



HAL
open science

pH-dependent absorption spectrum of oxyluciferin analogues in the active site of firefly luciferase

Nuno Manuel de Almeida Barbosa, Pauline Gosset, Eléonore Réal, Vincent Ledentu, Pascal Didier, Nicolas Ferré

► **To cite this version:**

Nuno Manuel de Almeida Barbosa, Pauline Gosset, Eléonore Réal, Vincent Ledentu, Pascal Didier, et al.. pH-dependent absorption spectrum of oxyluciferin analogues in the active site of firefly luciferase. Physical Chemistry Chemical Physics, 2020, 10.1039/D0CP02514C . hal-02954392

HAL Id: hal-02954392

<https://amu.hal.science/hal-02954392>

Submitted on 1 Oct 2020

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 active site of *re_γ* luciferase[†]

Nuno Manuel de Almeida Barbosa,^a Pauline Gosset,^b Eléonore Réal,^b Vincent Ledentu,^a Pascal Didier^b and Nicolas Ferré^{*a}

In the quest for the identification of the light emitter(s) responsible for the *re_γ* bioluminescence, the study of oxyluciferin analogues with controlled chemical and electronic structures is of particular importance. In this article, we report the results of our experimental and computational investigation of the pH-dependent absorption spectra characterizing three analogues bound into the luciferase cavity, together with adenosine-monophosphate (AMP). While the analogue microscopic pK_a values don't differ much from their reference values, it turns out that AMP protonation state is analogue-dependent and never doubly-deprotonated. Careful analysis of the interactions evidences the main role of E344 glutamic acid, as well as the flexibility of the cavity which can accommodate any oxyluciferin analogue. The consideration of the absorption spectra suggests that the oxyluciferin enolate form has to be excluded from the list of the bioluminescence reaction products.

1 Introduction

Bioluminescence, i.e. the ability of various living species to emit light, is characterized at the molecular scale by a complex biochemical reaction catalyzed by an enzyme called luciferase which eventually produces a light emitter called oxyluciferin¹⁻⁴. In the case of firefly and some other beetles, the oxyluciferin molecule, 2-(6'-hydroxybenzothiazol-2-yl)-4-hydroxythiazole, shows a chemical structure featuring two deprotonation sites and one tautomerization site⁵. Depending on the pH^{6,7} and on the electrostatic interactions with its surroundings^{8,9}, for instance with adenosine-monophosphate (AMP, Figure 1A) which is also produced during the bioluminescence reaction, oxyluciferin can formally exist in 6 different forms (Figure 1B), all being possibly involved into the light emission process¹⁰.

Accordingly, the elucidation of the oxyluciferin light-emitting form(s) has been, and still is, a long-lasting debate which fostered numerous experimental and theoretical investigations¹¹⁻¹³. While the phenolate-keto form is the most likely emitter up-to-date, the phenolate-enol and phenolate-enolate forms still remain potential candidates¹⁴. In order to disentangle the deprotonation

^a Aix-Marseille Univ, CNRS, ICR, Marseille, France. Tel: +33 413945889; E-mail: nicolas.ferre@univ-amu.fr

^b Université de Strasbourg, UMR 7021 CNRS, Laboratoire de Bioimagerie et Pathologies, Strasbourg, France.

[†] Electronic Supplementary Information (ESI) available: CpHMD parameters, de-tails of the titrating residue selection, titration curves, trajectory analysis, analysis of the interactions. See DOI: 10.1039/cXCP00000x/

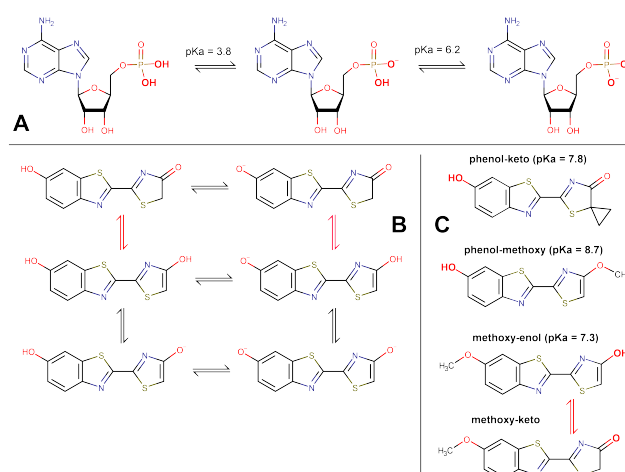


Fig. 1 A: adenosine monophosphate neutral, singly deprotonated and doubly protonated forms with titratable sites in bold. B: 6 possible forms of oxyluciferin (black double arrows indicate acid-base equilibrium, red ones indicate tautomerization). C: structures of the 3 considered titrating analogues together with the keto tautomer of the last analogue.

and tautomerization chemical reactions that are controlling the relative populations of the 6 oxyluciferin forms, several chemical analogues have been proposed, synthesized and studied spectroscopically and computationally¹⁵⁻²². In the present work, we focus on 3 analogues (Figure 1C) which have been characterized experimentally¹⁷. At variance with two of them (phenol-keto, phenol-methoxy), the third one (methoxy-enol) can un-

dergo tautomerization and co-exist with a fourth form (methoxy-keto) which does not titrate. Overall, these analogues are particularly suited to the investigation of pH effects.

The most important spectroscopic features of these analogues in water have already been determined by means of steady-state UV-visible absorption and emission spectroscopies^{15,17}, time-resolved spectroscopies²³, as well as using computational approaches, mainly rooted in the time-dependent density functional theory^{22,24}. Changing the pH has a clear impact on the analogue absorption spectra. When protonated, all analogues absorb in the 300–450 nm range with λ_{max} being about 380 nm. At pH values larger than 8, the phenol-keto form is fully deprotonated and exhibit a first absorption band with λ_{max} being 480 nm, ie 100 nm red-shifted with respect to the protonated form. Regarding the phenol-methoxy and the methoxy-enol analogues, their protonated and deprotonated corresponding spectra only differ by 50 nm. Considering the possible interaction between the analogues and AMP (still in water), it has been shown that AMP does not modify significantly the absorption spectra, at all pH values between 5 and 11, even when a bounded complex is formed between AMP and the analogue²².

Since the analogue photophysical properties are dependent on their environment, the next step consists in studying them when inserted into their natural host: the luciferase protein. In the present article, we report the results of our investigation of pH effects on the absorption spectrum of the 3 oxyluciferin analogues in the *Photinus pyralis* firefly luciferase, in the presence of AMP. To this end, we used steady-state absorption spectroscopy and the recently-reported CpHMD-then-QM/MM modeling protocol^{25,26} which enables to sample both the conformational and the protonation state spaces (Constant-pH Molecular Dynamics, CpHMD)^{27,28} before running thousands of hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) calculations²⁹.

2 Experimental Section

2.1 Preparation of oxyluciferin analogues in luciferase, experimental absorption spectra

Oxyluciferin analogues have been synthesized following the procedure previously reported¹⁷. Stock solutions of all compounds were prepared in spectroscopic grade DMSO (Sigma-Aldrich) and stored in several aliquots at 4°C. The buffers were prepared by diluting a 1M stock solution in deionized, Millipore-purified water (18.2M Ω). Buffered stock solutions were prepared as follows: 75mM NaCl/20mM KH₂PO₄/0.2mM MgCl₂ for pH=6 and 75mM NaCl/25mM tris(hydroxymethyl)aminomethane (TRIS)/0.2mM MgCl₂ for pH=9, and PBS for pH=7.5. The pH of the buffers was adjusted with 250mM HCl or with 250mM NaOH separately at 20°C. Luciferase was expressed in BL-21 (DE3*) *E. coli* bacteria cells using *Photinus Pyralis* plasmid and purified according to a previously published protocol³⁰. After purification, final solution was mixed with 10% of glycerol and aliquots were frozen with liquid nitrogen and stored at -80°C. The compounds stock solutions were finally diluted in the prepared aqueous buffer of different pH to a final concentration of 1 μ M (5% DMSO) with 10 μ M of luciferase and 3mM of AMP. A analogue/luciferase ratio

of 1/10 was used to avoid the contribution of unbound analogues. Absorption spectra were recorded with a Cary-4000 spectrometer (Agilent Technologies). The abrupt change in absorbance signal around 350 nm is an experimental artefact associated to the switch between the UV and visible lamps used in the spectrophotometer. The effect is more pronounced for diluted samples.

2.2 CpHMD and molecular parameters

In this work, 8 different systems were modeled considering the 3 different synthetic oxyluciferin analogues with one of them featuring two different tautomers, AMP in two possible acid-base equilibrium (neutral/singly deprotonated or singly deprotonated/fully deprotonated, respectively denoted AMP1 or AMP2 in the following) and the crystallographic protein structure of firefly luciferase in its second catalytic conformation (PDB 4G37)³¹. Similar setup was used in a previous study³². Aiming to parametrize non-standard residues as titratable ones in Amber18³³, partial atomic charges and reference energies were derived using the restrained electrostatic potential (RESP) method and the finddref tool respectively. Parametrization has been performed using molecular structures optimized at the B3LYP /6-311G(2d,p) level of theory when not available in previous studies^{21,22}. Parameters are reported in ESI first section†.

These systems were solvated in a cubic water box ensuring a distance of 20 Å between the protein and the box side, using the TIP3P water model. Each system was then equilibrated in the following way: (i) 20000-step energy minimized with restraints on the protein backbone atoms (5 kcal/mol/Å²), (ii) slowly heated up to 300 K and (iii) equilibrated for 5 ns without any restraints. In order to calculate the microscopic pK_a values of selected residues (including the analogue and AMP) along the pH interval of interest, we performed 6 CpHMD production runs of 60 ns for pH values between 6.5 and 9.0. Because of some convergence problems, the Replica Exchange protocol was used in the case of the phenol-keto/AMP1 system in order to improve the sampling of both the conformational and the protonation state spaces. All MD simulations were performed with Amber³³ software, using the ff14SB³⁴ force field parameters.

2.3 Iterative selection of the titrating residues

The very large number (157) of titratable residues in luciferase (Figure 2) does not allow a complete titration of the protein due to the computational cost of such calculation (the total number of protonation microstates is larger than 10⁴⁷). Furthermore, in the pH window we are interested in, some residues will remain either deprotonated or protonated. Thus, the careful selection of the most relevant residues to titrate is essential. Owing to our previous experience with Anabaena Sensory Rhodopsin, the application of our minimal electrostatic model³⁵ has shown that titratable residues located very far from the chromophore or at the border with the protein surroundings does not contribute much to the absorption spectrum. In the case of luciferase, such a tiny effect is fully screened by the one due to the titratable residues closer to the oxyluciferin analogue. Consequently, the selection process excludes titratable residues not close enough to the chro-

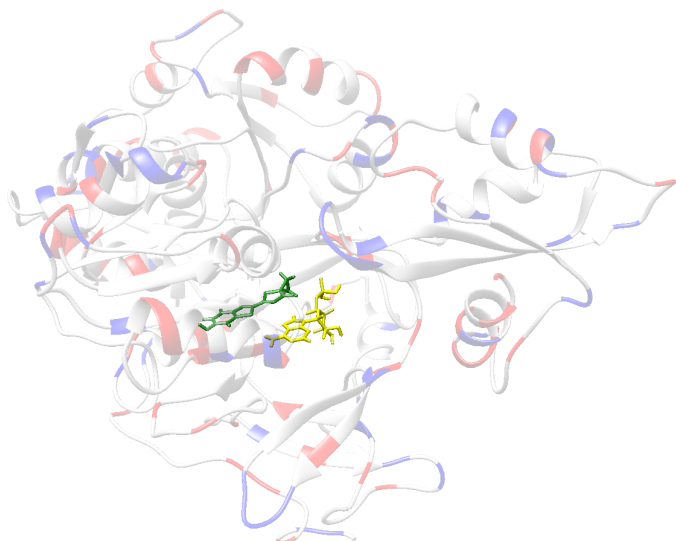


Fig. 2 Graphical representation of the luciferase protein, including all titratable residues in blue (positively charged) and red (negatively charged). AMP is colored in yellow and the oxyluciferin analogue is colored in green.

mophore and, for keeping the computational cost affordable, at most 20 residues have to be selected.

Besides the previous electrostatic point of view, the pK_a of the relevant residues should be such that they titrate in the 6.5 to 9.0 pH range we are interested in. Considering the pH-dependent deprotonated fraction x_i of the residue i ³⁶:

$$x_i(\text{pH}) = \frac{1}{1 + 10^{n_i(\text{pH} - pK_a^i)}} \quad (1)$$

we assume that the residue i co-exists in both the protonated and deprotonated forms (ie it titrates) if its x_i ranges from 0.1 to 0.9. According to Equation (1), this condition translates to $pK_a^i - 1/n_i \leq \text{pH} \leq pK_a^i + 1/n_i$.

Since the Hill factor n_i reflects the possible interactions of i with other titrating residues, as well as structural fluctuations of its surroundings, we have first estimated n_i using several representative geometries. Clustering of MD trajectory snapshots (using heavy-atom root mean square deviation lower than 1.8 Å as a selection threshold) has been performed in order to extract N (typically 10 to 20) clusters. An average deprotonated fraction is then calculated according to:

$$\langle x_i(\text{pH}) \rangle = \sum_{k=1}^N w_k x_{i,k}(\text{pH}) \quad (2)$$

where w_k is the statistical weight of the cluster k and $x_{i,k}(\text{pH})$ is the deprotonated fraction of the residue i in the fixed environment of the cluster k :

$$x_{i,k}(\text{pH}) = \frac{1}{1 + 10^{(pK_a^{i,k} - \text{pH})}} \quad (3)$$

Finally, for each titrating residue a Hill factor n_i is determined by fitting this mean deprotonated fraction with the Hill equation already discussed (Equation 1). All $pK_a^{i,k}$ value are obtained using the PROPKA3.1 software.³⁷

Using this set of n_i Hill factors, the workflow of the titrating residue selection goes as follows. First, we carry out a pK_a calculation on the most representative clusters extracted from a standard (fixed microstate) equilibrated MD trajectory. Second, a preliminary CpHMD run is performed considering a reduced set of 5 titrating residues. Afterwards, a new clustering is performed using the CpHMD trajectories before to calculate new pK_a values of all titratable residues. After examination of these pK_a values, the final set of titrating residues is selected (14 to 16 residues, depending on the considered system, the list is given in ESI Section 2†). Eventually a final CpHMD computation is performed.

2.4 QM/MM computations

Considering 30000 equally spaced snapshots extracted from the CpHMD trajectories, we calculated their QM/MM transition energies. The QM/MM partitioning defined the analogue and AMP as part of the quantum mechanics (QM) subsystem and the protein and solvent as part of the molecular mechanics (MM) subsystem. In order to describe the electrostatic interaction between QM/MM subsystems, we have included the MM external potential as extra diagonal elements in the semi-empirical PM7 Fock operator. Vertical excitation energies and transition oscillator strengths have been obtained using the Multi-Electron Configuration Interaction (MECI)³⁸ implemented in MOPAC 2016 program³⁹. The absorption spectra have been generated using the Newton-X 2.0 package⁴⁰.

3 Results and Discussion

3.1 CpHMD titration analysis

The pK_a values of the titrating sites (selected amino-acids, oxyluciferin analogue, AMP) can be obtained from CpHMD calculations by fitting individual microstate population curves using a generalized Henderson-Hasselbach equation in which collective effects are modeled thanks to the Hill factor (see ESI Section 3†). In order to understand the environmental effect on the pK_a values of oxyluciferin analogues and AMP, we have considered two different references: the free molecules in water and the analogue-AMP complexes in water. According to our previous study²², aqueous AMP only exists in deprotonated forms in the pH range we are interested in. In other words, only the second deprotonation equilibrium AMP2 is relevant in water (in the 6 to 9 pH range). According to Table 1, it turns out that the analogue pK_a values do not change much when they are interacting with AMP. Conversely, AMP pK_a always increases when it is associated with one analogue. Nevertheless, all analogues and AMP are titrating in the 6.5 to 9.0 pH window.

However, when AMP and one analogue are inserted into the luciferase cavity, their pK_a values change dramatically.

- *AMP + phenol-keto*. This analogue is characterized by a pK_a value which is almost independent on its environment. Being 7.8 in water, it shifts down to 7.2 when complexed with AMP (AMP2 model), still in water. Then, when it interacts with both AMP and luciferase, its pK_a value increases a bit, reaching 7.5 or 7.6, depending on the AMP model. On the other hand, AMP is characterized by two pK_a values

Table 1 Calculated oxyluciferin analogues and AMP pK_a values (AMP1 - first deprotonation and AMP2 - second deprotonation) using CpHMD calculations ("free" column: pK_a values of each free molecule in water; "complex": pK_a values of analogue-AMP complexes in water; "protein": pK_a values of analogue and AMP inside the luciferase cavity). Experimental pK_a values are 3.8/6.2, 7.8, 8.7, 7.3 for AMP, phenol-keto, phenol-methoxy, methoxy-enol, respectively.

	free	complex	protein
phenol-keto / AMP1	7.8 / 3.8	7.8 / < 5.0	7.5 / 6.8
phenol-keto / AMP2	7.8 / 6.2	7.2 / 6.7	7.6 / > 9.0
phenol-methoxy / AMP1	8.7 / 3.8	8.7 / 5.0	> 9.0 / > 9.0
phenol-methoxy / AMP2	8.7 / 6.2	8.7 / 7.7	> 9.0 / > 9.0
methoxy-enol / AMP1	7.3 / 3.8	7.7 / < 5.0	> 9.0 / < 6.0
methoxy-enol / AMP2	7.3 / 6.2	7.3 / 6.7	> 9.0 / > 9.0
methoxy-keto / AMP1	- / 3.8	- / < 5.0	- / 6.2
methoxy-keto / AMP2	- / 6.2	- / 6.7	> / > 9.0

which are far from the ones obtained previously in water: 6.8 (AMP1) and > 9.0 (AMP2). In the 6.5–9.0 pH range: (i) the doubly-deprotonated form of AMP does not exist, (ii) both phenol-keto and AMP are titrating, both molecules being either electrically neutral or negatively charged (-1).

- *AMP + phenol-methoxy.* The pK_a values of the analogue and AMP (whatever the model, AMP1 or AMP2) are calculated to be larger than 9.0. Accordingly, the two molecules do not titrate and are always fully protonated (electrically neutral) in the considered pH range.
- *AMP + methoxy-enol.* Similar to the phenol-methoxy analogue, the methoxy-enol one features a pK_a value larger than 9.0. Conversely, AMP pK_a depends on the model: smaller than 6.0 in the case of AMP1, larger than 9.0 in the case of AMP2. Hence, methoxy-enol is always protonated while AMP remains singly deprotonated between pH=6.5 and pH=9.0.
- *AMP + methoxy-keto.* Of course, this analogue does not titrate. However AMP first deprotonation occurs at much higher pH with respect to its reference pK_a while AMP second deprotonation (that we did not consider in CpHMD) should not occur in the considered pH range, given the results obtained for the other analogues.

In summary, only phenol-keto is titrating with a pK_a equal to 7.8 while the other two analogues do not titrate in the pH range considered for this study. These results highlight not only the importance of which oxyluciferin analogue is present inside of the luciferase active site, but also that thorough consideration of AMP protonation state is required, its pK_a value being analogue-dependent. Of course, these differences demonstrate the sensitivity of the pK_a value to the fine details of the interactions between the analogue and its surroundings: AMP, close titrating or non-titrating residues, luciferase conformations, ... as demonstrated in the following.

3.2 Analysis of the interactions in the active site

We hereafter present the most important interactions that can be attributed to the presence of the oxyluciferin analogue within the luciferase active site. We have selected some important residues using the criteria that their distances from the analogue and AMP titratable groups must be less than 16 Å (further titratable residues would always contribute negligibly to the pH-dependent absorption spectrum), as represented in Figure 3. As a first note,

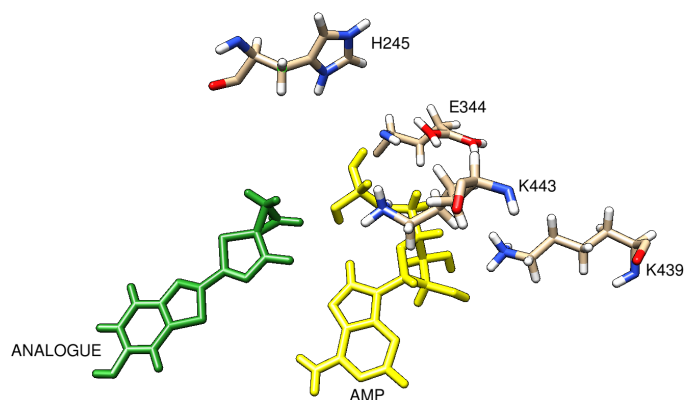


Fig. 3 Graphical representation of the 4 most important titrating residues around the oxyluciferin analogues (herein, the phenol-keto one) and AMP.

it should be mentioned that most of the selected titrating residues are closer to AMP than to the analogue. In the framework of the minimal electrostatic interaction model³⁵, this explains why AMP pK_a is much more pH-dependent than the analogue one (Table 1).

Small distances between AMP and titrating residues call for the consideration of possible proton exchange between titrating sites. Plotting protonation state populations along each CpHMD trajectory does not evidence any prevalence of such a mechanism (see ESI section 4†). The distance-based analysis is represented graphically in Figure 4. A thorough description of the analogue or AMP atoms participating to these interactions is included in ESI section 5†.

Starting with the phenol-keto analogue (Figure 4) top-left panel), the principal interactions are found along a rather long pathway, E311–R337–analogue–K443–AMP–E344, with an average inter-residue distance of 5 Å. On average, the analogue is slightly more often deprotonated than protonated, its negative charge being stabilized by R337 positive charge which, in turn, interacts with deprotonated E311. The analogue keto side also stays close to H245, K443 and AMP, highlighting the steric hindrance due to the cyclopropyl moiety (Figure 1C). AMP mainly carries a negative charge which is stabilized by its interaction with lysine K443 and to a lesser extent, K439. Interestingly, a close contact is evidenced between AMP and E344, both being mainly deprotonated, hence negatively charged.

In the case of phenol-methoxy analogue (Figure 4 top-right panel), several sets of close interactions ranging from 4 to 7 Å are found. The first one, very stable in the considered pH range, involves H245, the analogue and AMP, all of them being electrically neutral. The negative charge carried by E344 is mainly counterbalanced by K439 positive one. A third contact is found

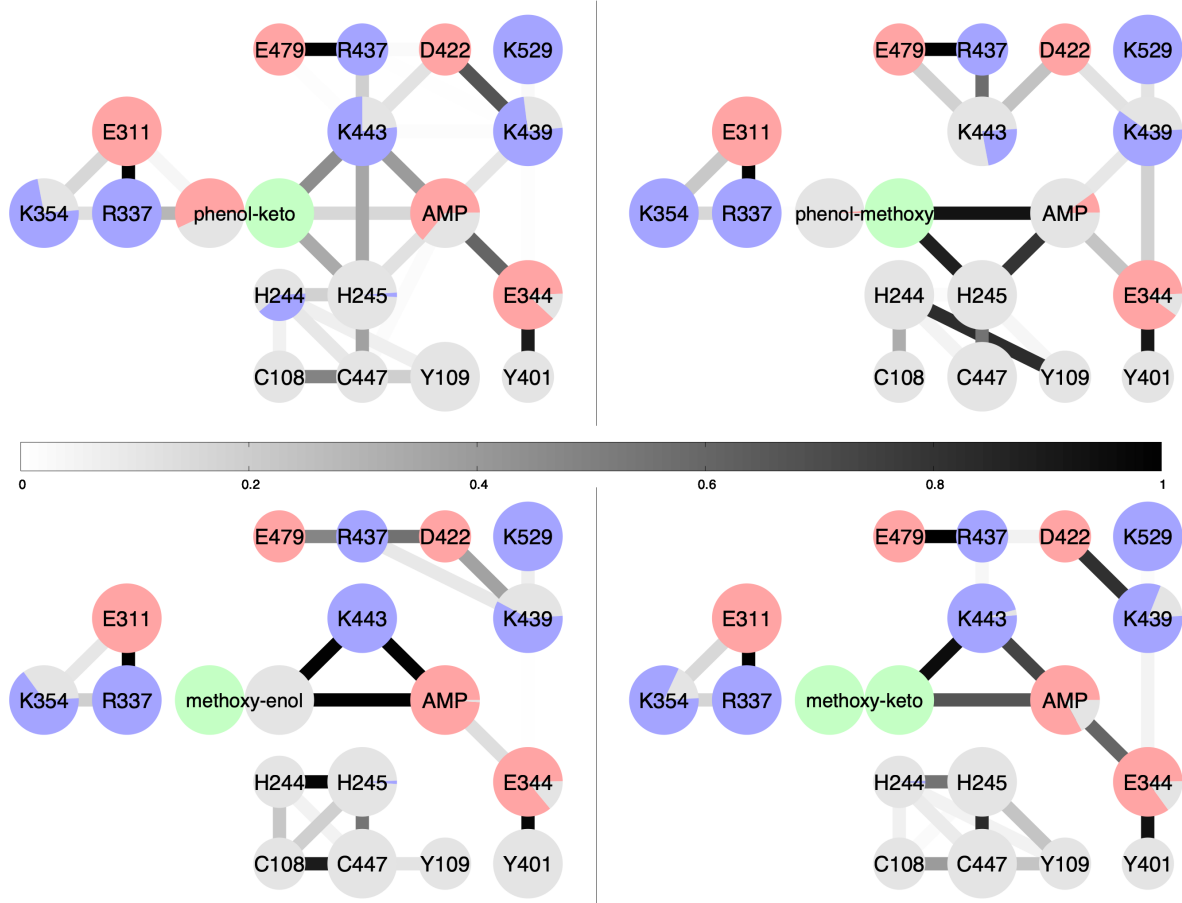


Fig. 4 Distance-based interaction analysis between analogues, AMP and relevant titratable residues. Blue, grey and red colors indicate positive, neutral and negative charged residues. Circle sub-surfaces are proportional to the average (de)protonated fraction. Analogue non-titratable moieties are represented in green. Inter-residue interactions are represented by segments whose color (from light gray to black) is proportional (central bar) to the fraction of close contacts ($< 5 \text{ \AA}$). Residues with titratable neighbors not reported on this diagram are represented by a reduced size circle, eg E479, D422, ...

between the electrically neutral H244 and Y109. K443 being 80% neutral, it remains close to the R437/E479 ion pair.

The methoxy-enol analogue and its tautomer form, methoxy-keto, share a common strong contact between the analogue, K443 and almost always deprotonated AMP. As a matter of fact, whatever its form, the analogue remains electrically neutral. AMP is more strongly interacting with E344 in the methoxy-keto case. Other electrically neutral groups can be found: E311–R337, R437–E479, D422–K439.

The general picture that can be drawn from this analysis involves local charge compensations, achieving local electroneutrality. For instance, when the analogue is partially negatively charged (phenol-keto), it interacts with R337. If not, the analogue is found closer to AMP (phenol-methoxy, methoxy-enol, methoxy-keto). Actually, the AMP case is more complex: depending on the analogue, it changes from neutral (phenol-methoxy) to partially negatively charged (phenol-keto and methoxy-keto) to fully deprotonated (methoxy-enol). If its charge neutralization is mainly due to K443, AMP is also interacting quite strongly with negatively charged E344, preventing local electroneutralization to occur. It turns out that steric hindrance explains this

behavior. The atomic density around E344 (Figure 5) is much higher than the one calculated for the other residues shown in Figure 3. This constrained environment induces concerted motions of the E344+AMP+Y401 cluster, the latter residue being hydrogen-bonded with E344 (Figure 6). Based on this analysis, we emphasize that the proximity of E344 negative charge with AMP explains why AMP second deprotonation does not occur in luciferase, at variance with what is found in water²².

Having considered the principal structural features within the full pH window, we now turn to a more detailed analysis of the pH effect on the analogue and AMP environment. Knowing the role that solvent has on protein structural stability and function, it is expected that pH can influence the water local concentration inside the active site of luciferase. Such property is monitored by means of radial distribution function (RDF) analysis between the oxygen atom of the analogue titratable group and the oxygen atom of water molecules (Figure 7). Given their identical titratable moieties, phenol-keto and phenol-methoxy analogues exhibit a first solvation shell with a maximum at the same distance of about 2.7 \AA . However, while the number of water molecules is pH-independent in the latter analogue (slightly larger than 1), this

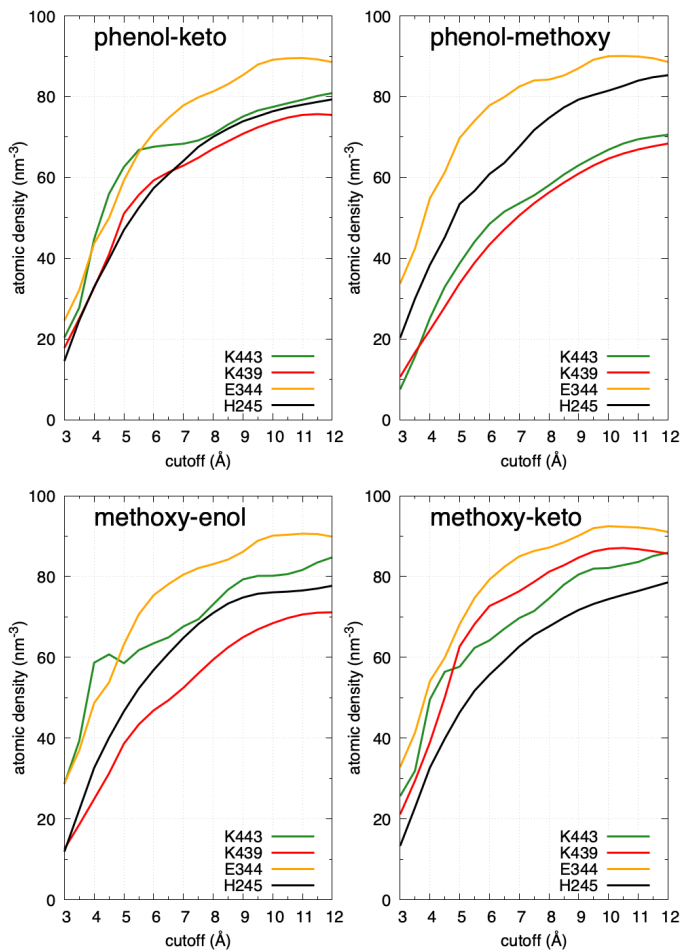


Fig. 5 Atomic density analysis for each analogue.

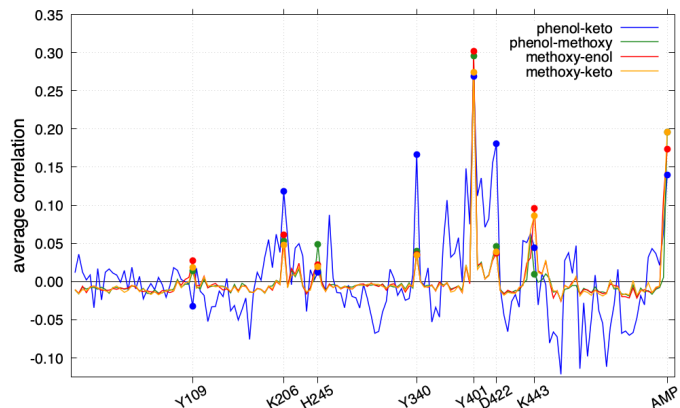


Fig. 6 E344 motion correlated with other residues in luciferase. In the case of phenol-keto, the replica-exchange technique used in the CpHMD calculation resulted in a larger sampling of the conformational space, explaining the observed larger motion correlations. In all, the main correlations are the same for all analogues and involve the residues explicitly mentioned below the x-axis.

value is much more changing in the former one: from less than 1 to more than 2 with increasing pH. This different behavior reflects the phenol deprotonation process in phenol-keto while phenol-methoxy remains neutral in the pH window. The methoxy-enol

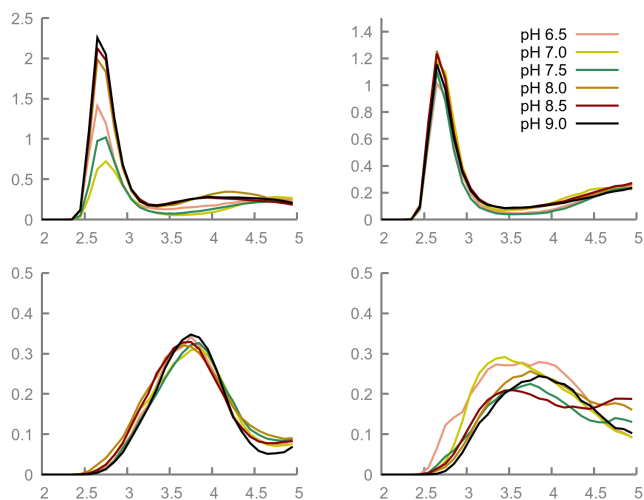


Fig. 7 Normalized radial distribution function of the analogue titrating O10/O11 water O interatomic distance (in Å): phenol-keto (top left), phenol-methoxy (top right), methoxy-enol (bottom left), methoxy-keto (bottom left). Analogue atom numbering available in ESI—

analogue shows a different solvent organization when compared with the other two analogues: its first solvation shell is located at around 3.8 Å from its titrating enol group and hosts less than 0.4 water molecules. Such a difference can be attributed to the analogue orientation in the luciferase cavity, its enol group facing the AMP deprotonated phosphate moiety. Similarly to methoxy-enol, methoxy-keto does not interact with much water molecules. Moreover, these interactions are weaker, resulting in a less structured solvent organization.

To complement the previous analysis, we have calculated the distribution of distances between the analogue (still using the oxygen atom carrying the labile proton) and AMP phosphate P atom (Figure 8), to better understand the possible motion that each analogue can experience, as a function of pH. Regarding the

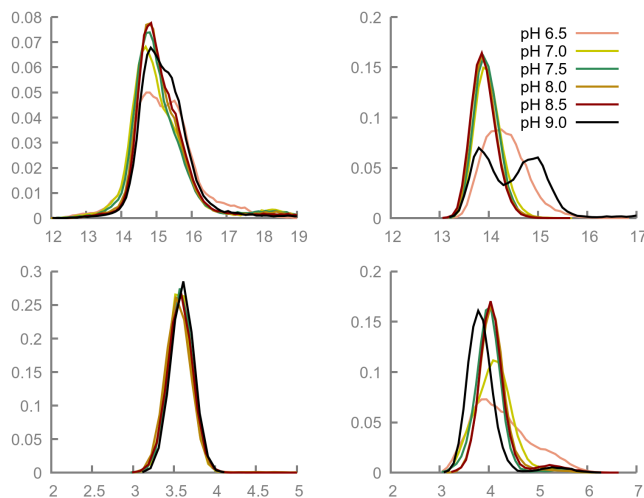


Fig. 8 Normalized radial distribution function of the analogue O10/O11 AMP PA interatomic distances (in Å): phenol-keto (top left), phenol-methoxy (top right), methoxy-enol (bottom left) and methoxy-keto (bottom right). Analogue atom numbering are available in ESI—

phenol-keto analogue, two contributions are found: the largest one is centered at around 14.8 Å, while the second one, only appears at 15.7 Å as a small shoulder for pH values 6.5 and 9. Phenol-methoxy-AMP presents 2 distinct populations at pH 9 (13.8 and 15 Å), whereas for the remaining of the considered pH range, only single distributions are observed at approximately 13.9 and 14.3 Å. Such a behavior is difficult to explain: both the analogue and AMP do not titrate. However, we can notice large variations of two lysine protonation state populations, K439 and K443 which, in turn, may induce some local structural modifications responsible for these modified phenol-methoxy-AMP interactions. The methoxy-enol shows virtually no pH-induced variation of its distance with AMP, its position is particularly stable inside of the cavity due the strong interactions with AMP and K443 (Figure 4). Methoxy-keto demonstrates almost no variations of its distance with AMP, whatever the pH value. The broader distribution found at pH=6.5 can be attributed to AMP titration, its pK_a being 6.2 (Table 1). Such analogue- and pH-dependent behaviors can be rationalized using even finer details in the interactions between the analogue, AMP and their surroundings. The first one involves the translation and rotation of the phosphate group in AMP (ESI section 5†). Actually, it turns out that AMP can exist in two different principal conformations, folded and extended as represented in ESI section 5†. The equilibrium between these two structures is particularly apparent in phenol-keto, in which the cyclopropyl moiety is preventing the analogue to interact with AMP. Also apparent in the analysis of the atomic density around the analogue (Figure 9) in which the AMP contribution is much lower for the phenol-keto analogue, it calls for a closer look to the cavity in which the analogue lies. Selecting the residues which are (i) the closest to AMP and the analogue, hence defining the cavity, and (ii) whose side-chains are clearly mobile (the contact analysis reported in ESI section 5† shows that the backbone atoms belonging to the cavity are rather fixed while the side-chain ones are not), the analysis of their motions demonstrates that the cavity is flexible enough to adapt its shape to the presence of any of the analogues.

3.3 Experimental and computational absorption spectra

In this section, we compare the experimental and theoretical absorption spectra of the analogues and AMP1 within the active site of the enzyme luciferase, at pH values ranging from 6.5 to 9.0. These spectra are reported in Figures 10, 11 and 12. Corresponding maximum absorption wavelengths are summarized in Table 2.

Remembering that phenol-keto is the only analogue titrating in the considered pH window, we expect its absorption spectrum to change dramatically with the pH. This is indeed the case, either looking at the experimental spectrum or at the calculated one (Figure 10). At acidic pH, two intense experimental bands appear with maximum wavelengths at 400 nm and 355 nm (Table 2), that could either correspond to the two first allowed electronic transitions, quite energetically close, or to some vibronic progression in the first allowed transition. In the same 300 to 450 nm window, the computed spectrum only shows one band

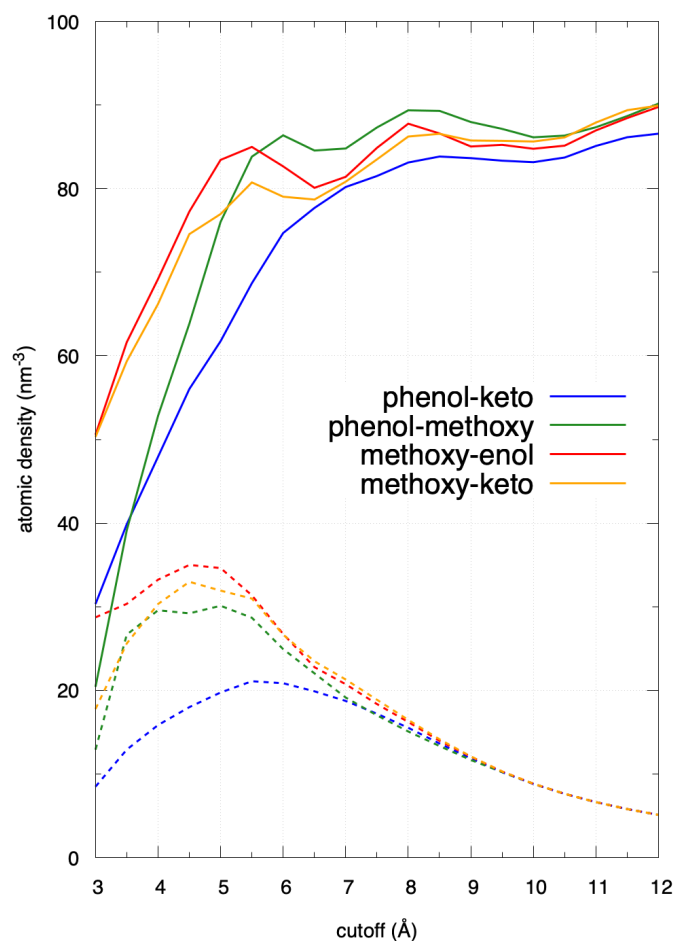


Fig. 9 Atomic density around the each analogue (using the oxygen atom closest to AMP as the reference). Plain lines: full densities; dashed lines: AMP contributions.

Table 2 Experimental and computed maximum absorption wavelengths in nm. In the case of the methoxy-enol analogue, the experimental spectrum includes the contribution from its tautomer, methoxy-keto, while both contributions are computed separately.

Analogue	Acidic pH		Neutral pH		Basic pH	
	Exp	Comp	Exp	Comp	Exp	Comp
phenol-keto	355	351	352	365	350	384
	400	562	399	505	483	481
phenol-methoxy	381	368	381	366	396	367
methoxy-enol	386	373	387	372	386	372
methoxy-keto		344		349		347

with maximum wavelength at 351 nm. Using the semi-empirical PM7 Hamiltonian does not allow to calculate possible vibronic couplings. Nevertheless, some of the vertical excitation calculations have been repeated using Time-Dependent Density Functional Theory B3LYP/6-311G(2d,p) on a restricted set of 1000 structures. Indeed, the corresponding spectrum shows two transitions with maxima at 371 and 400 nm (see ESI section 6†), in agreement with the experimental spectrum. We shall also notice

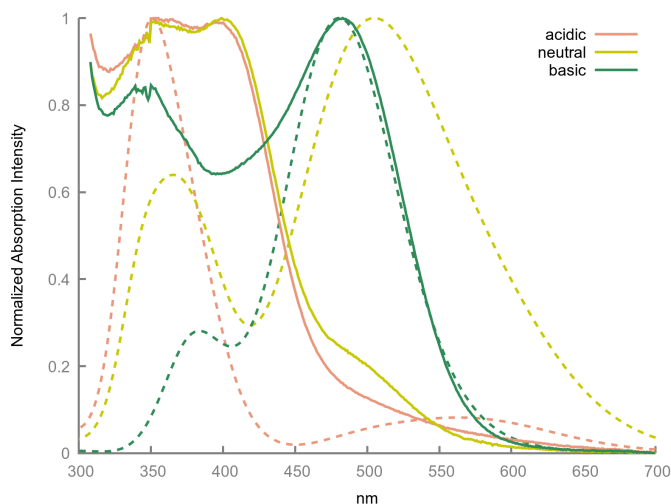


Fig. 10 Normalized experimental (full lines) and calculated (dash lines) absorption spectra of phenol-keto analogue. Experimental pH values: 6.0, 7.4 and 9.0; theoretical pH values: 6.5, 7.5 and 9.0.

the broad (450 to 700 nm) but not intense computed band. Analyzing the protonation microstate populations coming out of the corresponding CpHMD trajectory, it turns out that the very mod-est phenolate-keto population in acidic conditions is responsible for this band. At pH=9.0, the experimental spectrum presents a band with a maximum at 483 nm. This band is accurately reproduced theoretically. The CpHMMD-then-QMMM approach only overestimates by 2 nm the energy gap between the phenol-keto protonated and deprotonated absorption maxima. On the other hand, the computed spectrum features a much less intense band with maximum at 384 nm. This latter band also exists in the experimental spectrum at a much higher intensity and a maximum at about 350 nm. The pH=7.5 value is very close from the phenol-keto pK_a (Table 1). Accordingly, the corresponding absorption spectrum should appear as a combination of the acidic and basic ones. The experimental spectrum remains close to the acidic pH one, showing however a small shoulder (at about 500 nm) corresponding to the deprotonated form. Clearly the protonated form still prevails at neutral pH. However, the computed spectrum shows two distinct bands, hence indicating a more balanced population of the protonated and deprotonated forms. This theory vs experiment difference is certainly rooted into rather small deviations (0.1/0.2 units) between the calculated pK_a value and the experimental one. Given the very qualitative level of theory used in the QM/MM calculations, the small number of titrating residues and the rather short length of the CpHMD trajectories, the calculated spectrum is surprisingly in qualitative agreement with the experimental one.

According to the CpHMD results (Table 1), the phenol-methoxy does not titrate, nor does AMP. Accordingly, its absorption spectrum (Figure 11) should be mostly pH-independent. This isn't the case when looking at the experimental spectra. A single band appears at pH=6 and 7.5, with a maximum at 378 nm. Going to basic pH, the spectrum presents a large band, extending on more than 150 nm, with a maximum at 394 nm and a shoulder close

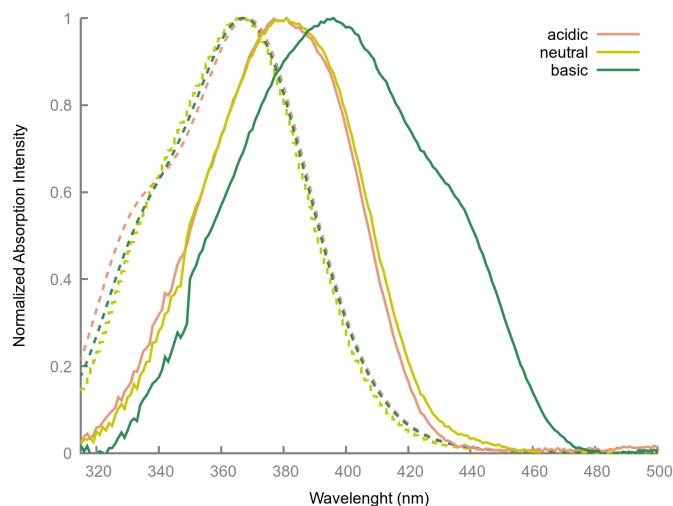


Fig. 11 Normalized experimental (full lines) and calculated (dash lines) absorption spectra of phenol-methoxy analogue. Experimental pH values: 6.0, 7.4 and 9.0; theoretical pH values: 6.5, 7.5 and 9.0.

to 440 nm. The 16 nm gap between these maxima is equivalent to the one reported for this analogue in water²². Accordingly, the spectrum at pH=9 reflects the presence of both its protonated and deprotonated forms and that the phenol-methoxy pK_a is probably not very different from 9 (and similar to its value in water: 8.7), at variance with the CpHMD results. Because of this inaccurate calculated pK_a , the computed spectra only show a single band, very close to the experimental one for the protonated phenol-methoxy analogue. Nevertheless, the quasi perfect superposition of the spectra for acidic and neutral pH demonstrates that the titration of other amino-acids does not modify the photophysical properties of phenol-methoxy.

The methoxy-enol analogue does not titrate in the considered pH window (Table 1) and interacts with singly-deprotonated AMP. Moreover, it can co-exist with its tautomer form, methoxy-keto, which cannot titrate either. Accordingly, its absorption spectrum shall be pH-independent. Indeed, the experimental spectrum (Figure 12) confirms this finding, showing a single band with a maximum wavelength at 391 nm. The spectra of the methoxy-enol form computed at the 3 different pH values share the same shape: a single band with a maximum at 372 nm, hence 19 nm blue-shifted with respect with the experimental value. The computed spectra of the tautomer form is also reported in Figure 12. Its single band shape is further blue-shifted by 23 nm with respect to the methoxy-enol one. There is no direct way to estimate the relative ratio between the methoxy-enol and the methoxy-keto forms using the results obtained in the present study. However, the computed maxima wavelengths suggest the predominance of the enol form. As a matter of fact, the experimental spectra show the apparition of some kind of noisy shoulder at wavelengths shorter than 350 nm. This may indicate the existence of the analogue keto form.

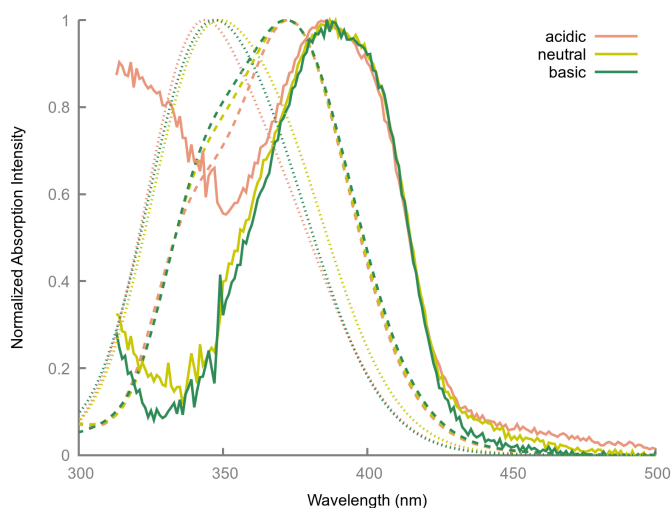


Fig. 12 Normalized experimental (full lines), calculated absorption spectrum of methoxy-enol analogue (dashed lines) and its methoxy-keto tau-tomer (dotted lines). Experimental pH values: 6.0, 7.4 and 9.0; theoretical pH values: 6.5, 7.5 and 9.0.

4 Conclusions

Using steady-state absorption spectroscopy and the CpHMD-then-QM/MM computational protocol, we have investigated the pH-dependent properties of three different oxyluciferin analogues inside luciferase, the enzyme involved in the light emission of firefly bioluminescence.

At variance with results already obtained with the same analogues in water^{17,22}, the present results evidence different behaviors for each analogue in the 6 to 9 pH range. Showing a pH-dependent spectrum in water and in luciferase, the phenol-keto analogue is titrating in this pH window. However, while AMP is also titrating, this molecule co-exists in neutral and singly-deprotonated forms in luciferase, instead of singly- and doubly-deprotonated forms in water. Indeed, the AMP second deprotonation is disfavoured in the protein because of the strong interaction between AMP and the negatively charged glutamate E344. The same applies to the other analogues for which the doubly-deprotonated AMP never exists.

In luciferase, the phenol-methoxy analogue spectrum is modified with the pH the same way it is in water. However, AMP remains always fully protonated for pH ranging from 6 to 9. While the CpHMD pK_a value of this analogue in luciferase is proposed to be much larger than 9, the modification observed in its experimental spectrum indicates a pK_a value not so different from 9.

The behavior of the third analogue, methoxy-enol, strongly depends on its environment. While it is titrating in water, this is no longer the case in luciferase: the enolate form is absent at any considered pH value. AMP does not titrate either, it is always singly-deprotonated. Accordingly, the absorption spectrum of this analogue, even if its methoxy-keto tautomer form is taken into account, does not change with the pH.

The absorption spectrum of the natural oxyluciferin in luciferase is pH-dependent⁷. According to our present results, especially the pH-independent absorption spectrum of the methoxy-

enol analogue, we would then suggest that the oxyluciferin enolate form does not exist as a product of the chemical reaction responsible for firefly bioluminescence. Nevertheless, we want to stress that the replacement of oxyluciferin with one analogue induces a significant reorganization of the flexible luciferase cavity, as the distance analysis has shown in the present work. Further developments would imply the characterization of the possible pH-dependency of the analogue fluorescence. However, the direct application of the CpHMD-then-QM/MM protocol is tedious, since it would require tens of thousands of excited state geometry optimizations. Progress along this line will be reported soon.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

All the authors thank the French Agence Nationale de la Recherche for funding (grant ANR-16-CE29-0013-01, project BIOLUM). Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high performance computing resources.

Notes and references

- 1 F. McCapra and M. J. Manning, *J. Chem. Soc., Chem. Commun.*, 1973, 467–468.
- 2 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 and C. Lowik, *Chemiluminescence and Bioluminescence*, The Royal Society of Chemistry, 2011, pp. P001–590.
- 3 J. Lee, *Photochem. Photobiol.*, 2017, **93**, 389–404.
- 4 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 and R. Lindh, *Chem. Rev.*, 2018, **118**, 6927–6974.
- 5 P. Naumov, Y. Ozawa, K. Ohkubo and S. Fukuzumi, *J. Am. Chem. Soc.*, 2009, **131**, 11590–11605.
- 6 E. E. Dahlke and C. J. Cramer, *J. Phys. Org. Chem.*, 2003, **16**, 336–347.
- 7 Y. Ando, K. Niwa, N. Yamada, T. Enomoto, T. Irie, H. Kubota, Y. Ohmiya and H. Akiyama, *Nat. Photonics*, 2007, **2**, 44–47.
- 8 T. Nakatsu, S. Ichiyama, J. Hiratake, A. Saldanha, N. Kobashi, K. Sakata and H. Kato, *Nature*, 2006, **440**, 372–376.
- 9 T. Hirano, H. Nagai, T. Matsuhashi, Y. Hasumi, S. Iwano, K. Ito, S. Maki, H. Niwa and V. R. Viviani, *Photochem. Photobiol. Sci.*, 2012, **11**, 1281–1284.
- 10 I. Navizet, Y.-J. Liu, N. Ferré, D. Roca-Sanjuán and R. Lindh, *ChemPhysChem*, 2011, **12**, 3064–3076.
- 11 G. Orlova, J. D. Goddard and L. Y. Brovko, *J. Am. Chem. Soc.*, 2003, **125**, 6962–6971.
- 12 L. P. da Silva and J. C. E. da Silva, *J. Chem. Theory Comput.*, 2011, **7**, 809–817.
- 13 L. P. da Silva and J. C. G. E. da Silva, *J. Phys. Chem. B*, 2015, **119**, 2140–2148.
- 14 V. R. Viviani, G. V. M. Gabriel, V. R. Bevilaqua, A. F. Simões, T. Hirano and P. S. L. de Oliveira, *Sci. Rep.*, 2018, **8**, 17594.

- 15 M. Rebarz, B.-M. Kukovec, O. V. Maltsev, C. Ruckebusch, L. Hintermann, P. Naumov and M. Sliwa, *Chem. Sci.*, 2013, **4**, 3803–3809.
- 16 X. Q. Ran, X. Zhou and J. D. Goddard, *ChemPhysChem*, 2014, **16**, 396–402.
- 17 A. Ghose, M. Rebarz, O. V. Maltsev, L. Hintermann, C. Ruckebusch, E. Fron, J. Hofkens, Y. Mély, P. Naumov, M. Sliwa and P. Didier, *J. Phys. Chem. B*, 2015, **119**, 2638–2649.
- 18 Y.-Y. Cheng and Y.-J. Liu, *J. Chem. Theory Comput.*, 2015, **11**, 5360–5370.
- 19 A. Ghose, O. V. Maltsev, N. Humbert, L. Hintermann, Y. Arntz, P. Naumov, Y. Mély and P. Didier, *J. Phys. Chem. B*, 2017, **121**, 1566–1575.
- 20 C.-G. Min, Q.-B. Liu, Y. Leng, S.-J. Huang, C.-X. Liu, X.-K. Yang, A.-M. Ren and L. P. da Silva, *Comput. Theor. Chem.*, 2018, **1133**, 18–24.
- 21 C. García-Iriepa, P. Gosset, R. Berraud-Pache, M. Zemmouche, G. Taupier, K. D. Dorkenoo, P. Didier, J. Léonard, N. Ferré and I. Navizet, *J. Chem. Theory Comput.*, 2018, **14**, 2117–2126.
- 22 N. M. de Almeida Barbosa, M. Zemmouche, P. Gosset, C. García-Iriepa, V. Ledentu, I. Navizet, P. Didier and N. Ferré, *ChemPhotoChem*, 2019, 1219–1230.
- 23 P. Gosset, G. Taupier, O. Crégut, J. Brazard, Y. Mély, K.-D. Dorkenoo, J. Léonard and P. Didier, *J. Phys. Chem. Lett.*, 2020, 3653–3659.
- 24 C. García-Iriepa, M. Zemmouche, M. Ponce-Vargas and I. Navizet, *Phys. Chem. Chem. Phys.*, 2019, **21**, 4613–4623.
- 25 E. Pieri, V. Ledentu, M. Huix-Rotllant and N. Ferré, *Phys. Chem. Chem. Phys.*, 2018, **20**, 23252–23261.
- 26 E. Pieri, V. Ledentu, M. Sahlin, F. Dehez, M. Olivucci and N. Ferré, *J. Chem. Theory Comput.*, 2019, **15**, 4535–4546.
- 27 J. M. Swails and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 4393–4404.
- 28 J. M. Swails, D. M. York and A. E. Roitberg, *J. Chem. Theory Comput.*, 2014, **10**, 1341–1352.
- 29 H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198–1229.
- 30 C. Carrasco-López, J. C. Ferreira, N. M. Lui, S. Schramm, R. Berraud-Pache, I. Navizet, S. Panjikar, P. Naumov and W. M. Rabeh, *Life Science Alliance*, 2018, **1**, e201800072.
- 31 J. A. Sundlov, D. M. Fontaine, T. L. Southworth, B. R. Branchini and A. M. Gulick, *Biochemistry*, 2012, **51**, 6493–6495. 32 C. García-Iriepa and I. Navizet, *Molecules*, 2019, **24**, 1565. 33 D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, I. T.E. Cheatham, V. Cruzeiro, T. Darden, R. Duke, D. Ghorishi, 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 and P. Kollman.
- 34 J. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 35 M. Stenrup, E. Pieri, V. Ledentu and N. Ferré, *Phys. Chem. Chem. Phys.*, 2017, **19**, 14073–14084.
- 36 E. Bombarda and G. M. Ullmann, *J. Phys. Chem. B*, 2010, **114**, 1994–2003.
- 37 C. R. Sndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 2284–2295.
- 38 D. R. Armstrong, R. Fortune, P. G. Perkins and J. J. P. Stewart, *J. Chem. Soc. Faraday Trans.*, 1972, **68**, 1839.
- 39 J. J. P. Stewart, *MOPAC 2016*, <http://openmopac.net/>, 2016. 40 M. Barbatti, M. Ruckebauer, F. Plasser, J. Pittner, G. Granucci, M. Persico and H. Lischka, *Rev.: Comput Mol Sci.*, 2014, **4**, 26–33.