

# PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

## QM/MM calculations on a new synthesised oxyluciferin substrate: new insights on the conformational effect<sup>†</sup>

Romain Berraud-Pache<sup>a</sup>, Isabelle Navizet<sup>\*a</sup>Received Date  
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

In this publication we conduct calculations on a new synthesised red-shifted emitter of the luciferin in order to understand what are the main contributions to the colour-shifting emission. Indeed the bioluminescence process, especially from fireflies, is one of the main source in medical imaging but its efficiency grandly depends on the wavelength of the emission. We performed classical molecular dynamics followed by quantum mechanics/molecular mechanics (QM/MM) calculations with either density functional theory or multiconfigurational reference second-order perturbation theory on different emitters to obtain the bioluminescence emission. We analysed the calculations and investigated the effects which play a non-negligible role on the emission, like the effect of the surrounding or the effect of the conformation of the emitter. Finally, in the absence of crystallographic structure, we proposed the luckiest conformation for the emitter in the bioluminescence process.

### Introduction

For several years, the modulation of the colour of the bioluminescence has been a trendy topic with a lot of publications either from an experimental or a theoretical point of view<sup>1–8</sup>. All kind

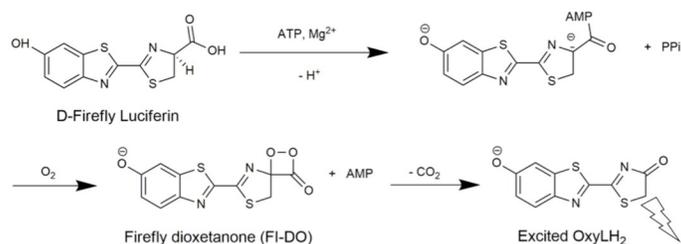
of emission colours can now be obtained, from the yellow-green tone of the wild-type substrate to a near IR one. Sadly, the mechanism of colour modulation is not well understood, despite some relevant insights<sup>9–15</sup>.

Fireflies are insects that can emit visible light with flashes. The bioluminescence is due to the interaction of a protein called luciferase with an organic molecule, the luciferin (LH<sub>2</sub>). The generally accepted mechanism of firefly bioluminescence is shown in Figure 1. During the first step, the protein catalyses the formation of a luciferyl-adenylate intermediate (luciferin-AMP). Then, the coordination of a dioxygen molecule to the intermediate leads to the formation of a dioxetanone compound. Thus, the decomposition of this dioxetanone gives an emissive molecule known as oxyluciferin (OxylH<sub>2</sub>) which releases the energy by emitting light<sup>16</sup>.

In 2014, the group of the Pr. Anderson and Dr. Pule published the synthesis of a new modified luciferin called infra-luciferin (iLH<sub>2</sub>)<sup>17</sup>. While the wild-type luciferin-luciferase system emits in the yellow-green range of the spectra, the iLH<sub>2</sub> molecule leads to a strong red emission during the bioluminescence process. This

<sup>a</sup> Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME, UMR 8208 CNRS, UPEM, 5 bd Descartes, 77454 Marne-la-Vallée, France; E-mail: isabelle.navizet@u-pem.fr

<sup>†</sup> Electronic Supplementary Information (ESI) available: Graphical representation of the molecule DLSA and the complex OxylH<sub>2</sub> + AMPH. TD-DFT/MM results of  $T_e$  with 6-31G(d,p) basis set. TD-DFT/MM results of  $T_e$  with 6-311G(2d,p) basis set for different functionals. TD-DFT/MM results of  $T_e$  with 6-311G(2d,p) basis set for different solvents. CAS orbitals of all emitters. MS-CASPT2/MM results of  $T_e$  with 2 and 4 roots used in the calculation with an IPEA shift of 0.25. Comparison of MS-CASPT2/MM results of  $T_e$  with 2 different IPEA shifts (0.25 and 0.1). Representation of the H-bond network in Ser\_OxylH<sub>2</sub> before and after classical molecular dynamics. Graphical representation of the free energy of the system during the MD on Ser\_OxylH<sub>2</sub>. Graphical representation of the dihedral angle N-CA-CB-OG in the Serine 284 during the MD of Ser\_OxylH<sub>2</sub>. QM/MM results of  $T_e$  for Ser\_OxylH<sub>2</sub> before and after classical molecular dynamics. Representation of the H-bond network of the model Ser\_OxylH<sub>2</sub>-2 and Ala\_OxylH<sub>2</sub>-2. Comparison of the structures for OxylH<sub>2</sub>. Schematic representation of OxylH<sub>2</sub>-Z-1 and OxylH<sub>2</sub>-Z-2, TD-DFT/MM results of  $T_e$  with 6-311G(2d,p) basis set for OxylH<sub>2</sub>-Z-1 and OxylH<sub>2</sub>-Z-2, Cartesian coordinates of OxylH<sub>2</sub> of the 6 models presented in Table 4. See DOI: 10.1039/b000000x/



**Fig. 1** Mechanism of firefly bioluminescence

range of emission in the red region of the visible spectra allows better penetration depths when used for medical imaging. By adding a double bond between the 2 cycles of the luciferin, the emission is red-shifted by 0.37 eV (112 nm). Furthermore, the authors also observed that by mutating the residue Serine 284 to Threonine (S284T) in the protein, which is only one methyl group longer, the emission for both the wild-type or the infrared luciferin is slightly red-shifted, between 0.1 eV and 0.2 eV. In this paper we characterise the interactions between the light emitter and the protein by using Quantum Mechanics/Molecular Mechanics (QM/MM) calculations with Time-Dependent Density Functional Theory (TD-DFT) or multi-configuration methods in order to rationalise the observed light colour shifting.

## Computational details

In the present publication we used a crystallographic structure (PDB 4G37) of the luciferase complex from the group of Prof. Branchini<sup>18,19</sup>. This structure comes from the North America firefly and has been chemically engineered. It has been designed to reproduce the conformation of the protein when the dioxygen binds to the luciferin. This conformation was created thanks to a disulfur bridge between two residues Isoleucine 108 and Tyrosine 447.

The PDB file 4G37.pdb was downloaded from the RSCB PDB website. Thus the missing loop was added with Disgro program<sup>20</sup>. Then, the 5'-O-[N-(dehydrolyciferin)-sulfamoyl]adenosine (DLSA) residue was replaced by adenosine monophosphate (AMP) and OxyLH<sub>2</sub> or OxyiLH<sub>2</sub>, the oxyluciferin resulting from iLH<sub>2</sub> (see Figure S1). The two SO<sub>4</sub><sup>-</sup> groups and the bridging residue XLX were removed.

The residues were protonated using Leap from Amber14 suite of program<sup>21</sup>. The contentious cases, especially for histidines were resolved by computing their pKa with the H++ program<sup>22</sup>. We have chosen to double protonated the following histidines: residues 76, 171, 221, 310, 332, 419, 461 and 489. AMP (charged -2) was protonated into AMPH (charged -1) and a Cl<sup>-</sup> was added in order to achieve a neutral charge for the system.

Classical dynamics simulations were made with Amber14 in order to obtain one snapshot for further QM/MM optimization. The model was solvated with TIP3P water molecules within a cuboid

box, ensuring a solvent shell of at least 15 Å around the solute. The resulting system contained roughly 28000 water molecules and 90000 atoms in total. The AMBER99ff was used to model the residues of the protein. The AMPH and the emitter (also called substrate in this paper) were described using parameters developed by our group<sup>9,15,23</sup>. As the parameters were not fully optimised for the excited state, the substrate structure was first obtained by QM/MM optimization of the first singlet excited state, and frozen in its excited state conformation during the whole dynamic. The system was heat from 100K to 300K in 20ps. Then, under NPT conditions with T=300K and P=1atm a 5ns dynamic using periodic boundary conditions was realised with a 2 fs time step. During these simulations pressure and temperature were maintained using the Berendsen algorithm<sup>24</sup> with a coupling constant of 5 ps and SHAKE constraints were applied to all bonds involving hydrogen atoms<sup>25</sup>.

The QM/MM calculations were performed using a QM/MM coupling scheme<sup>26</sup> between Gaussian<sup>27</sup> and Tinker<sup>28</sup> or Molcas<sup>29</sup> and Tinker (Gaussian 09d/Tinker or Molcas 8.0/Tinker). The electrostatic potential fitted (ESPF) method<sup>26</sup> was used to compute the interaction between the Mulliken charges of the QM parts and the external electrostatic potential of the MM parts within 9Å from the QM part. The microiterations technique<sup>30</sup> was used to converge the MM subsystem geometry for every QM minimization step. The emitter was selected as the QM part, while the rest of the system belonged to the MM part. The QM/MM optimisation of the first singlet excited state (S1) state was first performed followed by the calculation of the vertical difference of energies ( $T_e$ ) between the S1 and the ground state (S0) which corresponds to the fluorescence emission. Since in fireflies the fluorescence transition is the same as the bioluminescence one<sup>23</sup>, the calculated  $T_e$  value can be compared to the experimental emission energy.

The level of theory used for the QM part of the QM/MM calculations was chosen as followed. The first guess was to try some TD-DFT calculations with CAM-B3LYP functional<sup>31</sup> with the 6-31G(d,p) basis set, which has often been used for the QM system in previous QM/MM calculations<sup>2,32</sup> of such system.

The preliminary QM/MM results with the small basis set were not convincing (see Table S1). Thus as recommended in the publication<sup>33</sup>, a triple-zeta basis set was used as well as different functionals, i.e. B3LYP<sup>34-36</sup>, MO6<sup>37</sup>, MO62X<sup>37</sup> and  $\omega$ B97XD<sup>38</sup> (see Table S2). The chosen basis set was 6-311G(2d,p) as we did not use basis set with diffuse functions because these later ones can interact with the MM system while using an electrostatic embedding (ESPF method<sup>26</sup>). For all functionals and both basis sets, the optimised excited geometric conformations of the substrate are very similar to each other with a total Root-Mean-Square Deviation of atomic positions (RMSD) between the structures lower than 0.02Å.

Nature of the protein	Mutable residue (284)	Substrate	Assigned label
wild-type	Serine	Oxyluciferin	Ser_OxyLH <sub>2</sub>
wild-type	Serine	infraOxyluciferin-1	Ser_OxyiLH <sub>2</sub> -1
wild-type	Serine	infraOxyluciferin-2	Ser_OxyiLH <sub>2</sub> -2
mutated	Threonine	Oxyluciferin	Thr_OxyLH <sub>2</sub>
mutated	Threonine	infraOxyluciferin-1	Thr_OxyiLH <sub>2</sub> -1
mutated	Threonine	infraOxyluciferin-2	Thr_OxyiLH <sub>2</sub> -2

**Table 1** Description of the 6 models used. The "wild-type" name refers to the structure obtained from the 4G37 structure<sup>18</sup> and the "mutated" name refers to the same structure with the residue Serine 284 mutated to Threonine.

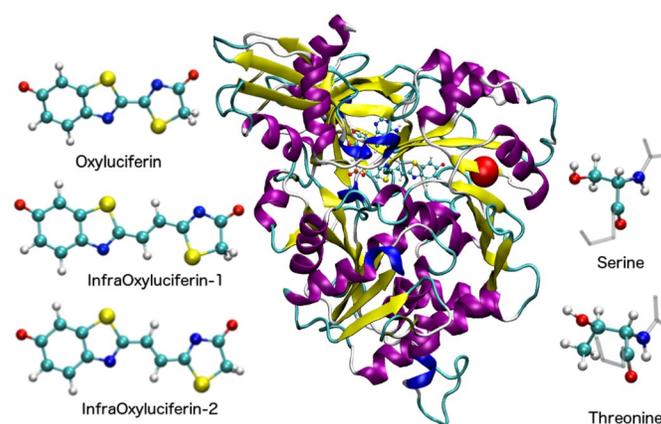
For the rest of the paper, we have chosen to compare CAM-B3LYP and B3LYP functionals with the triple zeta basis set, 6-311G(2d,p). Along with the QM/MM calculations we have also carried out some vacuum and in solvent calculations in a polar and non-polar solvent (see Table S3). The solvent effects were taken into account using Polarizable Continuum Model (PCM)<sup>39,40</sup>.

TD-DFT results were compared with more precise calculations like multiconfigurational reference second-order perturbation theory. The complete-active-space second-order perturbation theory (CASPT2)<sup>41</sup> was also used to compute the  $T_e$  with the triple-zeta basis set ANO-RCC-VTZP<sup>42</sup>. We performed single point CASPT2/MM calculations with the ANO-RCC-VTZP basis set on the optimised geometries from the CAM-B3LYP 6-31G(d,p)/MM calculations. As mentioned in the DFT paragraph, the geometric structure is the same regardless of the functional or basis set used.

For OxyLH<sub>2</sub>, 3 active spaces were tried: the 18 electrons in 15 orbitals choice (18-in-15) corresponds to all  $\pi$ -conjugated orbitals of OxyLH<sub>2</sub>. Then, a 16-in-14 and a 14-in-13 active spaces were created by removing  $\pi$  conjugated orbitals centred on sulfurs. The 16-in-14 corresponds to the remove of the  $\pi$  sulfur centred orbital on the thiazolone cycle while the 14-in-13 corresponds to the remove of both the  $\pi$  sulfur centred orbital on the thiazolone cycle and on the benzothiazole cycle. For OxyiLH<sub>2</sub>-1 and OxyiLH<sub>2</sub>-2 a 16-in-15 active space was chosen. It corresponds to all  $\pi$ -conjugated orbitals except those centred on both sulfurs due to computational limitation. The different active spaces are represented in Figure S2 to S4. The Multi-State CASPT2 (MS-CASPT2) calculations were computed either with the 4 first states or the 2 first states and a 0.1 level shift. Two different IPEA shift were tried, one at 0.25 and one at 0.1 (see Table S4 for the comparison between the 2 IPEA shift and Table S5 for the results with 4 roots).

In this paper we discuss on several models. The first model designed represents a structure with the wild-type protein asso-

ciated with the wild-type Oxyluciferin (OxyLH<sub>2</sub>) from the 4G37 crystallographic structure determined by the group of Prof. Branchini<sup>18,19</sup>. Unfortunately no crystallographic structures with OxyiLH<sub>2</sub> in protein are available. Moreover, the experimenters<sup>17</sup> have only studied the stereoisomer with the double bond in trans (E) position. We remained with the (E) stereoisomer in this study. Besides, there are not yet experimental evidences of the chemical conformation of the stereoisomer (E) of OxyiLH<sub>2</sub>. Indeed, the surrounding protein prevents the substrate from moving through sterical effects and thus from changing its conformation. Therefore each conformer has to be studied in details. One of the conformer seems at the first glance to fit better in the cavity and thus has been named infraOxyluciferin-1 (OxyiLH<sub>2</sub>-1) while the other one has been called infraOxyluciferin-2 (OxyiLH<sub>2</sub>-2) (see Figure 2).



**Fig. 2** Graphical representation of the different models used in this paper. On the left, the structures and the names of the 3 different substrates studied. Atoms are colour-coded as O red, H white, C cyan, N blue, S yellow. On the right, the structures and the names of the wild type (Serine) and mutated (Threonine) version of the residue 284. In the middle, the graphical representation of luciferase and OxyiLH<sub>2</sub>-1. The protein is drawn in ribbon with alpha helix in purple and beta sheets in yellow. The position of the residue 284 is highlighted in red.

The obtained model corresponding to the crystallographic form of the wild-type protein with the Oxyluciferin as described above is named Ser\_OxyLH<sub>2</sub> (Substrate OxyLH<sub>2</sub> inside the protein with Serine 284). For the others systems, modifications of either the emitter or the protein were made by hand. The model called Ser\_OxyiLH<sub>2</sub>-1 (respectively Ser\_OxyiLH<sub>2</sub>-2) corresponds to the wild-type protein with one of the studied conformer of OxyiLH<sub>2</sub> labelled OxyiLH<sub>2</sub>-1 emitter (respectively the other labelled OxyiLH<sub>2</sub>-2 emitter) (Figure 2). The models called Thr\_OxyLH<sub>2</sub>, Thr\_OxyiLH<sub>2</sub>-1 and Thr\_OxyiLH<sub>2</sub>-2 correspond to the mutated version of the protein, S284T (Serine mutated into Threonine) with respectively OxyLH<sub>2</sub>, OxyiLH<sub>2</sub>-1 or OxyiLH<sub>2</sub>-2. We have

**Table 2** TD-DFT/MM for electronic transition between S1 and S0 ( $T_e$ ) with 6-311G(2d,p) basis set and associated oscillator strength on structure optimised with the same level of theory.

Model	CAM-B3LYP eV (nm)	B3LYP eV (nm)	Exp. <sup>17</sup> eV (nm)
Ser_OxyLH <sub>2</sub>	2.53 (490) f=0.57	2.20 (564) f=0.33	2.22 (558)
Ser_OxyiLH <sub>2</sub> -1	2.23 (554) f=1.13	2.00 (616) f=0.53	1.85 (670)
Ser_OxyiLH <sub>2</sub> -2	2.10 (590) f=1.13	1.87 (663) f=0.53	
Thr_OxyLH <sub>2</sub>	2.43 (509) f=0.60	2.15 (577) f=0.34	2.04 (605)
Thr_OxyiLH <sub>2</sub> -1	2.24 (552) f=1.12	1.98 (625) f=0.47	1.75 (706)
Thr_OxyiLH <sub>2</sub> -2	2.16 (574) f=1.12	1.87 (661) f=0.47	
vacuum OxyLH <sub>2</sub>	2.52 (482) f=0.67	2.27 (547) f=0.41	
vacuum OxyiLH <sub>2</sub> -1	2.36 (525) f=1.14	2.13 (580) f=0.76	
vacuum OxyiLH <sub>2</sub> -2	2.26 (547) f=1.02	2.03 (608) f=0.65	

chosen to put the OH group of the Threonine in direction of the emitter. The infrared-Oxyluciferin structures were achieved by adding a double bond between the 2 cycles of the natural OxyLH<sub>2</sub>. Thus, the thiazolone cycle was frozen and the remaining part of the molecule was added. All structures are described in Table 1 and summarised in Figure 2.

## Results and discussion

For each emitters OxyLH<sub>2</sub>, OxyiLH<sub>2</sub>-1 and OxyiLH<sub>2</sub>-2,  $T_e$  was calculated in vacuum and in solvent. For each classical molecular dynamics (MD) done on models of Table 1, a snapshot corresponding to the minimum energy structure was selected from the MD, minimized at the MD level of theory and further optimised in the first singlet excited state with QM/MM methods to obtain  $T_e$  in protein.

### Vacuum and in solvent calculations

We first compare the optimised S1 structures and the calculated  $T_e$  values at the TD-DFT B3LYP/6-311G(2d,p) level for the 3 substrates obtained in either vacuum, with a polar solvent (water), or a non-polar solvent (chloroform). The optimised structures with the different PCM solvents are close to the one in vacuum with a RMSD lower than 0.02Å. The augmentation of the polarity around the substrate leads to a smaller  $T_e$ . In details, the val-

**Table 3** MS-CASPT2/MM for electronic transition between S1 and S0 ( $T_e$ ) with ANO-RCC-VTZP basis set, including 2 states and using a 0.1 level shift and 0.1 IPEA shift.

Model	Active Space	MS-CASPT2 eV (nm)	Exp. <sup>17</sup> eV (nm)
Ser_OxyLH <sub>2</sub>	14-in-13 16-in-14	2.24 (554) 2.14 (579)	2.22 (558)
Ser_OxyiLH <sub>2</sub> -1	16-in-15	2.19 (566)	1.85 (670)
Ser_OxyiLH <sub>2</sub> -2	16-in-15	1.92 (645)	
Thr_OxyLH <sub>2</sub>	14-in-13 16-in-14	2.07 (598) 2.15 (577)	2.04 (605)
Thr_OxyiLH <sub>2</sub> -1	16-in-15	2.15 (578)	1.75 (706)
Thr_OxyiLH <sub>2</sub> -2	16-in-15	1.93 (643)	
vacuum OxyLH <sub>2</sub>	14-in-13 16-in-14	2.33 (532) 2.25 (550)	
vacuum OxyiLH <sub>2</sub> -1	16-in-15	2.02 (614)	
vacuum OxyiLH <sub>2</sub> -2	16-in-15	1.93 (642)	

ues obtained with the chloroform are closer to the ones obtained taking into account the protein surrounding, with the QM/MM method (Table S3) than in the water or in vacuum.

### QM/MM

The utilisation of QM/MM calculations on the structures obtained after MD gives us good insights on the bioluminescence emission colour modulation. For each model, one snapshot of the MD has been optimised by a QM/MM scheme with the substrate in the first singlet excited state in order to get the  $T_e$ . In all models the studied transition corresponds mainly to a LUMO-HOMO transition. The experimental trends are explained by our theoretical results. For the DFT/MM calculations the difference of energy between the experimental and the calculated emission energies ( $\Delta T_e$ ) is around 0.35 eV for CAM-B3LYP while for B3LYP this shift ranges between 0.02 and 0.25 eV which is in agreement with the DFT level of error (see Table 2).

MS-CASPT2/MM calculations were also computed and the obtained results are similar with the one obtained with the B3LYP functional (see Table 3). Moreover, the  $\Delta T_e$  is between 0.05 and 0.4 eV. We obtained these results by fixing an empirical parameter called IPEA in the MS-CASPT2 program to 0.1 (default one is 0.25). Current research<sup>43</sup> shows that this parameter has a default value set too high. Indeed, the IPEA shift set at 0.25 instead of 0.1 induces a blue-shift of at least 0.15 eV for the  $T_e$  (see Table S4). Thus the  $\Delta T_e$  for the IPEA shift set at 0.25 is ranged between 0.1 and 0.6 eV (see Table S5). Others parameters can also influence the  $T_e$  such as the size of the active space (see Table S5) and the basis set.

When the  $\pi$  conjugated system in the emitter is increased i.e. from OxyLH<sub>2</sub> to OxyiLH<sub>2</sub> the calculated emission of energy  $T_e$  decreases leading to red-shifting of the emission, as seen in the experimental results. Besides, for both emitters the emission is accompanied with a charge transfer from the Highest Occupied Molecular Orbital (HOMO) to the Lowest Unoccupied Molecular Orbital (LUMO) corresponding to a charge transfer from the benzothiazole moiety to the thiazolone cycle (see Figure 4). Furthermore, the results show that the conformation of the double bond in OxyiLH<sub>2</sub> has an impact on the  $T_e$ . We observed that the OxyiLH<sub>2</sub>-2 structure's  $T_e$  is closer to the experimental emission than the one from OxyiLH<sub>2</sub>-1 structure. For example, the Ser\_OxyiLH<sub>2</sub>-1 has a B3LYP  $T_e$  of 1.98 eV while the Ser\_OxyiLH<sub>2</sub>-2 has a  $T_e$  of 1.87 eV, to be compared to the experimental value of 1.85 eV.

In conclusion, results obtained in TD-DFT/MM with B3LYP/6-311G(2d,p) are more reliable and faster. Thus, in the rest of the paper both structures and properties will be discussed with this level of theory.

### Classical Molecular Dynamic (MD)

During the MD, we were able to follow several elements involved in the interaction between the substrate and the environment (protein residues and water molecules). First, the equilibration is attained after 1ns for all MD (see Figure S6). During the MD, the coordinates of the substrate, i.e. OxyLH<sub>2</sub> or OxyiLH<sub>2</sub>, were restrained. Thus the structure of the substrate is kept at its initial S1 conformation. However, the AMPH was not restrained but thanks to strong hydrogen bonds and steric effects the molecule does not move a lot (RMSD < 0.5Å) on all MD. The snapshot used for the QM/MM calculations corresponds to a residue 284 turned toward the cavity. A longer MD simulation shows that the dihedral angle N-CA-CB-OG is stable to this direction during the 5 first ns and returns to the same value after 8.5ns (see Figure S7). In the cavity, an H-bond network interacts with the emitter and especially with the phenolate oxygen (O<sub>1</sub>) (see Figure 3, Figure 5 and Figure S5). The structures obtained after MD show that a H-bond network is created between the emitter and several residues especially the 284 which corresponds to the residue that can be mutated from Serine to Threonine. This H-bond network cannot be seen in the crystallographic structure as the water molecules are labile. A water molecules network has already been observed in our previous publications<sup>9</sup>.

### Influence of the H-bond network on the $T_e$

How does the H-bond network obtained with MD simulations can explain the experimental emission?

The effect of the creation of this H-bond network during the dynamic is characterised by a blue-shift of the emission for all

levels of theory computed (see Table S6). In comparison, the PCM calculations in water give a red-shift emission compared to the vacuum ones (see Table S3). However, in the PCM model the water solvent is model by a dielectric constant all around the substrate while in the QM/MM models the water molecules are localised, leading to non-homogenous polarity of the cavity.

In the QM/MM models, the H-bond network involved the phenolate group of the oxyluciferin. As the transition S1 to S0 involved a charge transfer from the thiazolone moiety to the benzothiazolate moiety (see HOMO and LUMO on Figure 4), the S0 state is stabilised by the creation of the H-bond network, compared to the S1 state. Therefore the transition energy increases and the light emission is shifted to the blue side of the spectra. The blue-shift induced by the presence of a water molecule near the phenolate group has been already observed experimentally<sup>44</sup>. Finally, the presence of the H-bond network is required to get the closest emission values comparable to experiments.

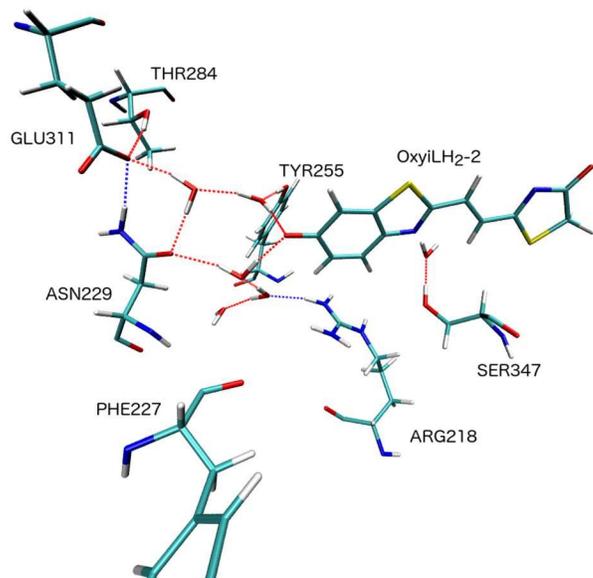
### Mutation of the protein

The effect of the mutation of the S284T has a different impact according to the emitter chosen. Indeed for the OxyLH<sub>2</sub> we observed a slight red-shift when the residue is mutated but not as consistent as the observed experimental red shift. When we look at the H-bond network created we observe that in the case of the Ser\_OxyLH<sub>2</sub> model there are two water molecules connected to the O<sub>1</sub> oxygen of OxyLH<sub>2</sub> while for Thr\_OxyLH<sub>2</sub> only one water molecule is present. The reduction of interaction with O<sub>1</sub>, due to the mutation leads to the destabilisation of the HOMO, resulting in a red shift.

For the OxyiLH<sub>2</sub>, the experimental results show also a red shift with the mutation. However, QM/MM energies  $T_e$  show that the mutation does not change significantly the emission. For the OxyiLH<sub>2</sub>-2 emitter, one water molecule interacts with the emitter in the Ser model while two of them are present in the Thr model. For the Thr\_OxyiLH<sub>2</sub>-2 model the  $T_e$  is slightly blue shifted compared to Ser\_OxyiLH<sub>2</sub>-2, which is the opposite of the experimental observation. Furthermore, to get more insight on the H-bond network we have mutated the residue 284 from Serine to Alanine. Unlike the previous mutation, the mutation S284A induces a modification of the surrounding. The alanine is a non-polar residue, thus it can not be involved in a H-bond network (see Figure S8). The TD-DFT/MM B3LYP  $T_e$  (1.89 eV, 655 nm) obtained for Ala\_OxyiLH<sub>2</sub>-2 (ie. OxyiLH<sub>2</sub>-2 with S284A luciferase) is similar to both  $T_e$  obtained in Ser\_OxyiLH<sub>2</sub>-2 and Thr\_OxyiLH<sub>2</sub>-2.

The present calculations were therefore not able to explain the red-shift due to the S284T mutation of the protein. A possible explanation could be that only one snapshot was taken from the dynamic. The use of many snapshots<sup>45 46 47</sup> may have given more

precise information but are time-consuming when dealing with emission.



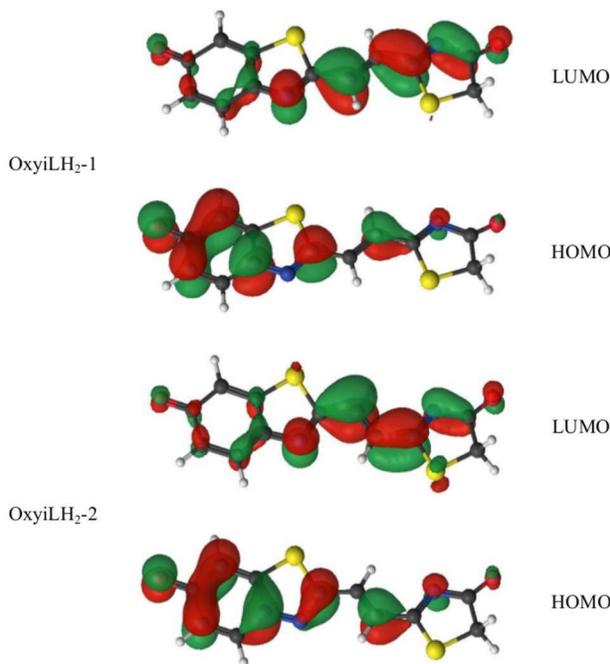
**Fig. 3** Representation of the H-bond network of the model Thr\_OxyiLH<sub>2</sub>-2 after classical molecular dynamics (MD). The residues or part of the residues represented are involved in the H-bond network. Atoms are colour-coded as O red, H white, C cyan, N blue, S yellow, and P tan.

### Decomposition of the surrounding contributions

In fireflies the bioluminescence reaction takes place when the oxyluciferin is located in the cavity of the protein. By doing so, the substrate adopts a certain conformation and creates some intermolecular interactions with the luciferase and water molecules inside the protein.

Two main contributions of the surrounding on the transition energy have to be taken into account; the electrostatic effect and the geometric constraint on the conformation of the molecule.

We first looked at the geometry constraint on the substrate. For the OxyLH<sub>2</sub> substrate this torsion is about 3° and is located between the two cycles<sup>9</sup>. Then we collect some geometry parameters of OxyiLH<sub>2</sub> substrates, optimised in vacuum or inside the mutated protein (see Table 4). There are no major differences on the C-O bonds lengths and therefore these factors are not contributing. In vacuum the optimised structure of the emitter is completely planar while the one optimised in the protein is twisted. In order to look at the torsion between the two moieties of the OxyiLH<sub>2</sub> molecule, we have reported the torsion angle between the benzothiazole moiety and the middle C=C double bond ( $\beta_{1-12}$ ) and the one between the C=C and the thiazolone moiety ( $\beta_{11-8}$ ) (see Figure 5). The twisting in OxyiLH<sub>2</sub>-1 is com-



**Fig. 4** Frontier's orbitals of OxyiLH<sub>2</sub>-1 (top) and OxyiLH<sub>2</sub>-2 (bot) obtained from the CASSCF calculation of the emitter in the protein (LUSCUS program<sup>48</sup>)

parable with the one of OxyLH<sub>2</sub> for both  $\beta$  while for OxyiLH<sub>2</sub>-2 the twisting is located around  $\beta_{1-12}$  and has a higher intensity of 9° (see Table 4). Furthermore, Figure 6 shows the superposition of the optimised OxyiLH<sub>2</sub> structures. When OxyiLH<sub>2</sub>-1 shows a distortion of the benzothiazole plane in the wild-type protein, the OxyiLH<sub>2</sub>-2 shows a rotation around C1-C12 bond. Therefore, the torsion between the 2 cycles called  $\beta_{2cycles}$  in Table 4 is very small for OxyiLH<sub>2</sub>-1 while for OxyiLH<sub>2</sub>-2 the value is around 20°. The torsion for OxyLH<sub>2</sub> is not as visible (Figure S9).

The constraint due to the environment lowered the conjugation pathway in the molecule compared to a planar structure but does not have a huge impact on the emission. Indeed, single point calculation of the  $T_e$  in vacuum ( $T_e^{vac}$ ) on both the in-vacuum-optimised structure and the in-protein optimised structure shows almost no difference, for OxyiLH<sub>2</sub>-1 it is +0.03 eV (8 nm) while for OxyiLH<sub>2</sub>-2 it is -0.05 eV (15 nm) (Table 4). Geometric effects are still very small. The difference between the 2 conformers can be explained as the following: in OxyiLH<sub>2</sub>-1, the two torsions are relatively small compared to OxyiLH<sub>2</sub>-2 and do not interact a lot with the HOMO and LUMO of OxyiLH<sub>2</sub>-1. In OxyiLH<sub>2</sub>-2, the biggest torsion is around C1-C12 at the edge of the electronic density of the LUMO which destabilizes this one.

The electrostatic effect can be quantified by comparing two calculations: the QM/MM calculated  $T_e$  on the optimised structure

**Table 4** Geometric parameters of OxyiLH<sub>2</sub> structures optimised inside the protein or in vacuum (optimisation with B3LYP functional and 6-311G(2d,p) basis set). Single point in vacuum energy  $T_e$  on these structures (B3LYP functional and 6-311G(2d,p) basis set).

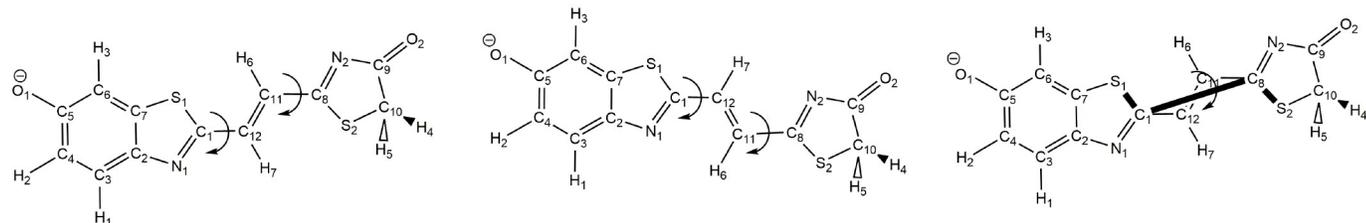
	OxyiLH <sub>2</sub> -1 optimised in wild protein	OxyiLH <sub>2</sub> -1 optimised in mutated protein	OxyiLH <sub>2</sub> -1 optimised in vacuum	OxyiLH <sub>2</sub> -2 optimised in wild protein	OxyiLH <sub>2</sub> -2 optimised in mutated protein	OxyiLH <sub>2</sub> -2 optimised in vacuum
$C_5-O_1^a$	1.25	1.26	1.25	1.25	1.26	1.25
$C_9-O_2^a$	1.24	1.24	1.22	1.22	1.23	1.22
$\beta_{1-12}^b$	1.52 <sup>c</sup>	2.84 <sup>c</sup>	0 <sup>c</sup>	-8.98 <sup>d</sup>	-8.62 <sup>d</sup>	0 <sup>d</sup>
$\beta_{11-8}^b$	-4.85 <sup>e</sup>	-3.13 <sup>e</sup>	0 <sup>e</sup>	-0.77 <sup>f</sup>	-0.33 <sup>f</sup>	0 <sup>f</sup>
$\beta_{2cycles}^{b,g}$	0.53	2.14	0	20.68	18.17	0
$T_e^{vac,h}$	2.16	2.13	2.13	2.00	1.98	2.03

<sup>a</sup> Bond length in Å; <sup>b</sup> Angle in degree

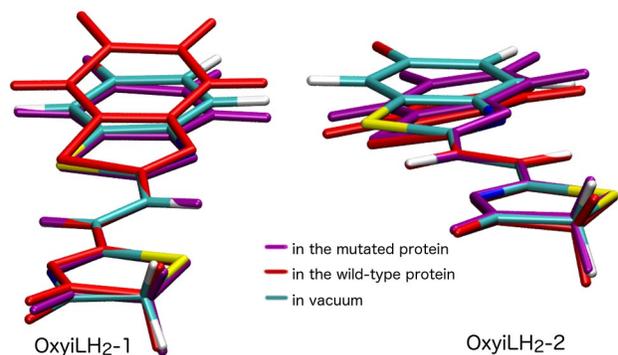
<sup>c</sup> Dihedral angle between S1-C1-C12-C11 <sup>d</sup> Dihedral angle between N1-C1-C12-C11

<sup>e</sup> Dihedral angle between C12-C11-C8-S2 <sup>f</sup> Dihedral angle between C12-C11-C8-N2

<sup>g</sup> Dihedral angle between S1-C1-C8-S2 <sup>h</sup> Single point in vacuum  $T_e$  in eV (to be compared with the value of 1.98 eV for the QM/MM transition for OxyiLH<sub>2</sub>-1 and 1.87 eV for OxyiLH<sub>2</sub>-2)



**Fig. 5** Structure of OxyiLH<sub>2</sub>-1 (left) with definition of  $\beta_{1-12}$  and  $\beta_{11-8}$ ; Structure of OxyiLH<sub>2</sub>-2 (middle) with definition of  $\beta_{1-12}$  and  $\beta_{11-8}$ ; Structure of OxyiLH<sub>2</sub>-1 (right) with definition of  $\beta_{2cycles}$ .



**Fig. 6** Comparison of the structures between OxyiLH<sub>2</sub>-1 (left) and OxyiLH<sub>2</sub>-2 (right). Molecules are colour-coded as: in vacuum cyan; in wild-type protein red; in mutated protein purple.

taking into account the protein environment and the single-point QM calculated  $T_e^{vac}$  on the same structure, without further optimisation. This has been done at the B3LYP/6-311G(2d,p) level with OxyiLH<sub>2</sub>-1 and OxyiLH<sub>2</sub>-2 derived from the corresponding models where the residue 284 is mutated.

The results show that the presence of the protein environment

reduces the gap between the ground and the first excited state of the emitter. For example, the electrostatic effect in Thr\_OxyiLH<sub>2</sub>-2 is obtained by calculating the difference between the  $T_e^{vac}$  (1.98 eV, in Table 4) and the  $T_e$  in protein (1.87 eV, in Table 2). Thus the electrostatic contribution corresponds to a lowering of 0.11 eV (an increase of 37 nm). For Thr\_OxyiLH<sub>2</sub>-1 this contribution is 2.13 - 1.98 = 0.15 eV (corresponding to an increase by 44 nm of the wavelength).

In summary, the surrounding effect can be decomposed into a geometrical constraint that leads to a very small red shift (lower than 0.05 eV) and a bigger electrostatic effect that leads to a total red shift of around 0.15 eV. The electronic effect is more important than the steric constraint.

### Conformation of the emitter

We have studied two E conformers of the OxyiLH<sub>2</sub>. The emission wavelength of the conformer OxyiLH<sub>2</sub>-2 is closer to the experimental one than the one of the OxyiLH<sub>2</sub>-1 conformer. Already in vacuum emission of OxyiLH<sub>2</sub>-2 is 0.1 eV lower than the one of OxyiLH<sub>2</sub>-1. When taking into account the environment, the emission energy is lowered further on by 0.15 eV. Thus, the difference of emission energy between these two conformers is coming from

the nature of the conformer itself. Between the two studied conformers, the OxyiLH<sub>2</sub>-2 seems to be the most likely emitter.

Z conformers were not studied in this paper as the geometrical constraints induce a bigger deformation of the cavity of the protein. Some calculations were performed on these Z conformers and results are collected in SI (see Figure S10 and Table S7).

## Conclusion

In this paper we give theoretical insights on the red-emission of a modified firefly emitter. MD simulations following by QM/MM calculations give fluorescence emission close to the experimental bioluminescence emission, for both the wild-type and the modified emitters. We also investigate the influence of the surrounding protein on the emission, thus two factors are mainly responsible for the colour-tuning. Indeed the presence of a H-bond network in the cavity of the emitter leads to a blue-shift of the emission. Besides, the mutation of specific residues linked to this network can also influence this transition wavelength. Secondly, the presence of the protein itself induces a red-shift. Therefore both these effects are mandatory to get close to the experimental situation.

A complete study has also been realised on the modified oxyluciferin substrate iLH<sub>2</sub>. We analyse 2 conformers and while both give a red-shifted emission compared to the wild-type substrate, one of them gives closer results compared to experimental values. In the end, without experimental evidences or crystallographic structure of the real emitter, our calculations show that one conformer is more likely to be the probable emitter.

The present study gives insight to new questions about the modified light emitter iLH<sub>2</sub>: which is the conformation of the double bond between the 2 cycles in the light emitter? What are the constraints inside the protein that influence the most likely conformer? These raised questions and their answers give new leads to experimental studies. For example, the mutation of polar (Glu 311, Asn 229) and non-polar (Ile 351) residues present in the cavity could be explored both experimentally and theoretically in order to a better understanding of the H-bonding network. The presence of several possible emitters makes iLH<sub>2</sub> an interesting substrate for both experimental and theoretical point of view in order to understand the colour-tuning of the bioluminescence.

## Acknowledgements

The author thanks Prof. Ferré N. from University of Marseille, France, group of Prof. Barone V. from Scuola Normale Superiore, Pisa, Italia, Prof. Lindh R. from Uppsala University, Sweden, Dr. Roca-Sanjuán D. from University of Valencia, Spain and group of Prof. Dr. Beckert R. from Friedrich-Schiller-University Jena, Germany for their advices and commentary.

## References

- 1 C. C. Woodrooffe, P. L. Meisenheimer, D. H. Klaubert, Y. Kovic, J. C. Rosenberg, C. E. Behney, T. L. Southworth and B. R. Branchini, *Biochemistry*, 2012, **51**, 9807–9813.
- 2 Y.-Y. Cheng, J. Zhu and Y.-J. Liu, *Chemical Physics Letters*, 2014, **591**, 156 – 160.
- 3 C. C. Woodrooffe, J. W. Shultz, M. G. Wood, J. Osterman, J. J. Cali, W. J. Daily, P. L. Meisenheimer and D. H. Klaubert, *Biochemistry*, 2008, **47**, 10383–10393.
- 4 X. Q. Ran, X. Zhou and J. D. Goddard, *ChemPhysChem*, 2015, **16**, 396–402.
- 5 L. P. da Silva and J. C. E. da Silva, *Journal of Chemical Theory and Computation*, 2011, **7**, 809–817.
- 6 R. C. Steinhart, J. M. O'Neill, C. M. Rathbun, D. C. McCutcheon, M. A. Paley and J. A. Prescher, *Chemistry - A European Journal*, 2016, n/a–n/a.
- 7 C. Miura, M. Kiyama, S. Iwano, K. Ito, R. Obata, T. Hirano, S. Maki and H. Niwa, *Tetrahedron*, 2013, **69**, 9726 – 9734.
- 8 M. I. Koksharov and N. N. Ugarova, *Photochem. Photobiol. Sci.*, 2013, **12**, 2016–2027.
- 9 I. Navizet, Y.-J. Liu, N. Ferré, H.-Y. Xiao, W.-H. Fang and R. Lindh, *Journal of the American Chemical Society*, 2010, **132**, 706–712.
- 10 A. Ghose, M. Rebarz, O. V. Maltsev, L. Hintermann, C. Ruckebusch, E. Fron, J. Hofkens, Y. Mély, P. Naumov, M. Sliwa and P. Didier, *The Journal of Physical Chemistry B*, 2015, **119**, 2638–2649.
- 11 T. Hirano, Y. Hasumi, K. Ohtsuka, S. Maki, H. Niwa, M. Yamaji and D. Hashizume, *Journal of the American Chemical Society*, 2009, **131**, 2385–2396.
- 12 L. P. da Silva and J. C. E. da Silva, *Journal of Chemical Theory and Computation*, 2011, **7**, 809–817.
- 13 S. Hosseinkhani, *Cellular and Molecular Life Sciences*, 2010, **68**, 1167–1182.
- 14 Y.-Y. Cheng and Y.-J. Liu, *Journal of Chemical Theory and Computation*, 2015, **11**, 5360–5370.
- 15 S.-F. Chen, Y.-J. Liu, I. Navizet, N. Ferré, W.-H. Fang and R. Lindh, *Journal of Chemical Theory and Computation*, 2011, **7**, 798–803.
- 16 I. Navizet, Y.-J. Liu, N. Ferré, D. Roca-Sanjuán and R. Lindh, *ChemPhysChem*, 2011, **12**, 3064–3076.
- 17 A. P. Jathoul, H. Grounds, J. C. Anderson and M. A. Pule, *Angewandte Chemie International Edition*, 2014, **53**, 13059–13063.
- 18 J. A. Sundlov, D. M. Fontaine, T. L. Southworth, B. R. Branchini and A. M. Gulick, *Biochemistry*, 2012, **51**, 6493–6495.
- 19 B. R. Branchini, J. C. Rosenberg, D. M. Fontaine, T. L. Southworth, C. E. Behney and L. Uzasci, *Journal of the American*

- Chemical Society*, 2011, **133**, 11088–11091.
- 20 K. Tang, J. Zhang and J. Liang, *PLoS Comput Biol*, 2014, **10**, e1003539.
- 21 D. Case and al, 2015.
- 22 R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Research*, 2012, **40**, W537–W541.
- 23 I. Navizet, D. Roca-Sanjuán, L. Yue, Y.-J. Liu, N. Ferré and R. Lindh, *Photochemistry and Photobiology*, 2013, **89**, 319–325.
- 24 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684–3690.
- 25 J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327 – 341.
- 26 N. Ferré and J. G. Ángyán, *Chemical Physics Letters*, 2002, **356**, 331 – 339.
- 27 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox.
- 28 J. W. Ponder, 2005.
- 29 F. Aquilante, L. De Vico, N. Ferré, G. Ghigo, P.-A. Malmqvist, P. Neogrády, T. B. Pedersen, M. Pitoňák, M. Reiher, B. O. Roos, L. Serrano-Andrés, M. Urban, V. Veryazov and R. Lindh, *Journal of Computational Chemistry*, 2010, **31**, 224–247.
- 30 F. Melaccio, M. Olivucci, R. Lindh and N. Ferré, *International Journal of Quantum Chemistry*, 2011, **111**, 3339–3346.
- 31 T. Yanai, D. P. Tew and N. C. Handy, *Chemical Physics Letters*, 2004, **393**, 51 – 57.
- 32 S. Chen, I. Navizet, R. Lindh, Y. Liu and N. Ferré, *The Journal of Physical Chemistry B*, 2014, **118**, 2896–2903.
- 33 A. D. Laurent, A. Blondel and D. Jacquemin, *Theoretical Chemistry Accounts*, 2015, **134**, 1–11.
- 34 A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648–5652.
- 35 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *The Journal of Physical Chemistry*, 1994, **98**, 11623–11627.
- 36 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
- 37 Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2008, **120**, 215–241.
- 38 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- 39 B. Mennucci, E. Cances, and J. Tomasi, *The Journal of Physical Chemistry B*, 1997, **101**, 10506–10517.
- 40 E. Cances, B. Mennucci and J. Tomasi, *The Journal of Chemical Physics*, 1997, **107**, 3032–3041.
- 41 K. Andersson, P. Malmqvist and B. O. Roos, *The Journal of Chemical Physics*, 1992, **96**, 1218–1226.
- 42 B. O. Roos, R. Lindh, P.-A. Malmqvist, V. Veryazov, and P.-O. Widmark, *The Journal of Physical Chemistry A*, 2004, **108**, 2851–2858.
- 43 J. P. Zobel, J. J. Nogueira and L. González, *in preparation*, 2016.
- 44 K. Stöckel, 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 and B. F. Milne, *Journal of the American Chemical Society*, 2013, **135**, 6485–6493.
- 45 N. A. Murugan, J. Kongsted, Z. Rinkevicius and H. Ågren, *Phys. Chem. Chem. Phys.*, 2012, **14**, 1107–1112.
- 46 K. Aidas, J. M. H. Olsen, J. Kongsted and H. Ågren, *The Journal of Physical Chemistry B*, 2013, **117**, 2069–2080.
- 47 N. A. Murugan, J. M. H. Olsen, J. Kongsted, Z. Rinkevicius, K. Aidas and H. Ågren, *The Journal of Physical Chemistry Letters*, 2013, **4**, 70–77.
- 48 G. Kovačević and V. Veryazov, *Journal of Cheminformatics*, 2015, **7**, 1–10.

For Table of Contents use only

Authors: Romain Berraud-Pache, Isabelle Navizet\*

QM/MM calculations and MD give insights on the light emission of the firefly oxyluciferin and on a modified red analogue.

