol oxygen–water oxygen distance). Even a slight rotation, e.g. due to the local solvent inhomogeneity, around the S–C–C–S bond would dramatically modify

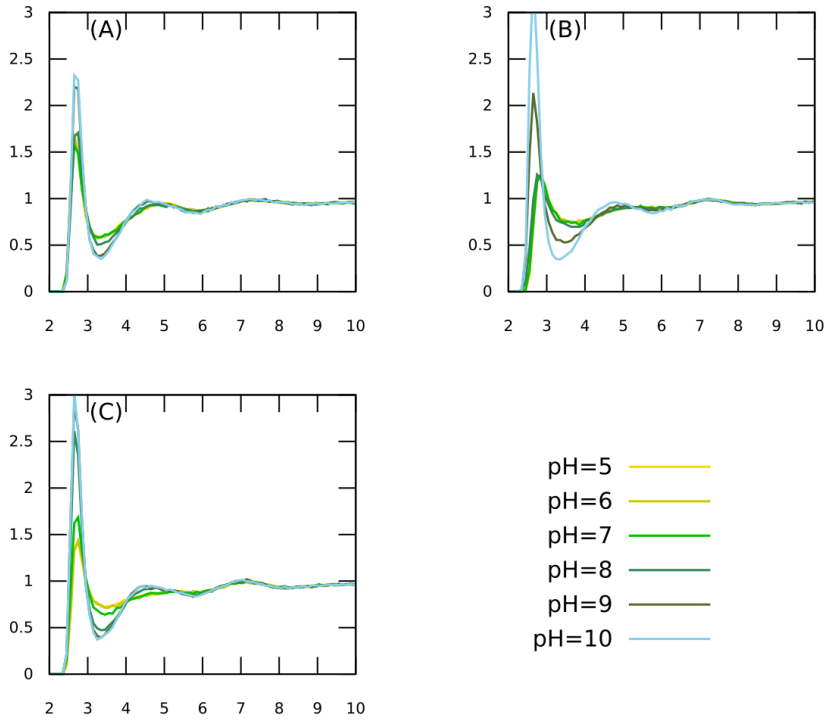


Figure 8: Normalized water oxygen – analogue alcohol oxygen RDF for the analogues phenol-keto (A), phenol-methoxy (B) and methoxy-enol (C). Distances are given in Å.

λ_{max} because of the conjugation between the two oxyluciferin analogue moieties. However, it turns out that the pH does not modify this parameter whose values ($180 \pm 10°$) indicate a planar geometry (data not shown) whatever the analog. However, there is indication that the pH can influence the distribution of the water molecules around the analogue titrating site (Figure 8). All the three analogues exhibit similar pH-dependent solvent organizations, with the same first solvation sphere at about 2.5Å from the titrating oxygen atom. While the phenol-keto analogue can accomodate between 1.5 water molecules at pH=5 and 2.3 water molecules at pH=10, the two other analogues can accept up to 3 water molecules at pH=10. Not only the role of the water molecules hydrogen-bonded to the analogue phenol/phenolate or enol/enolate moieties changes from mixed acceptor/donor at acidic pH to only donor at basic pH, but the presence of the large negative electric charge of the deprotonated forms is most probably attracting more water molecules. Such a complex solvent effect involves specific solvent–solute interactions which can't be accounted for in isotropic solvent models like the one we have used for computing vibrationally-resolved spectra.

2.2 Oxyluciferin analogue–AMP complexes in water

The presence of AMP is supposed to influence the oxyluciferin analogue absorption spectra, at least because of the presence of a -1 or -2 electric charge in the considered pH range (5 to 11). Note that AMP absorbs UV light below 300 nm (as we verified by computing its vibrationally-resolved spectra at the TDDFT level of theory, see Supplementary Information), significantly different from the analogue absorption. Nevertheless, we have considered two different computational cases: (i) the analogue and AMP in the QM subsystem; (ii) the analogue in the QM subsystem and AMP in the MM one (together with the water molecules).

Before to present the pH-dependent analogue-AMP complex spectra, we first analyze the CpHMD trajectories, in terms of deprotonation fractions for each of the 4 possible microstates (2 protonation states for AMP times 2 for the analogue). Using the fitting procedure developed by some of us (see the Experimental Section), we were able to derive the pKa values of each analogue, together with the AMP one (Table 2). CpHMD populations and the fitted curves are reported in Supplementary Information. Compared to the AMP pKa

	phenol-keto	phenol-methoxy	methoxy-enol
AMP	6.7	7.7	6.7
analogue	7.2	8.7	7.3

Table 2: Microscopic pKa values for AMP and each analogue in their complexes solvated in water.

(6.2, corresponding to its second deprotonation), the presence of the analogue results in a larger pKa, meaning that AMP becomes a weaker proton donnor. On the other hand, the analogue pKa values are barely affected by the presence of AMP, the largest deviation concerning the phenol-keto analogue (-0.6 pKa unit). From the titration point of view, the phenol-keto and the methoxy-enol analogues behave similarly when they interact with AMP.

The experimental and CpHMD-then-QMMM computed absorption spectra of the phenol-keto analogue in presence of AMP, at different pH values, are reported in Figure 9. Comparing Figures 3A and 9A, the experimental spectra do not significantly change with the complexation of the analogue with AMP. The differences introduced by AMP in the computed spectra remain small. First, a limited 8 nm blue-shift of the peak corresponding to the analogue deprotonated form can be noticed. Then, the protonated form band structure is lost when AMP is included into the QM subsystem. The QM/MM isobestic point is slightly blue-shifted with respect to the experimental one. The agreement between the experimental and computed pH=7 spectra has improved: both definitively show contributions from the protonated and depronated forms of the analogue, with relative intensities in good agreement.

In the case of the phenol-methoxy analogue (Figure 10), the influence of AMP is also negligible, as evidenced in the experimental spectra at all pH values. The computed spectra show a limited effect of AMP, through a 15 nm blue-

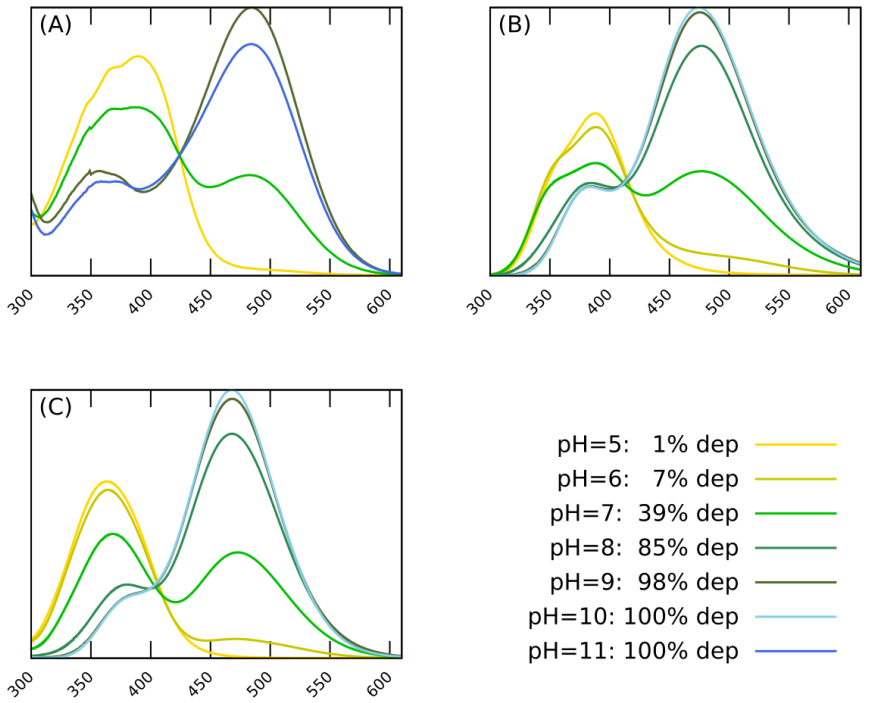


Figure 9: Absorption spectra of AMP + phenol-keto complex between pH=5 and pH=11. (A) Experimental unnormalized spectra. (B) Computed unnormalized spectra, AMP is part of the MM subsystem. (C) Computed unnormalized spectra, AMP is part of the QM subsystem. Wavelengths are given in nm. The deprotonated phenol-keto populations are calculated out of the CpHMD trajectories and are indicated as "% dep".

shift. Again, a 40 nm protonated to deprotonated analogue shift is measured, however the corresponding computed value is underestimated to less than 10 nm, whatever the chosen description (QM or MM) for AMP.

Even if the presence of AMP induces a slight shift of the methoxy-enol deprotonation (Figure 11) towards higher pH, the general features of its experimental absorption spectrum does not depend on AMP. The computed spectra are also independent of AMP. Only the level of theory for AMP induces a modification of the deprotonated second absorption ($\lambda_{max} = 340$nm) band shape, which appears as a shoulder when AMP is QM whereas it is a distinct peak when AMP belongs to the MM subsystem.

While AMP is always charged negatively (-1 or -2) in the 5 to 11 pH range, the analogues can be either neutral or negatively charged. Accordingly, it is expected that the distance between the two molecules increases with the pH.

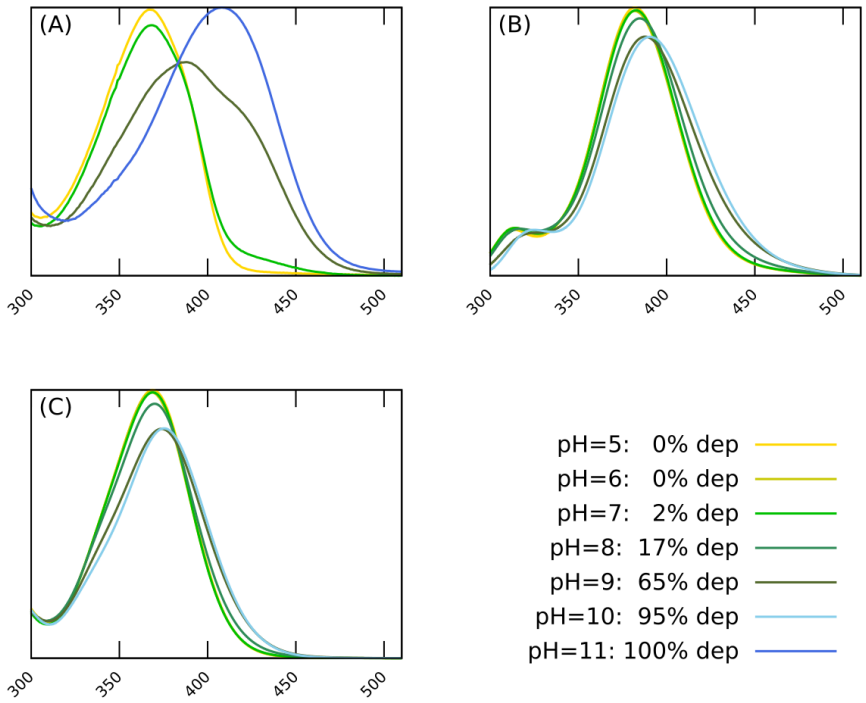


Figure 10: Absorption spectra of AMP + phenol-methoxy complex between pH=5 and pH=11. (A) Experimental unnormalized spectra. (B) Computed unnormalized spectra, AMP is part of the MM subsystem. (C) Computed unnormalized spectra, AMP is part of the QM subsystem. Wavelengths are given in nm. The deprotonated phenol-methoxy populations are calculated out of the CpHMD trajectories and are indicated as "% dep".

Having plotted the distance between their respective centers of mass (Figures S6-S8), the distributions of distances can be summarized as unimodal or bimodal distributions centered around 5 Å and/or around 30Å (in a region comprised between 25 to 40Å). No distinct trend can be discerned, whatever the pH or analogue. In other words, the two solutes can be either interacting or separated enough for their interaction to be completely screened by the solvent.

3 Conclusions

In this article, we have presented experimental and computational investigations of the pH-dependent absorption spectra for three aqueous oxyluciferin analogues in the presence of AMP. It has been shown that AMP does not mod-

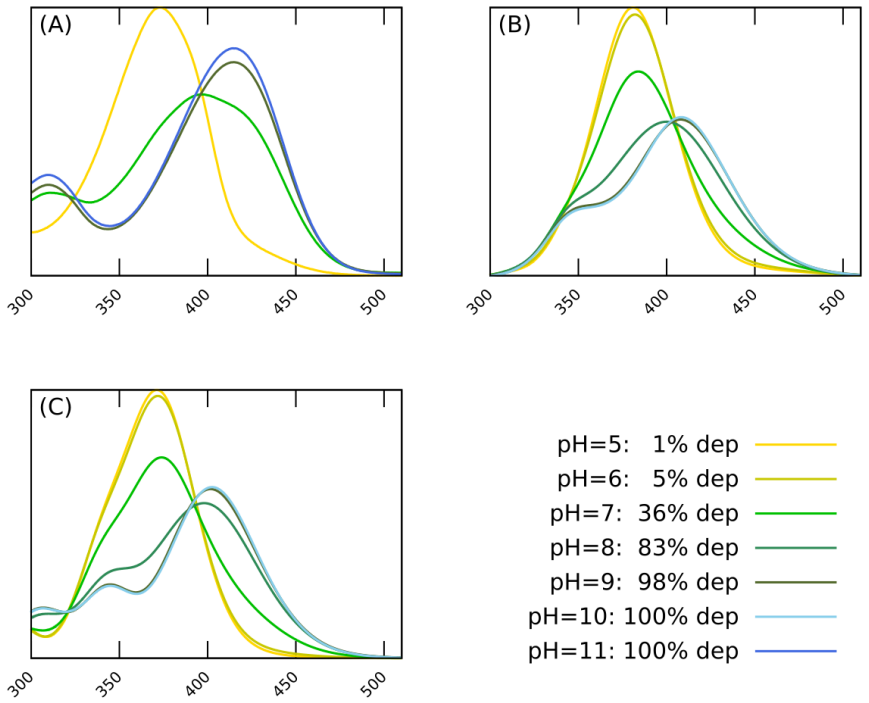


Figure 11: Absorption spectra of AMP + methoxy-enol complex between pH=5 and pH=11. (A) Experimental unnormalized spectra. (B) Computed unnormalized spectra, AMP is part of the QM subsystem. (C) Computed unnormalized spectra, AMP is part of the MM subsystem. Wavelengths are given in nm. The deprotonated methoxy-enol populations are calculated out of the CpHMD trajectories and are indicated as "% dep".

ify significantly the analogue light absorption, at all the considered pH values between 5 and 11. Despite its computational cost, the CpHMD-then-QM/MM approach is a particularly effective approach, able to capture the contributions of the different protonation microstates of the chromophore, interacting with another titratable molecule. Most notably, it allows to calculate the pH-dependent absorption spectrum as a weighted sum of contributions from individual protonation microstates. This converged statistical approach is utterly important for a correct evaluation of the pH effects. This last result calls for a generalization of such a pH-dependent approach for studying pH effects in a more complex environment like the oxyluciferin host, the luciferase protein, opening the way to the investigation of how pH can affect the firefly bioluminescence.

4 Experimental Section

Experimentally recorded absorption spectra Stock solutions of all compounds were prepared in spectroscopic grade DMSO (Sigma-Aldrich) and stored in several aliquots at 20°C to minimize the freezethaw cycle. They were further diluted 1000-fold to a final concentration of about a few μM in aqueous buffer with different pH. The buffers were prepared by diluting a 1 M stock solution in deionized, Millipore-purified water (18.2 MΩ). Buffered stock solutions were prepared as follows: 75 mM NaCl/20 mM KH_2PO_4/0.2 mM $MgCl_2$ for pH $\leq$ 7.0 and 75 mM NaCl/25 mM Tris(hydroxymethyl)- aminomethane (TRIS)/0.2 mM $MgCl_2$ for pH $>$ 7.0. The buffer was separated into several fractions, and their pH (error $\pm$0.02) was adjusted by 250 mM HCl or by 250 mM NaOH separately at 20°C (ΔpH 0.25). Final pH values are 5.0, 7.4, 9.1 and 11.1. Separate buffers at different pH were used for all spectral measurements. Absorption spectra were recorded with a Cary-4000 spectrometer (Agilent Technologies).

Determination of microscopic pKa values. When several titrating sites are interacting, the macroscopic parameter pKa becomes pH-dependent and can be replaced with the so-called $pK_{1/2}$ parameter which is formally equal to the pH value at which the protonated and deprotonated populations of a particular titrating site become equal, whatever the other populations. In the case of the analogue-AMP complex (2 titrating sites), using a 3-parameter fitting procedure of the 4 microstate population curves, it is possible determine the two individual $pK_{1/2}$ values, that we simply call pKa for the sake of simplicity (see the Supplementary Information).

CpHMD and molecular parameters. We have used the standard CpHMD method implemented in the Amber18 software[32]. During the simulation, solvent molecules are considered explicitly in each sampled protonation microstate. However, any protonation/deprotonation event, ie a change of protonation microstate, is performed in implicit solvent. Accordingly, the implicit water solvent model must be seen as an infinite reservoir of protons that translates statistically to a semi-grand canonical ensemble in which the hydronium chemical potential is constant. When a given residue undergoes a protonation state change, its atomic charges are modified in order to represent the new state. Titratable hydrogen atoms have zero charge in the deprotonated form. Since this change of net charge occurs in implicit solvent, all the charge-dependent finite size effects that would occur in explicit solvent are prevented: each MD partial trajectory is performed using a constant total charge (if not 0, the usual correction term is applied to the particle-mesh Ewald sums). Optimized structures of oxyluciferin analogues as well as AMP at the B3LYP /6-311G(2d,p) level of theory were obtained from a previous study [22]. In order to describe non-standard residues in CpHMD, partial atomic charges of all the structures were derived following the restrained electrostatic potential (RESP) approach. Reference energies were calculated using the finddgref tool implemented in the AMBER18

package[32]. The electrostatic contribution of the transition free energy for the deprotonation is calculated by taking into account the difference between the potential calculated with the charges for the proposed state and the potential calculated with the charges for the current protonation state, assuming that the nonelectrostatic transition free energy interactions are constant independently of the environment [33].

Oxyluciferin analogs and AMP parameters for each protonation state are reported as Supplementary Information.

The starting structures, analogues and complexes, were then solvated with TIP3P water molecules, extending for at least 20 Å from the structures to the box sides. All molecular dynamics simulations (MD) were performed using the AMBER18[32] software package. For each system, we followed the same 3 steps towards the equilibrium of the structures: i) an energy minimization of the whole system ii) a MD simulation gradually raising up the temperature of the system from 10 to 300 K iii) a 4 ns equilibration of the systems using a NPT ensemble. After the equilibration of the models, we applied the CpHMD method in explicit solvent using the discrete protonation state model. This method involves a MD simulation in explicit solvent followed by a Metropolis Monte Carlo-based evaluation of the Boltzmann distribution of protonated states for each residue in a random order. Willing to improve the sampling capabilities of this method, we used the replica exchange technique (pH-REMD)[34]. We took into account six 10 ns long pH-replicas between pH=5 and pH=10. Deprotonation fractions were checked to be converged much before reaching the 10 ns simulation length. All simulations were performed enforcing periodic boundary conditions.

QM/MM method. 10000 equally distributed snapshots were extracted from each trajectory and assigned to the corresponding protonation microstates. Three different QM/MM partitions were considered. In the absence of AMP, only the analogue belongs to the QM subsystem while the water molecules are always MM. When AMP is complexed with the analogues, we placed it either in the QM subsystem or in the MM subsystem. The electrostatic interaction between the QM and MM subsystems is taken into account by plugin the MM electrostatic potential into the QM one-electron hamiltonian. Two different QM hamiltonians have been used: the semi-empirical PM7 implemented in the MOPAC 2016 program [35] in which the MM potential enters in the diagonal Fock operator elements, or the DFT Khon-Sham one thanks to the ESPF method[36], implemented in a local version of Gaussian16 [37].

10000 vertical excitation energies and oscillator strengths have been computed using the MECI/PM7 approach in MOPAC [35]. The same properties have been determined at the TDDFT/TDA B3LYP/6-31G(d) level of theory using a restricted set of 1000 structures. The absorption spectra were generated from the first five excited states with a normalized Gaussian function using Newton-X [38].

Computed vibrationally-resolved absorption spectra. Vibrationally resolved absorption spectra have been computed on the basis of the TD-DFT vibrational signatures by using the Franck-Condon-Herzberg-Teller (FCHT) approximation and the Adiabatic Hessian method as implemented in Gaussian 16[37]. A comparison of the FranckCondon (FC) and FCHT vibrationally resolved absorption and emission spectra of oxyluciferin has been already discussed in a previous study[39], showing that the FCHT spectra resemble the FC ones. For this reason, only the FCHT spectra are simulated in this study. The band broadening was simulated by Gaussian functions with half-widths at half-maximum (HWHM) of 135 cm^{-1}. The vibrationally-resolved absorption spectra have been simulated in vacuum. All the structures have been optimized in the ground and excited state at the DFT and TD-DFT levels of theory respectively, using the B3LYP functional and the 6-31+G(d,p) basis set.

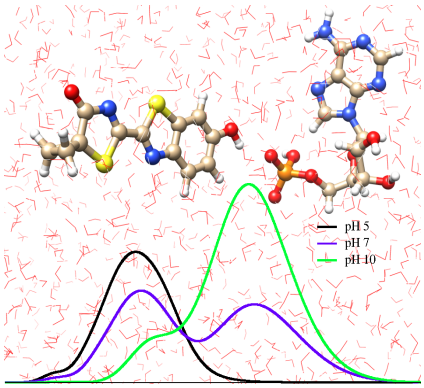
5 Acknowledgment

Agence Nationale de la Recherche is acknowledged for grant ANR-16-CE29-0013-01. Cristina Garcia-Iriepa acknowledges Fundación Ramón Areces for a postdoctoral fellowship. This work was granted access to the HPC resources of Aix-Marseille Université financed by the project Equip@Meso (ANR-10-EQPX-29-01) of the program "Investissements d'Avenir" supervised by the Agence Nationale de la Recherche.

6 Keywords

Oxyluciferin analogue, absorption spectrum, multiscale simulations, pH

7 TOC



Experimental (steady-state absorption spectroscopy) and computational (constant-pH molecular dynamics and QM/MM) insights on the pH-dependent absorption spectra of aqeuous oxyluciferin analogues in presence of adenosine monophosphate (AMP) are reported. Most notably, the results demonstrate the adequacy of the computational approach to reproduce pH-dependent absorption spectrum as well as the negligible effect of AMP in the absorption spectra of the oxyluciferin analogues.

References

[1] B. Branchini, S. Daunert, L. Blum, L. J. Kricka, G. Zomer, J. W. Hastings, F. Berthold, A. Lundin, A. M. G. Campana, R. Niessner, T. K. Christopolous, C. Lowik, *Chemiluminescence and Bioluminescence*, (Ed.: A. Roda), The Royal Society of Chemistry, **2011**, P001–590.

[2] M. Vacher, I. F. Galván, B.-W. Ding, S. Schramm, R. Berraud-Pache, P. Naumov, N. Ferré, Y.-J. Liu, I. Navizet, D. Roca-Sanjuán, W. J. Baader, R. Lindh, *Chem. Rev.* **2018**, *118*, 6927–6974.

[3] G. Orlova, J. D. Goddard, L. Y. Brovko, *J. Am. Chem. Soc.* **2003**, *125*, 6962–6971.

[4] S.-F. Chen, Y.-J. Liu, I. Navizet, N. Ferré, W.-H. Fang, R. Lindh, *J. Chem. Theory Comput.* **2011**, *7*, 798–803.

[5] L. P. da Silva, J. C. E. da Silva, *J. Chem. Theory Comput.* **2011**, *7*, 809–817.

[6] Y.-Y. Cheng, Y.-J. Liu, *J. Chem. Theory Comput.* **2015**, *11*, 5360–5370.

[7] Y. Ando, K. Niwa, N. Yamada, T. Enomoto, T. Irie, H. Kubota, Y. Ohmiya, H. Akiyama, *Nat. Photonics* **2007**, *2*, 44–47.

[8] Y. Yanagisawa, T. Kageyama, N. Wada, M. Tanaka, S.-y. Ohno, *Photochem Photobiol* **2013**, *89*, 1490–1496.

[9] B. F. Milne, M. A. L. Marques, F. Nogueira, *Phys. Chem. Chem. Phys.* **2010**, *12*, 14285.

[10] L. P. da Silva, J. C. G. E. da Silva, *The Journal of Physical Chemistry B* **2015**, *119*, 2140–2148.

[11] K. Støchkel, C. N. Hansen, J. Houmøller, L. M. Nielsen, K. Anggara, M. Linares, P. Norman, F. Nogueira, O. V. Maltsev, L. Hintermann, S. B. Nielsen, P. Naumov, B. F. Milne, *J. Am. Chem. Soc.* **2013**, *135*, 6485–6493.

[12] H. W. Kim, Y. M. Rhee, *The Journal of Physical Chemistry B* **2013**, *117*, 7260–7269.

[13] V. R. Viviani, G. V. M. Gabriel, V. R. Bevilaqua, A. F. Simões, T. Hirano, P. S. Lopes-de-Oliveira, *Sci. Rep.* **2018**, *8*, DOI `10.1038/s41598-018-33252-x`.

[14] C. Garcia-Iriepa, I. Navizet, *Molecules* **2019**, *24*, 1565.

[15] M. Rebarz, B.-M. Kukovec, O. V. Maltsev, C. Ruckebusch, L. Hintermann, P. Naumov, M. Sliwa, *Chem. Sci.* **2013**, *4*, 3803–3809.

[16] J. J. Snellenburg, S. P. Laptenok, R. J. DeSa, P. Naumov, K. M. Solntsev, *J. Am. Chem. Soc.* **2016**, *138*, 16252–16258.

[17] L. P. da Silva, A. J. M. Santos, J. C. G. E. da Silva, *J. Phys. Chem. A* **2013**, *117*, 94–100.

[18] M. Kakiuchi, S. Ito, M. Yamaji, V. R. Viviani, S. Maki, T. Hirano, *Photochem. Photobiol.* **2016**, *93*, 486–494.

[19] C.-G. Min, Q.-B. Liu, Y. Leng, S.-J. Huang, C.-X. Liu, X.-K. Yang, A.-M. Ren, L. P. da Silva, *Comput. Theor. Chem.* **2018**, *1133*, 18–24.

[20] A. Ghose, M. Rebarz, O. V. Maltsev, L. Hintermann, C. Ruckebusch, E. Fron, J. Hofkens, Y. Mély, P. Naumov, M. Sliwa, P. Didier, *J. Phys. Chem. B* **2015**, *119*, 2638–2649.

[21] A. Ghose, O. V. Maltsev, N. Humbert, L. Hintermann, Y. Arntz, P. Naumov, Y. Mély, P. Didier, *J. Phys. Chem. B* **2017**, *121*, 1566–1575.

[22] C. García-Iriepa, P. Gosset, R. Berraud-Pache, M. Zemmouche, G. Taupier, K. D. Dorkenoo, P. Didier, J. Léonard, N. Ferré, I. Navizet, *J. Chem. Theory Comput.* **2018**, *14*, 2117–2126.

[23] M. E. Casida, M. Huix-Rotllant in *Annual review of physical chemistry, vol 63*, (Eds.: M. A. Johnson, T. J. Martinez), Annual Review of Physical Chemistry BAI05, Annual Reviews, **2012**, pp. 287–323.

[24] S. Gozem, F. Melaccio, R. Lindh, A. I. Krylov, A. A. Granovsky, C. Angeli, M. Olivucci, *J. Chem. Theory Comput.* **2013**, *9*, 4495–4506.

[25] M. Caricato, B. Mennucci, G. Scalmani, G. W. Trucks, M. J. Frisch, *J. Chem. Phys.* **2010**, *132*, 084102.

[26] F. Santoro, D. Jacquemin, *WIREs Comput Mol Sci* **2016**, *6*, 460–486.

[27] U. N. Morzan, D. J. A. de Armiño, N. O. Foglia, F. Ramírez, M. C. G. Lebrero, D. A. Scherlis, D. A. Estrin, *Chem. Rev.* **2018**, *118*, 4071–4113.

[28] E. Pieri, V. Ledentu, M. Huix-Rotllant, N. Ferré, *Phys. Chem. Chem. Phys.* **2018**, *20*, 23252–23261.

[29] E. Pieri, V. Ledentu, M. Sahlin, F. Dehez, M. Olivucci, N. Ferré, *J. Chem. Theory Comput.* **2019**, *15*, 4535–4546.

[30] T. J. Zuehlsdorff, C. M. Isborn, *J. Chem. Phys.* **2018**, *148*, 024110.

[31] F. Santoro, A. Lami, R. Improta, J. Bloino, V. Barone, *J. Chem. Phys.* **2008**, *128*, 224311.

[32] D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, I. T.E. Cheatham, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson, H. Gohlke, A. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, L. Xiao, D. York, P. Kollman.

[33] J. Mongan, D. A. Case, J. A. McCammon, *J. Comput. Chem.* **2004**, *25*, 2038–2048.

[34] D. M. Y. Jason M. Swails, A. E. Roitberg, *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352.

[35] J. J. P. Stewart, MOPAC 2016, http://openmopac.net/, **2016**.

[36] N. Ferré, J. G.Ángyán, *Chem. Phys. Lett.* **2002**, *356*, 331–339.

[37] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 16 Revision B.01, Gaussian Inc. Wallingford CT, **2016**.

[38] M. Barbatti, M. Ruckenbauer, F. Plasser, J. Pittner, G. Granucci, M. Persico, H. Lischka, *Rev.: Comput Mol Sci.* **2014**, *4*, 26–33.

[39] C. García-Iriepa, M. Zemmouche, M. Ponce-Vargas, I. Navizet, *Phys. Chem. Chem. Phys.* **2019**, *21*, 4613–4623.