



**HAL**  
open science

## Etude QM/MM de systèmes bioluminescents

Romain Berraud-Pache

► **To cite this version:**

Romain Berraud-Pache. Etude QM/MM de systèmes bioluminescents. Théorie spectrale [math.SP]. Université Paris-Est, 2017. Français. NNT : 2017PESC1033 . tel-01679340

**HAL Id: tel-01679340**

**<https://theses.hal.science/tel-01679340>**

Submitted on 9 Jan 2018

**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.



UNIVERSITÉ —  
— PARIS-EST



Thèse présentée pour obtenir le grade de

**Docteur de l'Université Paris-Est**

Spécialité : Chimie

par

**Romain Berraud-Pache**

Ecole Doctorale : SCIENCES, INGÉNIERIE ET ENVIRONNEMENT

***Etude QM/MM de systèmes bioluminescents***

Thèse soutenue le 6 octobre 2017 devant le jury composé de :

Isabelle Demachy	Université Paris-Sud	<i>Rapporteur</i>
Etienne Derat	Université Pierre et Marie Curie	<i>Rapporteur</i>
Jérémie Leonard	Université de Strasbourg	<i>Examineur</i>
Elise Dumont	Ecole Normale Supérieure de Lyon	<i>Examineur</i>
Marco Garavelli	Università degli Studi di Bologna	<i>Président du jury</i>
Isabelle Navizet	Université Paris-Est Marne-la-Vallée	<i>Directeur de thèse</i>



*Tell me and I forget. Teach me and I remember. Involve me and I learn.*

Benjamin Franklin



# Remerciements

Ce manuscrit correspond à la thèse que j'ai menée (avec plaisir) au sein de l'équipe de Chimie Théorique (CT) du Laboratoire de Modélisation Simulation Multi-Echelle (MSME) de l'Université Paris-Est Marne-la-Vallée.

Dans un premier temps je voudrais remercier les membres de mon jury de thèse. Je remercie le Prof. Isabelle Demachy et le Dr. Etienne Derat pour avoir lu et corrigé ma thèse ainsi que pour leurs retours lors de la soutenance de thèse. Je remercie également les trois examinateurs, le Pr. Elise Dumont, le Dr. Jérémy Léonard et le Prof. Marco Garavelli. J'ai beaucoup apprécié les questions et les discussions lors de ma soutenance de thèse.

Je remercie ensuite les personnes qui m'ont permis de faire ma thèse dans ce laboratoire : le Pr. Salah Naili, directeur du laboratoire MSME et les membres de l'école doctorale SIE, notamment Mme Cécile Blanchemanche. Merci également pour tous les financements qui m'ont été accordés pour aller en conférence.

Je remercie le Prof. Isabelle Navizet, ma directrice de thèse pour m'avoir fait confiance dès le départ, avant même le début de thèse. Merci également pour ces très bons moments, discussions et échanges. Tu m'as toujours laissé explorer et proposer de nouveaux projets et idées, et cette indépendance est vraiment ce que j'ai le plus aimé lors de la thèse. Merci encore pour toujours avoir pris le temps de m'écouter quand j'avais des problèmes ou interrogations. Ce fut un vrai plaisir de travailler avec toi et j'espère que cela continuera.

Lors de cette thèse j'ai été entouré de plein de personnes extraordinaires avec qui j'ai pu échanger et partager plein de très bon moments :

La nouvelle Dr. Pricilia Kutudila, avec qui j'ai partagé pendant 3 ans cette incroyable (mais fatigante) aventure. Merci pour les supers souvenirs, ton soutien, ton implication et ta joie omniprésente.

## REMERCIEMENTS

Le Dr. Eddy Bernard, pour nos discussions (plus ou moins en rapport avec le boulot) et j'attends toujours ton livre de maths. Merci également pour le prêt de ton portable de thèse avec lequel j'écris ces remerciements (et RIP à mon mac parti en plein milieu de mon écriture de thèse).

Le Dr. Cristina Garcia Iriepa pour la bonne ambiance dans le bureau et pour m'avoir soutenu dans l'écriture de la thèse.

Je remercie également tous les actuels et anciens doctorants, post-doctorants et stagiaires du laboratoire : Ayad Bellili, Romain Grenier, Séphora Mondésir, Mohamed Cheraki, Miguel Ponce Vargas, Marco Marazzi, Madjid Zemmouche (courage pour la thèse), Zied Gouid, Rahma Dahmani, Nihab, Dorra Khiri, Mohamed Achref Gannouni, Tarek Trabelsi, Reda Boulmene, Karim Boussouf, Tarek Ayari, Onsi Sghier, Massoud Sharoki, Prakash Muthuramalingam, Riadh Hanachi...

Merci à tous les membres de l'équipe Chimie Théorique pour leur écoute, leur aide et conseils tout le long de la thèse. Merci à Gilberte Chambaud, Majdi Hochlaf, Frédéric Le Quéré, Marie Guitou, Céline Léonard, Halima Mouhib, Alexander (Sacha) Mistrushenkov, Roberto Linguerri, Marius Lewerenz et Mirjana Madlenovic.

Merci à Guillaume Sauvenay pour avoir résolu nombre de mes problèmes informatiques ainsi qu'aux membres du CRI. Je remercie également le personnel enseignant de la licence de chimie qui m'ont beaucoup aidé lors de mes enseignements, notamment Béatrice Bassinet et Jean-Michel Riom. Merci finalement au personnel administratif du laboratoire, notamment Stéphanie Domise.

Je voudrais remercier le Prof. Dr. Beckert et son équipe de l'institut für organische chemie und makromolekulare chemie de Iena en Allemagne. Merci également au Dr. Stefan Schramm pour les moments passés à Iena.

J'aimerais aussi remercier des personnes qui le long de ma scolarité m'ont donné l'envie de travailler dans le domaine de la chimie et devenir chercheur. Merci au Dr. Bernard Boitrel et au Dr. Stéphane Le Gac de l'insitut des sciences chimiques de Rennes pour m'avoir accueilli pour mon premier stage de recherche et pour m'avoir donné envie de continuer. Merci également au Dr. Boris Le Guennic, de l'insitut des sciences chimiques de Rennes pour m'avoir fait confiance pour le stage de Master 2 et au Dr. Julie Jung pour son aide durant le stage.

Un remerciement spécial au Dr. Fanny Vazart pour m'avoir "encourager" à suivre les cours de chimie théorique en Master 2 et ainsi m'engager dans cette voie.

Je remercie tous mes amis qui m'ont soutenu et encouragé lors de ces trois ans. Je remercie aussi toutes les personnes qui m'ont suggéré des idées concernant ma recherche sur les lucioles (la peinture et tatouages bioluminescents).

## REMERCIEMENTS

Enfin je remercie ma famille et mes parents, qui m'ont toujours soutenu et encouragé à poursuivre dans ce qui me tiens à cœur.



# Table des matières

Remerciements

## Table des matières

<b>Introduction</b>	<b>1</b>
<b>I Méthodes</b>	<b>5</b>
<b>1 Méthodes de type ab initio</b>	<b>7</b>
1.1 Méthode Hartree-Fock . . . . .	8
1.1.1 Equation de Roothan . . . . .	9
1.1.2 La procédure SCF . . . . .	11
1.1.3 L'énergie de corrélation . . . . .	11
1.2 Méthodes Post Hartree-Fock . . . . .	11
1.2.1 La méthode d'interaction de configuration (IC) . . . . .	12
1.2.2 La méthode CASSCF (Complete Active Space Self Consistent Field) . . . . .	12
1.2.3 La méthode CASPT2 . . . . .	14
1.2.4 La méthode MS-CASPT2 . . . . .	15
1.3 Méthodes basées sur la densité électronique . . . . .	16
1.3.1 La méthode DFT (Density Functional Theory) . . . . .	16
1.3.2 Les fonctionnelles d'échange et de corrélation . . . . .	17
1.3.3 La méthode TD-DFT (Time-Dependent DFT) . . . . .	19
1.4 Les bases atomiques d'orbitales . . . . .	20
1.4.1 Les bases de type split-valence . . . . .	21
1.4.2 Les bases de type ANO . . . . .	22
<b>2 Méthodes basées sur la mécanique moléculaire</b>	<b>25</b>
2.1 Les champs de force . . . . .	25
2.1.1 La fonction d'énergie potentielle . . . . .	26
2.1.2 Les termes liés . . . . .	27
2.1.3 Les termes non-liés . . . . .	28
2.1.4 Des exemples de champ de force . . . . .	28
2.1.4.1 Champ de force classique . . . . .	28

## TABLE DES MATIÈRES

2.1.4.2	Champ de force polarisable . . . . .	29
2.1.5	Le traitement de l'eau . . . . .	29
2.2	Dynamique moléculaire . . . . .	30
2.2.1	Le principe . . . . .	30
2.2.2	Les paramètres de la simulation . . . . .	31
2.2.2.1	Les ensembles thermodynamiques . . . . .	31
2.2.2.2	Les conditions périodiques . . . . .	32
<b>3</b>	<b>Méthode hybride QM/MM</b>	<b>33</b>
3.1	Le principe . . . . .	34
3.2	Les schémas QM/MM . . . . .	34
3.2.1	Le schéma soustractif . . . . .	34
3.2.2	Le schéma additif . . . . .	34
3.2.3	Les interactions QM/MM . . . . .	35
3.2.3.1	Le mechanical embedding (ME) . . . . .	35
3.2.3.2	L'electrostatical embedding (EE) . . . . .	35
3.2.3.3	Le polarisable embedding (PE) . . . . .	36
3.2.4	Définir la partie QM/MM . . . . .	36
3.2.4.1	Le link-atom (LA) . . . . .	37
3.2.4.2	Les orbitales localisées . . . . .	37
<b>4</b>	<b>Modéliser l'environnement</b>	<b>39</b>
4.1	Le modèle implicite . . . . .	39
4.2	Le modèle explicite . . . . .	40
4.3	Calculer l'influence de l'environnement . . . . .	41
<b>5</b>	<b>Modéliser des chemins réactionnels</b>	<b>43</b>
5.1	Un mécanisme réactionnel simple . . . . .	43
5.2	La réactivité dans les états excités . . . . .	44
<b>II</b>	<b>La luminescence</b>	<b>47</b>
<b>6</b>	<b>Les différentes sortes de luminescence</b>	<b>49</b>
6.1	La photoluminescence . . . . .	50
6.2	La chimiluminescence . . . . .	53
6.2.1	Exemples de chimiluminescence . . . . .	53
6.2.2	Les 2-coumaranones . . . . .	54
6.3	La bioluminescence . . . . .	54
6.3.1	Les espèces bioluminescentes . . . . .	55
6.3.2	Les applications dans le domaine scientifique . . . . .	55
<b>III</b>	<b>Les protéines</b>	<b>59</b>
<b>7</b>	<b>Généralités sur les protéines</b>	<b>63</b>

## TABLE DES MATIÈRES

7.1	La structure des protéines . . . . .	63
7.1.1	Les acides aminés, structure primaire . . . . .	63
7.1.2	Structure secondaire . . . . .	65
7.1.3	Structure tertiaire . . . . .	66
7.1.4	Structure quaternaire . . . . .	67
7.2	L'utilité des protéines . . . . .	67
<b>8</b>	<b>Comment modéliser une protéine</b>	<b>71</b>
8.1	Modéliser et représenter un système protéique . . . . .	71
8.2	Les modèles disponibles . . . . .	72
8.3	Applications . . . . .	73
8.3.1	Umbrella sampling . . . . .	73
8.3.2	Docking . . . . .	73
8.3.3	Processus biologiques . . . . .	74
8.3.4	Prédiction et repliement . . . . .	75
<b>IV</b>	<b>Comment modifier la couleur de la bioluminescence</b>	<b>77</b>
<b>9</b>	<b>Modulation grâce à un analogue de la luciférine</b>	<b>81</b>
9.1	Contexte . . . . .	81
9.2	Problématiques . . . . .	82
9.3	Résultats . . . . .	83
9.3.1	Benchmark des méthodes . . . . .	83
9.3.2	Emission sans protéine . . . . .	86
9.3.3	L'influence de la dynamique moléculaire . . . . .	87
9.3.4	Le résidu 284 et le $T_e$ . . . . .	87
9.3.5	L'impact de l'environnement protéique . . . . .	89
9.3.6	L'étude conformationnelle . . . . .	91
9.4	Conclusion . . . . .	93
9.5	Détails de calculs . . . . .	94
9.5.1	Partie MM . . . . .	94
9.5.2	Partie MD . . . . .	95
9.5.3	Partie QM/MM . . . . .	96
<b>10</b>	<b>Modélisation et compréhension de nouvelles luciférases</b>	<b>99</b>
10.1	Contexte . . . . .	99
10.2	Etude de la $GB_{Av}$ , l'impact de la conformation du site actif . . . . .	100
10.2.1	Présentation du système . . . . .	100
10.2.2	Est-il possible de passer de $GB_{Av}$ -A à $GB_{Av}$ -B? . . . . .	101
10.2.3	Calcul de la longueur d'onde de bioluminescence . . . . .	102
10.2.3.1	Calcul de $T_e$ à partir des structures cristallographiques	102
10.2.3.2	Comprendre le déplacement vers le bleu . . . . .	104
10.2.3.3	Conclusion . . . . .	105
10.2.4	Le repliement de la protéine . . . . .	106
10.2.5	Calcul de l'émission dans les structures repliées . . . . .	107

TABLE DES MATIÈRES

10.2.5.1	Le modèle GB <sub>Av</sub> -B-closed_w_OxyLH <sub>2</sub> . . . . .	107
10.2.5.2	Le modèle GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed . . . . .	108
10.2.5.3	Conclusion . . . . .	111
10.2.6	Etude de mutations chez GB <sub>Av</sub> . . . . .	111
10.2.6.1	R332L . . . . .	112
10.2.6.2	I346L . . . . .	114
10.2.6.3	Insertion de R351 . . . . .	114
10.3	Prédire le comportement de l'environnement . . . . .	116
10.4	Conclusion . . . . .	118
10.5	Détails de calculs . . . . .	118
10.5.1	Partie MM . . . . .	118
10.5.2	Partie MD . . . . .	118
10.5.3	Partie QM/MM . . . . .	120

**V Etude du mécanisme réactionnel de la bio et chimiluminescence 121**

**11 La chimiluminescence des 2-coumaranones 125**

11.1	Contexte . . . . .	125
11.2	Résultats . . . . .	125
11.3	La synthèse de nouvelles 2-coumaranones . . . . .	127
11.4	Détails de calculs . . . . .	128

**12 Etude de la tautomérisation keto-enol 129**

12.1	Contexte . . . . .	129
12.2	La structure de départ . . . . .	130
12.3	Dans l'eau en solvant implicite (PCM) . . . . .	131
12.3.1	Géométrie des états de transition . . . . .	131
12.3.2	Les IRC . . . . .	133
12.3.3	Conclusion et perspectives . . . . .	135
12.4	Dans la protéine . . . . .	136
12.4.1	Géométrie des états de transition . . . . .	136
12.4.2	Les IRC . . . . .	138
12.4.2.1	Le TS1 . . . . .	138
12.4.2.2	Le TS2 . . . . .	140
12.5	Rotation . . . . .	142
12.6	Discussion . . . . .	143
12.7	Conclusion et perspective . . . . .	147
12.8	Détails de calculs . . . . .	147
12.8.1	Partie MM . . . . .	147
12.8.2	Partie MD . . . . .	148
12.8.3	Partie QM et QM/MM . . . . .	148
12.8.4	Discussion sur les méthodes . . . . .	150

**13 Etude de l'étape de coordination du dioxygène lors de la bioluminescence 151**

TABLE DES MATIÈRES

13.1	Contexte . . . . .	151
13.2	Déplacement de O <sub>2</sub> dans la protéine . . . . .	152
13.3	La coodination de O <sub>2</sub> . . . . .	154
13.3.1	Calcul préliminaire : minimisation par DFT/MM . . . . .	154
13.3.2	L'approche MS-CASPT2/MM . . . . .	155
13.3.2.1	Le choix de l'espace actif . . . . .	155
13.3.2.2	Le chemin réactionnel entre 3.5 Å et 2.3 Å. . . . .	157
13.3.2.3	Le chemin réactionnel entre 2.3 Å et 1.5 Å. . . . .	158
13.3.2.4	Retour sur les états singulets . . . . .	161
13.3.2.5	Chemin réactionnel après le départ de l'AMP . . . . .	161
13.4	Discussion . . . . .	165
13.4.1	L'espace actif en CASSCF . . . . .	166
13.4.2	La formation du dioxétanone . . . . .	167
13.5	Conclusion . . . . .	169
13.6	Perspective . . . . .	169
13.7	Détails de calculs . . . . .	170
13.7.1	Partie MM . . . . .	170
13.7.2	Partie MD . . . . .	171
13.7.3	Partie QM/MM . . . . .	171
	<b>Conclusion</b>	<b>175</b>
	<b>Bibliographie</b>	<b>191</b>
	<b>Annexes</b>	<b>195</b>
	<b>Abréviations</b>	<b>195</b>
	<b>Liste des protéines</b>	<b>197</b>
	<b>Orbitales CAS</b>	<b>199</b>
	<b>Publications</b>	<b>201</b>
	<b>Table des figures</b>	<b>204</b>
	<b>Table des tableaux</b>	<b>207</b>



# Introduction

Les lucioles sont un des plus anciens systèmes bioluminescents connus. Aristote évoquait déjà l'existence de cet insecte lumineux dans son livre *De anima*<sup>1</sup>. De nos jours, les recherches ayant pour thème la bioluminescence des lucioles jouissent toujours d'une popularité importante, avec de nombreuses applications issues de ce domaine. Ainsi depuis une quinzaine d'années, la sphère de la bioluminescence a été investie par les chimistes expérimentateurs puis par les chimistes théoriciens.

La bioluminescence chez les lucioles correspond à la réaction enzymatique d'un substrat appelé luciférine ( $LH_2$ ) catalysée par une protéine nommée luciférase. Le mécanisme réactionnel est présenté dans la Figure 1. Lors de la première étape, la luciférine réagit avec l'adénosine tri-phosphate (ATP), la molécule fournissant de l'énergie aux systèmes biologiques. Ces deux molécules pénètrent dans le site actif de la luciférase et adoptent leurs positions et conformations privilégiées. Lors de cette étape, la luciférase est dite en position ouverte avec son domaine C-terminal arrangé de tel façon à garantir une ouverture vers le site actif. La présence d'ion magnésium  $Mg^{2+}$  permet ensuite la réaction de l'ATP avec la luciférine afin de former l'intermédiaire de réaction luciférine-AMP par une réaction de type phosphorylation. La deuxième étape du mécanisme réactionnel fait ensuite intervenir un co-réactif, le dioxygène. La conformation de la luciférase évolue ensuite vers une position dite semi-fermée. L'intérêt de cette conformation de la protéine est de disposer d'un petit canal entre le site actif et le solvant permettant le transport du dioxygène du milieu extérieur vers l'intérieur de l'enzyme<sup>2</sup>. Le mécanisme réactionnel de la coordination du dioxygène sur la luciférine est encore mal compris, mais il fait intervenir la perte d'un proton sur le substrat afin de pouvoir fixer  $O_2$ . Une étude plus complète sur ce mécanisme est détaillée dans cette thèse afin d'avoir un point de vue plus explicite de cette étape. Finalement l'action du dioxygène entraîne la formation d'un cycle dioxétanone, présent dans la majorité des réactions bioluminescentes. Ce cycle est peu stable dû à la présence d'une liaison peroxyde O-O et se décompose par une coupure homolytique entre les deux oxygènes. Cette coupure libère du dioxyde de carbone et une nouvelle molécule, l'oxyluciférine ( $OxyLH_2$ ), la molécule émettrice de la bioluminescence chez les lucioles. Cette dernière est obtenue dans son 1<sup>er</sup> état singulet excité, et libère ensuite son énergie, donnant lieu à une émission de lumière visible de couleur verte et d'une longueur d'onde de l'ordre de 550 nm. Jusqu'au chapitre 12 le terme oxyluciférine ( $OxyLH_2$ ) correspond à la conformation dite phénolate-kéto du substrat comme représentée dans la Figure 1. L'explosion du nombre de recherches dans ce domaine a été permise par la découverte de la structure chimique du substrat et à la cristallisation de la structure protéique de la

## INTRODUCTION

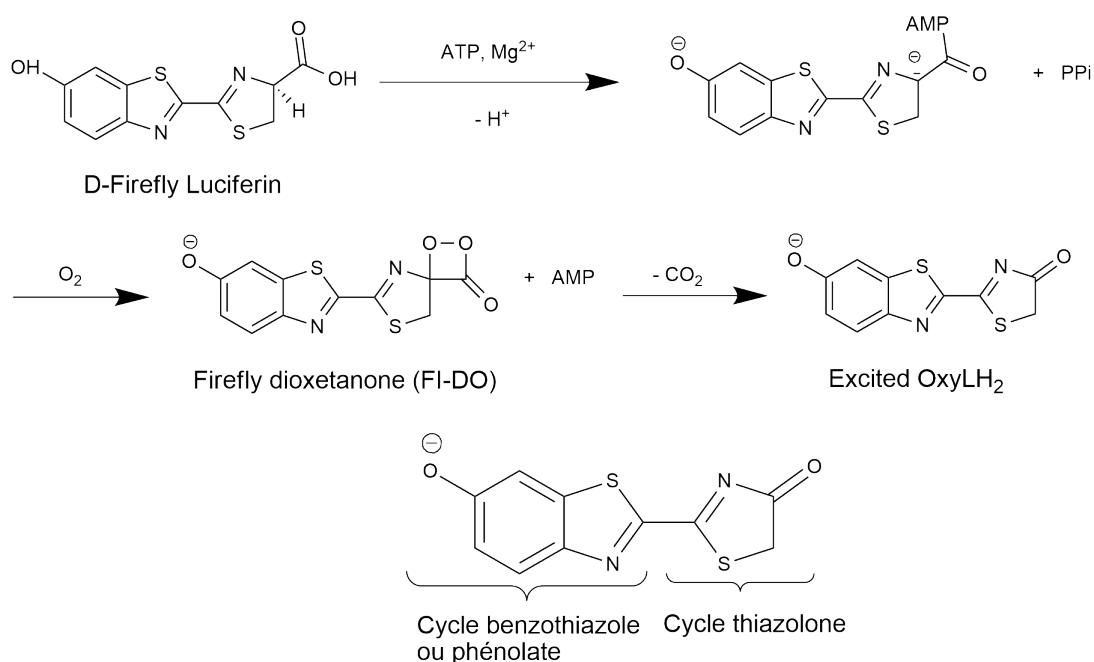


FIGURE 1 – Mécanisme de la bioluminescence chez les lucioles et structure chimique du substrat oxyluciférine (OxyLH<sub>2</sub>) dans la conformation phénolate-kéto.

luciférase<sup>3</sup>. Ainsi, la découverte de la structure chimique du substrat a permis la synthèse de ce composé<sup>4</sup>. Il n'est aujourd'hui plus nécessaire d'extraire la luciférine et la luciférase en broyant des lucioles. De plus, un nouveau domaine basé sur l'ingénierie chimique de ce composé a vu le jour. Il existe maintenant des dizaines d'analogues de la luciférine, plus ou moins différents et ayant des propriétés diverses<sup>5-7 8-12</sup>. Outre une modification de la longueur d'onde d'émission de la bioluminescence, certains substrats possèdent un nouveau groupement chimique permettant par exemple de les lier à d'autres molécules<sup>10</sup>. D'autre part, l'obtention de la structure cristallographique de la luciférase<sup>3</sup> permet de visualiser la position du site actif de la protéine et des différentes interactions pouvant s'exercer avec le substrat. La structure cristallographique donne aussi l'opportunité de mieux comprendre le mécanisme chimique conduisant au phénomène de la bioluminescence. Le nombre croissant de publications sur ce sujet permet aujourd'hui de cartographier les acides aminés ayant un impact sur la bioluminescence<sup>13-16</sup>. En outre, la multiplication des structures cristallographiques issues de différentes espèces bioluminescentes permet d'identifier les mutations naturelles ayant un rôle dans la réaction de bioluminescence<sup>17</sup>.

L'essor des méthodes QM/MM ainsi que l'augmentation de la puissance de calculs des ordinateurs autorisent maintenant des études théoriques complètes et précises des systèmes biochimiques. Les précédentes recherches réalisées par le Pr. Navizet et collaborateurs ont permis de donner un certains nombres d'indices pour expliquer une partie du mécanisme réactionnel. Je présente ici quelques anciens résultats importants pour la compréhension du système.



Dans la publication de Liu *et al.*<sup>18</sup> une étude de type CASPT2//CASSCF a été réalisée sur la dernière étape du mécanisme réactionnel, la formation de l'oxyluciférine à partir du cycle dioxétanone. Lors de cette étude plusieurs points cruciaux ont permis une meilleure compréhension du système. Le premier traite de la cassure du cycle dioxétanone. Cette dernière peut se dérouler de façon concertée ou en plusieurs temps selon la protonation du cycle benzothiazole. La perte de l'hydrogène sur ce cycle permet l'apparition d'un transfert de charge entre un électron de l'oxygène, du cycle benzothiazole chargé négativement de la luciférine et un oxygène du dioxétanone. Cela conduit à un mécanisme de cassure des liaisons C-C et O-O concerté et réduit par la même occasion la barrière d'activation de ce mécanisme par rapport au substrat avec l'oxygène du benzothiazole protoné. Il est alors primordial d'avoir une luciférine avec le fragment benzothiazole déprotoné, appelé par la suite forme phénolate, pour que la réaction se fasse. Le deuxième résultat obtenu montre que le cycle thiazolone adopte une conformation de type keto à la fin de la réaction de cassure. Ces deux résultats nous incitent à modéliser le système par un substrat de type phénolate-kéto pour calculer la longueur d'onde d'émission. Ce résultat sera discuté durant la thèse en comparant notamment la différence d'énergie entre l'émission obtenue sous une forme kéto et une forme énol et le mécanisme liant ces deux tautomères.

Un autre point important est discuté dans l'article de Navizet *et al.*<sup>19</sup>. Dans ce dernier, le calcul du chemin réactionnel des états électroniques les plus bas en énergie a été réalisé entre deux structures, celle de l'intermédiaire possédant le cycle dioxétanone et la structure du produit de réaction : l'oxyluciférine. Ainsi, il est possible de comparer le mécanisme aboutissant à la bioluminescence (produit de la réaction chimique) et celui obtenu lors d'une expérience de fluorescence (excitation lumineuse du produit keto de la réaction). Dans le cas de l'émetteur phénolate-kéto, quelque soit le mécanisme d'excitation, par absorption de photons ou par réaction chimiluminescente, le même minimum dans le premier état excité singulet est atteint après relaxation. On en déduit alors que la désexcitation du substrat, par bioluminescence ou par fluorescence, conduit à la même différence d'énergie, et donc à la même longueur d'onde d'émission. Ce résultat est très important car on peut donc obtenir la valeur de la longueur d'onde de la bioluminescence en calculant la valeur de la longueur d'onde d'émission de fluorescence.

Les résultats obtenus lors de cette thèse seront classés en 2 parties, notée IV et V. Dans la première on étudiera les différents facteurs pouvant affecter et modifier la longueur d'onde d'émission de la bioluminescence. On montrera notamment l'effet suscité par des analogues de l'oxyluciférine mais également l'impact des différentes modifications présentes dans des luciférases issues d'autres espèces bioluminescentes. L'étude d'un analogue de la luciférine a fait l'objet d'un article dans PCCP<sup>20</sup> et celle sur différentes espèces de luciférases est en cours de rédaction pour une letter dans Nature. Dans la seconde partie, on traitera plus en détail du mécanisme réactionnel de la bioluminescence chez les lucioles. Dans un premier temps on développera l'étape où le dioxygène se lie à la luciférine. Puis on discutera

## INTRODUCTION

de la réaction de tautomérisation kéto-énol du substrat dans la protéine et du rôle de ces deux formes comme possible émetteur de la bioluminescence. Ces deux études feront chacune l'objet d'un article, qui sont en cours de rédaction.

Avant de discuter de ces résultats, la partie I se focalisera sur les méthodes de calculs utilisées lors de cette thèse. Puis on traitera des différents type de luminescence dans les systèmes moléculaires dans la partie II avant de s'intéresser au fonctionnement des protéines et comment les modéliser dans la partie III.

# **Première partie**

## **Méthodes**



# Chapitre 1

## Méthodes de type ab initio

En chimie quantique, les propriétés moléculaires peuvent être déterminées par la résolution de l'équation de Schrödinger<sup>21</sup> dépendante du temps notée :

$$\hat{H}\Psi(r, t) = i\hbar \frac{\delta}{\delta t} \Psi(r, t) \quad (1.1)$$

où  $\Psi(r, t)$  est la fonction d'onde dépendante du temps et  $\hat{H}$  l'opérateur Hamiltonien.

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{Ne} + \hat{V}_{NN} + \hat{V}_{ee} + \hat{V}_{ext} \quad (1.2)$$

Pour le reste de ce manuscrit, l'état d'un système sera constitué de  $N$  électrons, de masse  $m_e$  et de charge  $-e$  ( $e$  désignant la charge électrique élémentaire), ainsi que de  $M$  noyaux, de masse  $M_A$  et de charge  $Z_A e$ . Enfin on appelle  $r_i$  les coordonnées spatiales des électrons et  $R_A$  les coordonnées spatiales des noyaux. On utilisera également le système d'unité atomique. Dans l'équation 1.2,  $\hat{T}_e$  et  $\hat{T}_N$  sont respectivement les opérateurs de l'énergie cinétique des électrons et des noyaux.

Ces termes sont définis de telles manières :

$$\hat{T}_e = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 \quad \text{et} \quad \hat{T}_N = -\frac{1}{2} \sum_A \frac{\nabla_A^2}{M_A} \quad (1.3)$$

Quant à  $\hat{V}_{Ne}$ ,  $\hat{V}_{NN}$  et  $\hat{V}_{ee}$ , ils représentent respectivement les opérateurs d'énergie potentiel d'interaction noyau-électron, noyau-noyau et électron-électron, et ont pour expression :

$$\hat{V}_{Ne} = -\sum_{i=1}^N \sum_A \frac{Z_A}{|r_{iA}|}, \quad \hat{V}_{NN} = \sum_A \sum_{B>A} \frac{Z_A Z_B}{|R_{AB}|}, \quad \hat{V}_{ee} = \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|r_{ij}|} \quad (1.4)$$

avec  $r_{iA}$  la distance entre l'électron  $i$  et le noyau  $A$ ,  $r_{AB}$  la distance entre le noyau  $A$  et le noyau  $B$  et  $r_{ij}$  la distance entre l'électron  $i$  et l'électron  $j$ .

Le dernier terme  $\hat{V}_{ext}$  représente le potentiel d'interaction des électrons et des noyaux avec l'extérieur, par exemple avec un champ électrique ou magnétique. Si ce dernier est nul, le système chimique est alors isolé et l'Hamiltonien est indépendant du temps.

L'équation de Schrödinger détaillé en 1.1 est souvent insolvable. Il est alors nécessaire d'inclure plusieurs approximations afin de pouvoir résoudre cette équation mathématique. Une de ces approximations est appelée l'approximation de Born-Oppenheimer<sup>22</sup>. Dans un système chimique la masse d'un électron est beaucoup plus faible que la masse d'un noyau, environ 1836 fois. Les électrons se déplacent alors beaucoup plus vite que les noyaux. L'approximation est donc de considérer les noyaux fixes et les électrons se déplaçant autour. On simplifie l'équation de Schrödinger en enlevant le terme  $\hat{T}_N$  et le terme  $\hat{V}_{NN}$  qui sont alors constants.

L'opérateur Hamiltonien s'écrit alors :

$$H_e = \hat{T}_e + \hat{V}_{Ne} + \hat{V}_{ee} \quad (1.5)$$

et  $H_e$  est nommé Hamiltonien électronique.

Dans une première partie nous détaillerons les méthodes utilisant la fonction d'onde afin de résoudre l'équation de Schrödinger comme la méthode Hartree-Fock (HF) ou MCSCF (Multi-Configuration Self-consistent Field). Puis dans un second temps on présentera les méthodes basées sur la densité électronique que sont la Théorie de la Fonctionnelle de la Densité (DFT) et la Théorie de la Fonctionnelle de la Densité Dépendante du Temps (TD-DFT). Ce chapitre a été écrit en se servant du livre Modern Quantum Chemistry<sup>23</sup> de A. Szabo et N. Ostlund et du livre Computational photochemistry<sup>24</sup> de M. Olivucci.

## 1.1 Méthode Hartree-Fock

La méthode Hartree-Fock est une approche variationnelle utilisée afin de décrire l'état électronique fondamental d'un système. C'est également une méthode basée sur un seul déterminant de Slater<sup>25</sup>, et donc nommée mono-déterminantale.

La fonction d'onde d'un système à N électrons est alors écrite comme un produit de N spin-orbitales moléculaires monoélectroniques antisymétrisées aussi appelé déterminant de Slater.

$$\Psi_e(1,2,\dots,N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \Phi_1(1) & \Phi_2(1) & \dots & \Phi_N(1) \\ \Phi_1(2) & \Phi_2(2) & \dots & \Phi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(N) & \Phi_2(N) & \dots & \Phi_N(N) \end{vmatrix} \quad (1.6)$$

avec  $\Phi_k$  une spin-orbitale moléculaire défini par :

$$\Phi_k = \phi_k \sigma_s \quad (1.7)$$

avec  $\phi_k$  une orbitale d'espace,  $s$  le spin de l'électron et  $\sigma_s$  la fonction de spin. Il existe deux fonctions de spin, une avec un spin up ( $\alpha$ ) et une avec un spin down ( $\beta$ ).

Ainsi avec un jeu de K orbitales il est possible de créer une jeu de 2K spin-orbitales.

La minimisation de l'énergie de ces spin-orbitales nous amène à l'équation aux valeurs propres :

$$\hat{f}(i)|\Phi_k\rangle = \varepsilon_k|\Phi_k\rangle \quad (1.8)$$

où  $\varepsilon_k$  est l'énergie de la spin-orbitale  $\Phi_k$  et  $\hat{f}(i)$  est appelé opérateur de Fock mono-électronique pour l'électron  $i$ . Il est défini par :

$$\hat{f}(i) = -\frac{\nabla_i^2}{2} - \sum_A \frac{Z_A}{r_{iA}} + \sum_l [\hat{J}_l(i) - \hat{K}_l(i)] \quad (1.9)$$

où  $\hat{J}_l$  est appelé opérateur de Coulomb et est défini à partir de son action sur  $\Phi_k(i)$  par :

$$\hat{J}_l(i)\Phi_k(i) = \left[ \int \frac{\Phi_l^*(j)\Phi_l(j)}{r_{ij}} dj \right] \Phi_k(i) \quad (1.10)$$

et  $\hat{K}_l$  est appelé opérateur d'échange et est défini à partir de son action sur  $\Phi_k(i)$  par :

$$\hat{K}_l(i)\Phi_k(i) = \left[ \int \frac{\Phi_l^*(j)\Phi_k(j)}{r_{ij}} dj \right] \Phi_l(i) \quad (1.11)$$

On utilise par la suite l'approche de type Restricted Closed-Shell Hartree-Fock, ou Hartree-Fock à couches fermées. Dans cette approche les spin-orbitales avec les spins  $\alpha$  et  $\beta$  possèdent la même fonction d'espace. De plus, le système étudié doit posséder un nombre pair d'électrons et ceux-ci sont appariés de façon à obtenir  $n$  orbitales doublement occupées avec  $n$  égal à  $N/2$ . L'équation (1.9) peut alors être exprimée de cette façon avec  $h^c$  appelé l'Hamiltonien de cœur :

$$\varepsilon(i) = h^c + \sum_k \sum_l^n [2J_{kl}(i) - K_{kl}(i)] \quad (1.12)$$

avec  $J_{kl}$  et  $K_{kl}$  correspondant respectivement à

$$J_{kl} = \langle \Phi_k | \hat{J}_l | \Phi_k \rangle \quad (1.13)$$

$$K_{kl} = \langle \Phi_k | \hat{K}_l | \Phi_k \rangle \quad (1.14)$$

Cette approche permet également d'éliminer le spin des équations aux valeurs propres. Les fonctions propres de l'opérateur de Fock deviennent les fonctions d'espace et (1.8) devient :

$$\hat{f}(i)|\Phi_k\rangle = \varepsilon_k|\Phi_k\rangle \quad (1.15)$$

### 1.1.1 Equation de Roothan

Afin de résoudre ces équations numériquement, Roothan<sup>26,27</sup> a introduit une base d'orbitales atomiques (OA)  $\{\chi_\mu\}$  sur laquelle on exprime les orbitales moléculaires (OM)  $\phi_k$  comme une combinaison linéaire de ces orbitales atomiques :

$$\phi_k = \sum_{\mu=1}^K C_{\mu k} \chi_\mu \quad (1.16)$$

les  $C_{\mu k}$  étant les coefficients de la combinaison linéaire et  $K$  le nombre d'orbitales atomiques dans la base. Si la base est complète, on aura un développement exact,

mais dans la pratique cela ne sera jamais le cas. Les OM qui possèdent des électrons sont dites "occupées" et les OM vides "virtuelles". On introduit cette base d'orbitales atomiques dans l'équation d'Hartree-Fock

$$\sum_{v=1}^K F_{\mu v} C_{v k} = \epsilon_k \sum_{v=1}^K S_{\mu v} C_{v k} \quad (1.17)$$

où les  $\epsilon_k$  sont les énergies des orbitales moléculaires,  $S_{\mu v}$  est la matrice symétrique de recouvrement des orbitales atomiques et  $F_{\mu v}$  est la matrice de Fock :

$$S_{\mu v} = \langle \chi_{\mu} | \chi_{\nu} \rangle \quad (1.18)$$

$$F_{\mu v} = \langle \chi_{\mu} | \hat{h}^c + \sum_l^n [2\hat{J}_l - \hat{K}_l] | \chi_{\nu} \rangle \quad (1.19)$$

$$F_{\mu v} = H_{\mu v}^c + \sum_l^n \sum_{\lambda \sigma} C_{\lambda l} C_{\sigma l} [2(\mu \nu | \lambda \sigma) - (\mu \sigma | \lambda \nu)] \quad (1.20)$$

$$F_{\mu v} = H_{\mu v}^c + \sum_{\lambda \sigma} P_{\lambda \sigma} [(\mu \nu | \lambda \sigma) - \frac{1}{2}(\mu \sigma | \lambda \nu)] \quad (1.21)$$

$$F_{\mu v} = H_{\mu v}^c + G_{\mu v}. \quad (1.22)$$

avec P la matrice densité définie comme :

$$P_{\lambda \sigma} = 2 \sum_l^n C_{\lambda l} C_{\sigma l} \quad (1.23)$$

Afin de résoudre l'équation (1.17) qui est non linéaire, il est nécessaire d'utiliser une approche itérative autocohérente appelée procédure SCF (champs autocohérent). En effet la matrice de Fock dépend des coefficients C de la combinaison linéaire. Avant d'appliquer cette procédure itérative on va transformer l'équation matricielle :

$$FC = \epsilon SC \quad (1.24)$$

en :

$$F'C' = \epsilon C' \quad (1.25)$$

avec  $C = XC'$  et  $F' = X^{\dagger}FX$ .

On utilise pour cela une matrice de transformation notée X qui permet d'obtenir un ensemble de fonctions de base orthonormale. Cette matrice est construite selon :

$$X^{\dagger}SX = I \quad (1.26)$$

avec I la matrice identité et  $X^{\dagger}$  la matrice adjointe de X.

Finalement les  $C'$  et  $\epsilon$  sont obtenus en diagonalisant la matrice  $F'$ .



### 1.1.2 La procédure SCF

Afin de résoudre (1.25), on utilise la procédure SCF en suivant l'algorithme suivant :

- 1- Définition du système étudié, c'est-à-dire expliciter la position et la nature des atomes, le nombre d'électrons total du système et une base atomique d'essai initiale.
- 2- Calcul des intégrales moléculaires que sont  $S$ ,  $H^c$ ,  $G$ ...
- 3- Diagonalisation de la matrice de recouvrement  $S$  afin d'obtenir la matrice  $X$ .
- 4- 1<sup>ère</sup> estimation de la valeur de la matrice densité  $P$ .
- 5- Calcul de la matrice de Fock  $F = H^c + G$ .
- 6- Transformation de  $F$  en  $F'$  dans la base orthogonale :  $F' = X^\dagger F X$ .
- 7- Diagonalisation de  $F'$  pour obtenir  $C'$  et  $\epsilon$ .
- 8- Calcul des coefficients  $C$  à partir de  $C'$  puis de la nouvelle matrice densité  $P$ .
- 9- Vérification de la convergence du calcul. Elle est atteinte si la différence entre deux pas consécutifs de la matrice  $P$  est inférieure à un seuil fixé.
- 10- Si la convergence du calcul n'est pas atteinte, retour au point 5 avec la nouvelle matrice  $P$ . Si la convergence est atteinte, fin du calcul.

### 1.1.3 L'énergie de corrélation

La méthode Hartree-Fock à couches fermées permet d'obtenir une bonne approximation de la fonction d'onde du système étudié en plus d'être une approche cohérente en taille. Cependant c'est une méthode dite monodéterminantale, basée sur un seul déterminant de Slater. Cette approximation peut engendrer des problèmes lors de calculs de la fonction d'onde d'états excités ou de systèmes à couches ouvertes. En outre, l'interaction électron-électron est traitée de façon moyennée. Ainsi deux électrons de spins différents peuvent occuper la même position dans l'espace.

L'énergie exacte  $E_0$  de l'état fondamental est différente de l'énergie calculée par la méthode Hartree-Fock  $E_{HF}$  d'une énergie appelée énergie de corrélation  $E_{corr}$  :

$$E_{corr} = E_0 - E_{HF} \quad (1.27)$$

La méthode Hartree-Fock est un bon point de départ pour les calculs d'états électroniques et a permis le développement de méthodes plus élaborées, appelées méthodes post-Hartree-Fock qui permettent d'incorporer cette énergie de corrélation.

## 1.2 Méthodes Post Hartree-Fock

Dans cette partie, on discute l'apport des méthodes post Hartree-Fock dans le traitement de systèmes moléculaires complexes, notamment dans le traitement des états excités.

### 1.2.1 La méthode d'interaction de configuration (IC)

La méthode d'interaction de configuration (IC ou CI) est une méthode où la fonction d'onde du système s'écrit comme la somme de plusieurs déterminants de Slater. Ainsi on ne tient pas seulement compte du déterminant de Slater de l'état fondamental appelée ici  $\phi_0$  mais également de tous les déterminants excités, également appelés configurations. On peut alors écrire :

$$|\Phi_{IC}\rangle = \sum_k C_k |k\rangle, \quad (1.28)$$

où les  $C_k$  sont les coefficients du développement,  $|k\rangle$  les configurations.

Si l'on considère le cas où l'on prend en compte toutes les configurations possibles, on effectue une interaction de configuration complète (full-CI). Le calcul permet d'obtenir la fonction d'onde exacte du système mais aussi son énergie. La corrélation électronique est alors traitée dans son intégralité. Cependant le nombre de déterminants augmente de façon exponentiel avec la taille du système et le nombre d'états excités considérés fait que la méthode full-CI ne peut être réalisée que sur de petits systèmes. Une solution est d'effectuer une troncature sur le nombre de déterminants choisis afin de limiter le temps de calcul tout en conservant une partie de la corrélation électronique. On peut ainsi tronquer les déterminants sur la base d'un critère énergétique, ou en ne conservant que certaines excitations.

$$|\Phi_{SD}\rangle = |\phi_0\rangle + \sum_{ar} C_a^r |\phi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} C_{ab}^{rs} |\phi_{ab}^{rs}\rangle \quad (1.29)$$

Ici par exemple  $|\Phi_{SD}\rangle$  correspond à la fonction d'onde du système en tenant compte des simples et doubles excitations. On a donc effectué une troncature au niveau des doubles excitations.  $|\phi_a^r\rangle$  correspond à la simple excitation d'un électron de la spin-orbitale  $\chi_a$  vers la spin-orbitale  $\chi_r$  avec le coefficient  $C_a^r$ .

### 1.2.2 La méthode CASSCF (Complete Active Space Self Consistent Field)

Les méthodes CASSCF<sup>28,29</sup> et plus généralement MCSCF (Champs Multi-Configurationnel Auto-Cohérent) sont des méthodes post-Hartree-Fock dites multi-configurationnelles où l'on a effectué une troncature sur la fonction d'onde que l'on peut écrire :

$$|\Phi_{MCSCF}\rangle = \sum_k C_k |\phi_k\rangle, \quad (1.30)$$

La différence avec la méthode IC précédemment explicitée est que l'on va optimiser dans le même temps les coefficients d'expansion  $C_k$  et les coefficients associés aux spin-orbitales. On récupère ainsi la majorité de la corrélation électronique malgré la troncature effectuée. En outre, une approche variationnelle et auto-cohérente permet d'obtenir de bon résultats tout en considérant peu de déterminants.

Lors de l'approche CASSCF on choisit un certain nombre de déterminants afin de construire un ensemble restreint d'orbitales, appelé espace actif. Les orbitales situées à l'extérieur de cet espace actif sont dites inactives si elles sont doublements occupées et virtuelles si elles sont inoccupées (voir Figure 1.1).

La fonction d'onde CASSCF est représentée par la combinaison linéaire de toutes les configurations possibles lors de la répartition des électrons actifs dans les différentes orbitales de l'espace actif. Si pour cet espace actif on considère tous les déterminants possibles on parle alors d'espace actif complet (CAS). Le reste des électrons présents dans les orbitales inactives sont optimisés par un procédé variationnel de type Hartree-Fock à couche fermée.

Afin de traiter les états excités on peut utiliser deux approches : une appelée state-specific et une dite state-average. Dans l'approche state-specific ou SS-CASSCF chaque état électronique est traité indifféremment des autres, il possède alors son propre jeu d'orbital. Dans l'approche state-average ou SA-CASSCF, on utilise un ensemble d'états électroniques de même symétrie spatiale ou de spin afin de calculer une énergie moyennée de ces états :

$$E_{average} = \sum_I \omega_I E_I \quad (1.31)$$

avec  $\omega_I$  le facteur du poids relatif de chaque état I. On obtient un jeu d'orbital moyennée ainsi qu'un nombre de fonctions d'ondes orthogonales égale au nombre d'états considérés dans le calcul.

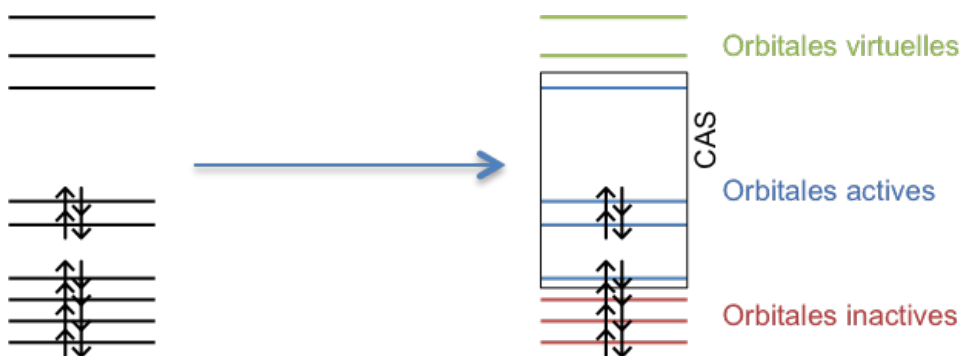


FIGURE 1.1 – Représentation schématique de l'approche CASSCF sur un système moléculaire donné. L'espace actif CAS contient 6 électrons dans 4 orbitales actives et est alors noté CAS(6-in-4). Les orbitales inactives et virtuelles correspondent respectivement aux orbitales doublement occupées et inoccupées situées hors de l'espace actif.

Le choix des orbitales à placer dans l'espace actif n'est pas évident. Une bonne connaissance du système et de la propriété étudiée sont nécessaires afin d'obtenir des résultats pertinents. La méthode CASSCF tiens compte de la corrélation non-dynamique pour les orbitales de l'espace actif. En revanche, les effets de corrélations dynamiques ne sont pas inclus.

### 1.2.3 La méthode CASPT2

La méthode CASPT2 (Complete Active Space Perturbation Theory 2) est une approche perturbative appliquée à une unique fonction d'onde CASSCF permettant d'ajouter les effets de type corrélation dynamique.

Dans cette méthode l'Hamiltonien du système peut être exprimé par la somme d'un Hamiltonien électronique de référence et d'un opérateur de perturbation faible  $\hat{V}$ .

$$\hat{H}(\lambda) = \hat{H}_0 + \lambda \hat{V} \quad (1.32)$$

Dans cette équation le  $\lambda$  est compris entre 0 ou 1. Pour  $\lambda = 1$  le système subit une perturbation totale alors que pour  $\lambda = 0$  on se retrouve dans le cas de la méthode CASSCF.

Le développement limité en  $\lambda$  permet d'obtenir :

$$|\Psi_k\rangle = \sum_{n=0}^{\infty} \lambda^n |\Psi_k^{(n)}\rangle \quad (1.33)$$

$$E_k = \sum_{n=0}^{\infty} \lambda^n \epsilon_k^{(n)} \quad (1.34)$$

Puis on remplace ces expressions dans l'équation de Schrödinger :

$$(\hat{H}_0 + \lambda \hat{V}) \left( \sum_{n=0}^{\infty} \lambda^n |\Psi_k^{(n)}\rangle \right) = \left( \sum_{n=0}^{\infty} \lambda^n \epsilon_k^{(n)} \right) \left( \sum_{n=0}^{\infty} \lambda^n |\Psi_k^{(n)}\rangle \right) \quad (1.35)$$

Dans l'approche CASPT2<sup>30</sup> on utilise un traitement perturbatif à l'ordre 2 dérivé de la méthode perturbative Møller-Plesset<sup>31</sup> qui donne une correction au second ordre valant :

$$\epsilon^{(2)} = \sum_{l \neq k} \frac{|\langle \Psi_k^{(0)} | \hat{V} | \Psi_l^{(0)} \rangle|^2}{\epsilon_k^{(0)} - \epsilon_l^{(0)}} \quad (1.36)$$

La méthode CASPT2 nécessite une fonction d'onde multi-déterminantale de type CASSCF à la place de la fonction d'onde monodéterminantale issue de la méthode Hartree-Fock comme dans le cas de MP2. La perturbation  $\hat{V}$  tient compte d'un nombre d'excitations beaucoup plus important et ces dernières doivent être traitées différemment selon l'OM de départ et d'arrivée. Les fonctions issues de simples et doubles excitations sont incluses dans  $V_{SD}$ , l'espace d'interaction de premier ordre. Cet espace est ensuite divisé en 8 sous-espaces selon la nature de l'excitation de la configuration donnée (voir Figure 1.2).

On retrouve cependant certains problèmes intrinsèques à la méthode CASPT2. Des états intrusifs peuvent apparaître pendant la correction au second ordre et induire une différence d'énergie importante entre l'énergie CASSCF et l'énergie CASPT2. Cela est possible car l'énergie de ces états intrusifs est proche de l'état de référence sans pour autant être inclus dans la fonction d'onde de référence. Ils peuvent être détectés par un poids quasi-nul dans la fonction d'onde CASPT2. Afin de traiter ces états intrusifs, il est possible d'utiliser plusieurs type de "shift" afin de décaler la position de ces états.

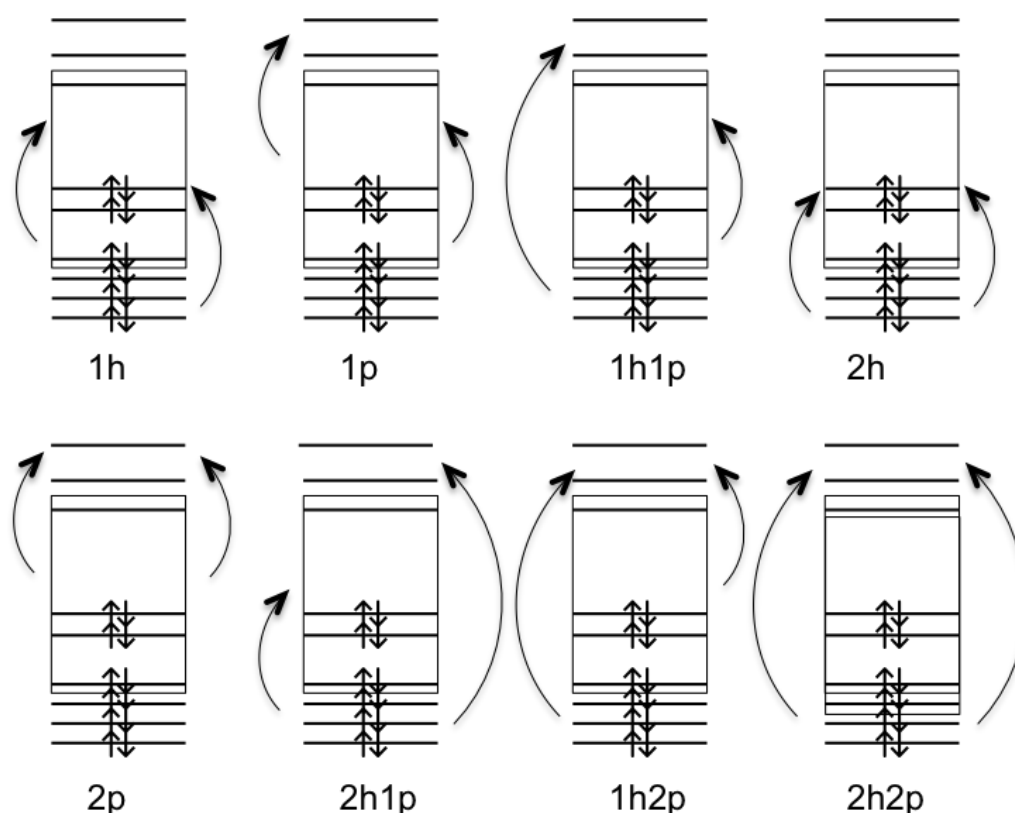


FIGURE 1.2 – Représentation schématique des doubles excitations incluses dans  $V_{SD}$ . L'espace actif est inclus à l'intérieur de la boîte dessinée, sous cette boîte se trouve les orbitales inactives et au-dessus les orbitales virtuelles. Ces doubles excitations sont nommées selon le nombre de trou (hole, h) dans l'espace inactif et particule (p) dans l'espace virtuel créés par ces doubles excitations. Ainsi la double excitation 2h1p signifie que cette dernière crée 2 trous dans les orbitales inactives et ajoute 1 particule dans les orbitales virtuelles.

Le développement de deux différents shifts, le level shift et l'IPEA shift est présenté ci-dessous.

Le level-shift CASPT2<sup>32</sup> (LS-CASPT2) introduit un paramètre de shift  $\epsilon$  dans l'Hamiltonien non perturbé de telle sorte que :

$$E^2 \approx \tilde{E}^2 - \epsilon \left( \frac{1}{\tilde{\omega}} - 1 \right) \equiv E_{LS}^2 \quad (1.37)$$

où  $\tilde{E}^2$  est l'énergie approchée de la correction au second ordre et  $\tilde{\omega}$  le poids de la fonction d'onde CASSCF. Ce shift est fixé par défaut à 0 et ne doit pas être trop élevé, une valeur de  $\epsilon$  comprise entre 0.1 et 0.4 est suffisante pour obtenir des valeurs de LS-CASPT2 cohérentes.

Le shift dit IPEA<sup>33</sup> peut être défini par :

$$\epsilon = (IP)_t - (EA)_t \quad (1.38)$$

où IP est le potentiel de ionisation, l'énergie nécessaire pour arracher un électron de l'orbitale  $t$ . EA est l'affinité électronique, l'énergie nécessaire pour attacher un électron extérieur à l'orbitale  $t$ . Ce shift varie bien entendu selon le système étudié. Ainsi une valeur moyenne  $\epsilon = 0.25$  a été choisi après confrontation avec des résultats expérimentaux.

### 1.2.4 La méthode MS-CASPT2

La méthode Multi-State CASPT2<sup>34</sup> (MS-CASPT2) est une extension de la méthode CASPT2 afin de traiter des systèmes comportant deux états de référence ou plus. Parmi les exemples les plus connus on peut utiliser la méthode MS-CASPT2 pour l'optimisation d'une intersection conique ou d'états de Rydberg.

La méthode MS-CASPT2 utilise un Hamiltonien effectif où les éléments diagonaux sont les énergies CASPT2 et les éléments non-diagonaux le couplage entre les différents états.

Par exemple, considérons un calcul MS-CASPT2 avec deux états de références. L'Hamiltonien effectif sera alors :

$$H^{eff} = \begin{Bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{Bmatrix} \quad (1.39)$$

et  $H_{21} = H_{12}$

Dans notre exemple  $H_{11}$  et  $H_{22}$  sont les énergies CASPT2 des états 1 et 2 respectivement. De plus,  $H_{21}$  est l'énergie de couplage entre l'état 1 et 2.

Il est également important que la valeur des couplages entre les différents états soit faibles sinon les résultats obtenus peuvent être peu fiables. Afin d'éviter ce problème, on peut soit augmenter le nombre d'états électroniques inséré dans le calcul SA-CASSCF, soit augmenter la taille de l'espace actif.

## 1.3 Méthodes basées sur la densité électronique

### 1.3.1 La méthode DFT (Density Functional Theory)

En parallèle des méthodes de calcul de structure électronique utilisant la fonction d'onde, il est possible de considérer l'énergie du système comme une fonctionnelle de sa densité électronique. Cette méthode se nomme la théorie de la fonctionnelle de la densité (DFT). La densité d'un système à  $N$  électrons est définie par l'équation suivante :

$$\rho(r_1) = N \int \dots \int |\Psi_e(x_1, x_2, \dots, x_N)|^2 d\sigma_{s1} dx_2 \dots dx_N \quad (1.40)$$

où  $\Psi_e$  est la fonction d'onde électronique et  $x_i = (r_i, \sigma_{si})$  regroupe les coordonnées d'espace et de spin de l'électron  $i$ . De plus si  $\rho \geq 0$  alors :

$$\int \rho(r) dr = N \quad (1.41)$$

La DFT repose sur la base de deux théorèmes démontrés par Hohenberg et Kohn<sup>35</sup>. Le théorème 1 montre que pour tout système polyélectronique soumis à un potentiel extérieur  $v(r)$ , ce dernier est déterminé, à une constante près, par la densité électronique de l'état fondamental. Il est alors possible d'écrire :

$$\rho(r) \rightarrow v(r) \rightarrow \hat{H} \rightarrow E \quad (1.42)$$

Les flèches signifient permettant de calculer.

Ainsi, la connaissance de la densité électronique permet d'accéder à la valeur du potentiel et ainsi de suite. La théorème 2 montre que l'énergie de l'état fondamental  $E_0$  est une fonctionnelle de la densité électronique  $E(\rho_0)$ . En conséquence la densité  $\rho$  de l'état fondamental minimise la fonctionnelle  $E(\rho)$ . Il est donc possible d'utiliser un principe variationnel sur une densité quelconque afin d'approcher l'énergie  $E_0$ .

$$E_0 = \min_{\rho} E[\rho] \quad (1.43)$$

L'énergie peut alors être écrite comme la somme du terme d'énergie cinétique des électrons  $T$ , l'énergie potentielle électrons-noyaux  $V_{ne}$  et l'énergie potentielle électron-électron  $V_{ee}$ .

$$E[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \quad (1.44)$$

Le problème est que seule l'expression de  $V_{ne}$  est connue. Afin de résoudre ce problème Kohn et Sham<sup>36</sup> ont émis l'hypothèse que les électrons peuvent être traités comme des particules indépendantes évoluant dans un potentiel externe. Il est alors possible de réexprimer l'énergie du système en des termes connus :

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + V_{ee}^{classique}[\rho] + E_{xc}[\rho] \quad (1.45)$$

avec  $T_s[\rho]$  l'énergie cinétique des électrons sans interaction et :

$$E_{xc}[\rho] = V_{ee}^{non-classique}[\rho] + \Delta T \quad (1.46)$$

$E_{xc}[\rho]$  est appelé fonctionnelle d'échange-corrélation et comprend les termes d'échange et de corrélation contenus dans  $V_{ee}^{non-classique}$  ainsi que  $\Delta T = T - T_s$  l'énergie cinétique correspondant à l'interaction entre les électrons. On peut maintenant résoudre les équations dites de Kohn-Sham. Posons

$$V_{ext}[\rho] = V_{ne}[\rho] + V_{ee}^{classique}[\rho] + E_{xc}[\rho] \quad (1.47)$$

On se retrouve alors avec :

$$\hat{f}_{KS} \phi_k = [T_s + V_{ext}] \phi_k = \epsilon_k \phi_k \quad (1.48)$$

avec  $\phi_k$  les orbitales de Kohn et Sham défini comme :

$$\rho = \sum_{k=1}^N \phi_k^* \phi_k \quad (1.49)$$

On se retrouve dans le même cas que lors de la méthode Hartree-Fock. On peut alors utiliser une procédure de champs-autocohérent afin de résoudre ces équations.

### 1.3.2 Les fonctionnelles d'échange et de corrélation

La méthode DFT se repose sur le terme d'échange et de corrélation  $E_{xc}[\rho]$  qui n'est pas connu et doit alors être approximé. Les fonctionnelles construites en conséquence dérivent d'un raisonnement purement physique ou sont fittées à partir de résultats expérimentaux.

Les fonctionnelles sont ainsi classées selon une échelle dite de Jacob où l'ascension de cette échelle correspond à une qualité de calcul croissante. Enfin, une fonctionnelle est souvent cohérente pour une propriété donnée mais donne des résultats médiocres pour d'autres propriétés. Il est important de tester plusieurs fonctionnelles afin d'être rigoureux. Un mini-benchmark sera étudié dans la partie résultat.

La classification de l'échelle de Jacob commence par la "terre" correspondant au niveau d'un calcul Hartree-Fock et se termine par le "paradis", représentant un résultat exact<sup>37,38</sup>.

Le premier barreau correspond à la Local Density Approximation (LDA). Dans cette approximation la densité électronique est décrite comme un gaz d'électron uniforme. Le spin des électrons peut également être pris en compte grâce à une variante de cette méthode la Local Spin Density Approximation (LSDA). L'énergie d'échange peut être écrite comme :

$$E_x^{\text{LSDA}}[\rho] = -2^{1/3} * C_x \int \rho_\alpha^{4/3} * \rho_\beta^{4/3} dr \quad (1.50)$$

avec  $C_x$  un coefficient.

Quant à l'énergie de corrélation, elle n'est connue que dans la limite des hautes et basses densités électroniques. Cependant des calculs de type Monte Carlo quantique ont permis la paramétrisation de fonctionnelles reproduisant l'énergie de corrélation. Ainsi plusieurs fonctionnelles de corrélations ont vu le jour comme par exemple la fonctionnelle VWN<sup>39</sup>. La LDA est toujours utilisée pour de grands systèmes où la densité électronique est homogène. Dans le cas d'une densité hétérogène il est nécessaire d'utiliser d'autres approximations.

Le deuxième barreau est l'approximation du gradient généralisé (GGA) et permet de tenir compte de la variation de la densité électronique au niveau local dans le système. Ainsi l'énergie d'échange corrélation tient compte de la densité électronique mais également de son gradient :

$$E_{xc}^{\text{GGA}}[\rho] = \int \rho \epsilon_{xc}[\rho, |\nabla\rho|] dr \quad (1.51)$$

Cette approximation est souvent utilisée comme correction de la fonctionnelle LSDA, le terme d'échange corrélation s'exprime alors comme :

$$E_{xc}^{\text{GGA}}[\rho] = E_{xc}^{\text{LSDA}} + \Delta E_{xc} \frac{|\nabla\rho|}{\rho^{4/3}} \quad (1.52)$$



On peut par exemple citer la fonctionnelle Perdew-Burke-Ernzerhof (PBE)<sup>40</sup> ou Perdew-Wang (PW)<sup>41</sup>.

Le troisième barreau correspond aux fonctionnelles de type meta-GGA. Ces fonctionnelles font intervenir le terme du laplacien de la densité en plus du gradient dans le terme d'échange corrélation. Les fonctionnelles de la famille des M06<sup>42</sup> développées par le groupe de Truhlar en sont un exemple.

Le quatrième barreau correspond aux fonctionnelles hybrides. Une partie de l'énergie d'échange calculée au niveau Hartree-Fock est inclus dans le terme d'échange de la fonctionnelle. On fait varier un paramètre  $\lambda$  entre 0 et 1 représentant le pourcentage d'interaction des électrons entre eux. On peut écrire :

$$E_{xc}[\rho] = \int_0^1 \langle \Psi_\lambda | V_{xc} | \Psi_\lambda \rangle d\lambda \quad (1.53)$$

On fixe alors un pourcentage d'échange exact à inclure dans la fonctionnelle à construire. Les deux fonctionnelles hybrides les plus utilisées sont B3LYP et PBE0. PBE0<sup>43</sup> est une fonctionnelle purement physique c'est à dire sans paramètres de fitting et inclut un échange Hartree-Fock de 25%. Pour B3LYP<sup>44,45</sup>, on combine la fonctionnelle d'échange de Becke<sup>46</sup> à 3 paramètres (B) avec 20% d'échange Hartree-Fock et la fonctionnelle de corrélation de Lee-Yang-Parr<sup>47</sup> (LYP). Ainsi la formule générale de la fonctionnelle B3LYP est :

$$E_{xc}^{B3LYP} = aE_x^{HF} + (1-a)E_x^{LSDA} + b\Delta E_x^B + cE_c^{LYP} + (1-c)E_c^{LSDA} \quad (1.54)$$

avec  $a=0.2$ ,  $b=0.72$  et  $c=0.81$  On peut aussi citer M06-2X<sup>42</sup> qui comporte 54% d'échange Hartree-Fock.

Par la suite plusieurs types de corrections ont été incluses dans de nouvelles fonctionnelles. Il existe par exemple des fonctionnelles avec des corrections à longue distance (LC-) ou incluant les effets de dispersions(-D).

Le principe des fonctionnelles avec correction à longue distance est de faire varier le pourcentage d'échange Hartree-Fock inclus dans la fonctionnelle selon la distance inter-électronique. On peut citer la fonctionnelle CAM-B3LYP<sup>48</sup> qui est dérivée de B3LYP. Les fonctionnelles avec dispersion incluent un terme supplémentaire de type van der Waals paramétrisé et possèdent une fonction d'atténuation. Les fonctionnelles de Grimme DFT-D<sup>49</sup> incluent cette dispersion à différents ordres.

Enfin au cinquième barreau de l'échelle de Jacob on trouve les fonctionnelles hyper-GGA ou double-hybride qui utilise l'information des orbitales de Kohn-Sham occupées. On trouve, par exemple les fonctionnelles B2PLYP<sup>50,51</sup> de Grimme utilisant des calculs de type MP2 afin de calculer l'énergie de corrélation.

### 1.3.3 La méthode TD-DFT (Time-Dependent DFT)

La TD-DFT est le pendant de la DFT afin de calculer les états excités d'un système. Cette méthode est basée sur un seul déterminant de Slater donc est moins coûteuse

que des méthodes de type MCSCF étudiées précédemment.

L'existence de la TD-DFT est postulée par le théorème de Runge-Gross. Il montre que si deux potentiels externes dépendant du temps  $t$  ne diffèrent que par une constante dépendant du temps alors ces deux potentiels possèdent la même densité électronique. On en déduit que les fonctions d'onde créées par ces deux potentiels ne diffèrent que par un facteur de phase près.

Par la suite Van Leeuwen a montré que le théorème de Runge-Gross peut s'appliquer sur les systèmes moléculaires. On peut alors résoudre les équations de Kohn et Sham comme dans le cas de la DFT. La différence est que dans le cas de la TD-DFT ce sont des équations dépendantes du temps.

Les équations de Kohn et Sham dépendantes du temps peuvent alors être écrites comme :

$$E[\rho, t] = T_s[\rho, t] + V_{ext}[\rho, t] \quad (1.55)$$

$$V_{ext}[\rho, t] = V_{ne}[\rho, t] + V_{ee}^{classique}[\rho, t] + E_{xc}[\rho, t] \quad (1.56)$$

et donc l'équation de Schrödinger dépendante du temps est :

$$\hat{f}_{KS}\phi_k(t) = [T_s + V_{ext}]\phi_k(t) = \varepsilon_k\phi_k(t) \quad (1.57)$$

Il est possible de résoudre cette équation en utilisant l'approximation de la réponse linéaire.

L'approximation de la réponse linéaire suppose que le système subit une perturbation de faible amplitude  $V_{ext}$ . On peut alors analyser la réponse linéaire de la densité à cette perturbation ainsi que la fonction de réponse des orbitales Kohn Sham. Le développement à l'ordre 1 de ces équations permet d'obtenir les énergies d'excitation et donc les énergies des états excités du système.

## 1.4 Les bases atomiques d'orbitales

Comme décrit dans la partie 1.1.1 l'utilisation d'une combinaison linéaire d'une base d'orbitales atomiques  $\chi_\mu$  permet la définition des orbitales moléculaire  $\phi_j$ . Il est alors nécessaire de trouver des fonctions mathématiques décrivant bien le comportement d'une orbitale moléculaire.

Deux types de fonctions mathématiques ont été historiquement utilisées : les orbitales slatériennes et les orbitales gaussiennes. Elles peuvent être définies respectivement pour une orbitale 1s centrée en  $R_A$  par :

$$\chi_{1s}^{SF}(\zeta, r - R_A) = (\zeta^3/\Pi)^{1/2} e^{-\zeta|r-R_A|} \quad (1.58)$$

avec  $\zeta$  le coefficient de l'orbitale slatérienne.

$$\chi_{1s}^{GF}(\alpha, r - R_A) = (2\alpha/\Pi)^{3/2} e^{-\alpha|r-R_A|^2} \quad (1.59)$$

avec  $\alpha$  le coefficient de l'orbitale gaussienne.

Les orbitales slatériennes décrivent très précisément les orbitales moléculaires avec un nombre minimal de fonctions à considérer mais la procédure SCF inclut le calcul des intégrales à deux électrons (1.21). Or il est très difficile et long de calculer ces intégrales avec des orbitales slatériennes. Le choix d'orbitales gaussiennes est quand à lui justifié car le calcul de ces intégrales doubles est simple. En effet, le produit de deux gaussiennes de centres différents est une troisième gaussienne plus une constante. Cependant les fonctions gaussiennes décrivent mal les orbitales moléculaires.

Dans la pratique, le choix des orbitales gaussiennes s'est vite démocratisé. En utilisant les propriétés précédentes, on peut décrire une orbitale slatérienne par une combinaison linéaire d'orbitales gaussiennes. Cette combinaison linéaire est appelée fonctions gaussiennes contractées CGF et définies comme :

$$\chi_{\mu}^{\text{CGF}}(r - R_A) = \sum_{p=1}^L d_{p\mu} \chi_p^{\text{GF}}(\alpha_{p\mu}, r - R_A) \quad (1.60)$$

avec  $L$  la taille de la contraction et  $d_{p\mu}$  le coefficient de contraction.

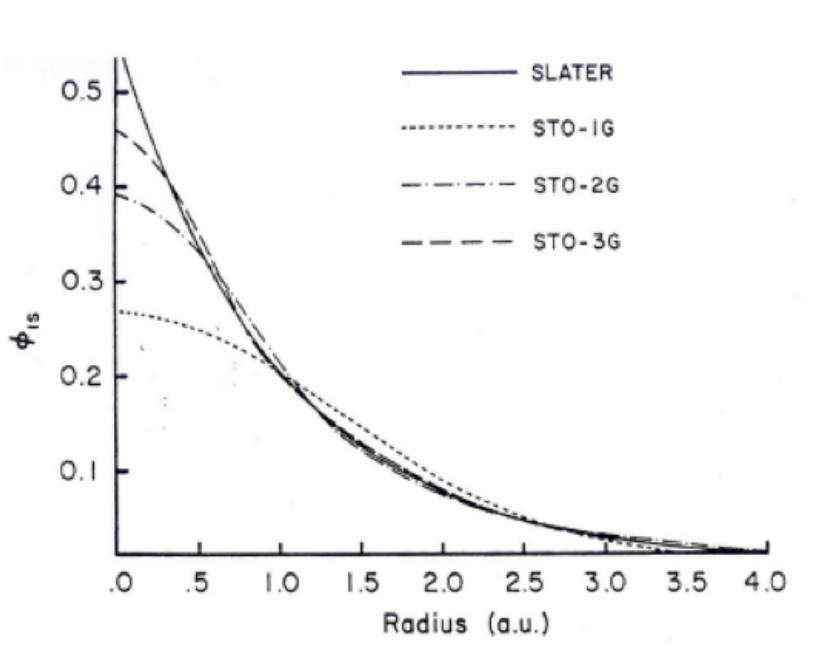


FIGURE 1.3 – Comparaison de la qualité d'une fonction de Slater 1s par rapport à des fonctions STO-1G, STO-2G et STO-3G. Cette figure est extraite du livre<sup>23</sup>, page 158.

On définit alors les bases de type STO-LG pour contracted slater type orbital avec  $L$  le nombre d'orbitales gaussiennes. La figure 1.3 montre le comportement des bases STO-LG comparées à une orbitale slatérienne. La base STO-3G qui se rapproche le

plus d'une orbitale slatérienne 1s quand est définie par les coefficients suivants :

$$\chi_{1s}^{\text{CGF}}(\zeta = 1.0, \text{STO} - 3\text{G}) = 0.444635\chi_{1s}^{\text{GF}}(0.109818) + 0.535328\chi_{1s}^{\text{GF}}(0.405771) + 0.154329\chi_{1s}^{\text{GF}}(0.2.22766) \quad (1.61)$$

La base STO-3G est appelée base minimale car chaque orbitale moléculaire n'est définie que par une seule fonction CGF. L'utilisation d'une base minimale diminue le temps de calcul mais donne des résultats peu précis. Par exemple, des atomes de la même période mais avec un nombre d'électrons différent sont décrits de la même façon alors qu'ils présentent des comportements différents. D'autres défauts sont également à noter comme l'impossibilité de faire varier le coefficient  $\alpha$  pendant le calcul ou que seules les densités électroniques sphériques peuvent être décrites.

### 1.4.1 Les bases de type split-valence

La solution mise en place est la création de bases dites N-zéta avec N le nombre de fonctions choisies pour décrire une orbitale moléculaire. On peut donc parler de base dite double-zéta, triple-zéta... où chaque orbitale est décrite par deux fonctions ou trois fonctions CGF respectivement.

On s'intéresse plus particulièrement à la base 6-311G qui est souvent utilisée dans le reste de ce manuscrit. Cette base appartient à la famille des bases triple-zéta et peut-être exprimée par :

- Les orbitales de cœur sont décrites par 6 fonctions gaussiennes primitives
- Les orbitales de valence sont décrites par 3 fonctions gaussiennes différentes : la première est constituée de 3 fonctions gaussiennes primitives, la deuxième et la troisième par une orbitale gaussienne non-contractée.

Par cette description on remarque qu'il est incorrect de dire que cette base est triple-zéta car les orbitales de cœur ne sont décrites que par une seule gaussienne contractée. Cependant, il est raisonnable de penser que les orbitales de valences sont majoritairement responsables des propriétés chimiques. Ces bases sont nommées bases de type split-valence.

Par la suite d'autres effets on été introduits pour améliorer l'efficacité des bases. Les effets dits de polarisations permettent d'inclure dans les bases atomiques une fonction de polarisation. En effet, dans un système chimique la densité électronique d'un atome n'est plus symétrique car elle entre en interaction avec les atomes voisins. Les orbitales sont alors déformées et un moyen de mieux les représenter est d'utiliser une fonction de polarisation. Par exemple, la base 6-311G\* contient une fonction de polarisation pour chaque atome lourd (hors hydrogène) correspondant à l'insertion d'une fonction gaussienne issue des orbitales d pour les éléments de la seconde période et une fonction gaussienne issue des orbitales p pour les éléments de la première période. La base 6-311G\*\* inclut également une fonction de polarisation sur les hydrogènes. Dans notre manuscrit, la base la plus utilisée sera 6-311G(2d,p) signifiant que 2 fonctions gaussiennes supplémentaires issues

d'orbitales d sont ajoutées pour les éléments lourds et une fonction gaussienne de type p pour les orbitales d'hydrogènes.

Finalement des fonctions dites diffuses sont également apparues. Elles servent à représenter des systèmes dont une partie importante de la densité électronique est située loin du noyau comme dans les anions ou les molécules possédant une paire libre. La base 6-311+G possède une fonction diffuse supplémentaire sur les fonctions gaussiennes de type s et p et pour les atomes lourds. La base 6-311++G possède la même fonction diffuse mais sur tous les atomes (hydrogènes compris).

### 1.4.2 Les bases de type ANO

Il existe également un autre type de base, appelée ANO pour Atomic Natural Basis où la contracted gaussian fonction (CGF) est construite différemment. Dans le cas des ANO, la CGF correspond à une orbitale atomique, chaque orbitale gaussienne participe alors à plusieurs CGF. Cette méthode présente plusieurs avantages :

- une meilleure description des orbitales atomiques
- un sens physique aux CGF
- une réduction de l'erreur de superposition de base

Cependant le nombre de fonctions de bases nécessaires est beaucoup plus importants, augmentant le temps de calcul.

La construction de ces orbitales ANO<sup>52,53</sup> se fait par des calculs hautement corrélés. Considérons la base ANO-RCC-VTZP utilisée dans la partie IV et V. Cette base triple-zéta à été définie à partir de calcul CASPT2//CASSCF pour la matrice densité de l'état fondamental, du 1<sup>er</sup> état excité, des ions positifs et négatifs et du dimère de l'atome considéré. En outre ces calculs tiennent également compte du couplage spin-orbite (via l'hamiltonien DKH). Cette base contient également des fonctions de polarisations et des fonctions diffuses.



# Chapitre 2

## Méthodes basées sur la mécanique moléculaire

Dans le chapitre précédent on a développé les méthodes ab-initio basées sur l'étude des électrons du système étudié. Malheureusement, même en utilisant des méthodes quantiques peu onéreuses comme la DFT, il est difficile d'étudier des molécules possédant plus d'une centaine d'atomes. Il est alors nécessaire de trouver une autre façon de décrire de larges systèmes comme par exemple les molécules d'intérêt biologique que sont les protéines. Ce chapitre a été écrit en se servant du livre *Molecular Modelling : Principles and Applications*<sup>54</sup> de A. Leach.

### 2.1 Les champs de force

En mécanique moléculaire, l'énergie potentielle d'un système est calculée grâce à l'action d'un champ de force (CF). Ces champs de force contiennent un ensemble de paramètres permettant de décrire les interactions à l'intérieur d'un système donné. Contrairement aux méthodes quantiques, l'énergie du système est calculée à partir des positions atomiques des noyaux, n'utilisant la position des électrons que de façon implicite. Il est alors impossible d'accéder aux informations obtenues dans la matrice densité électronique comme par exemple les excitations électroniques ou les énergies de polarisation.

L'utilisation de CF requiert un certain nombre d'approximations. La première est l'approximation de Born-Oppenheimer permettant de découpler les contributions nucléaires et électroniques. La deuxième est le principe de transférabilité. En effet, les paramètres des CF sont construits sur de petits modèles afin d'obtenir des valeurs très précises. Cependant, les systèmes décrits par des CF sont souvent issus des systèmes biologiques et sont en conséquence très volumineux. Par exemple, la protéine luciférase décrite dans les parties IV et V comporte plus de 8000 atomes. Il est alors obligatoire que les paramètres construits sur les petits systèmes présentent un comportement similaire pour tous les systèmes et pour toutes les tailles. C'est ce qu'on appelle la transférabilité des CF.

Les paramètres des CF sont classés en deux types : les termes liés et les termes

non liés. Parmi les termes liés on retrouve les énergies des longueurs de liaisons, les énergies de la valeurs des angles et des torsions entre atomes. Dans les termes non-liés, il s'agit principalement des énergies de répulsion (van der Waals) et des interactions de type électrostatiques.

Il existe également deux catégories principales de CF, les classiques et les polarisables. Les champs de force polarisables (CFP) tiennent également compte de la réponse de l'environnement dans les termes non-liés. Les champs de forces classiques (CFC) sont les plus courants et les plus utilisés mais ne possèdent pas cet effet de polarisation.

Par la suite nous détaillerons la fonction d'énergie potentielle utilisée dans les CF ainsi que le développement mathématique des différents paramètres avant de lister quelques exemples de champs de force et de détailler la modélisation du solvant en mécanique moléculaire.

### 2.1.1 La fonction d'énergie potentielle

Dans un CFC, la fonction d'énergie potentielle regroupe des termes associés à des atomes liés entre eux de façon covalente. En générale on considère ces interactions pour des groupes de 4 atomes au maximum. En plus de ces termes liés, il existe des termes non-liés regroupant les interactions électrostatique et de van der Waals (vdW). Dans un système composé de N atomes de coordonnées  $r$  on a :

$$V_{MM}(r_N) = \underbrace{V_{liaisons} + V_{angles} + V_{dièdres} + V_{impropres}}_{V_{liés}} + \underbrace{V_{vdW} + V_{el}}_{V_{non-liés}} \quad (2.1)$$

Une notion importante des CF est la définition des atomes. Alors qu'explicitier la charge globale ou la multiplicité (du spin) d'un système n'est pas obligatoire il est nécessaire de spécifier les différents types d'atomes. Outre le nom et le numéro atomique, il est aussi important de préciser l'état d'hybridation de ces atomes. La définition des atomes peut être difficile car leurs types dépend de l'environnement local. On peut par exemple citer les différentes représentations de l'histidine, un acide aminé. Il est possible de représenter cette molécule sous trois formes différentes selon la position des hydrogènes. Or selon l'emplacement de la protonation, le carbone situé en bêta de l'azote n'a pas le même environnement chimique, donnant lieu à un type d'atome différent (voir Figure 2.1).

### 2.1.2 Les termes liés

Les termes liés assimilent les noyaux des atomes à des sphères liées entre elles par des ressorts, voir figure 2.2. On peut alors utiliser la théorie de l'oscillateur harmonique afin de modéliser ces interactions.

$$V_{liaison} = \sum_{liaisons} \frac{1}{2} k_l (l - l_0)^2 \quad (2.2)$$



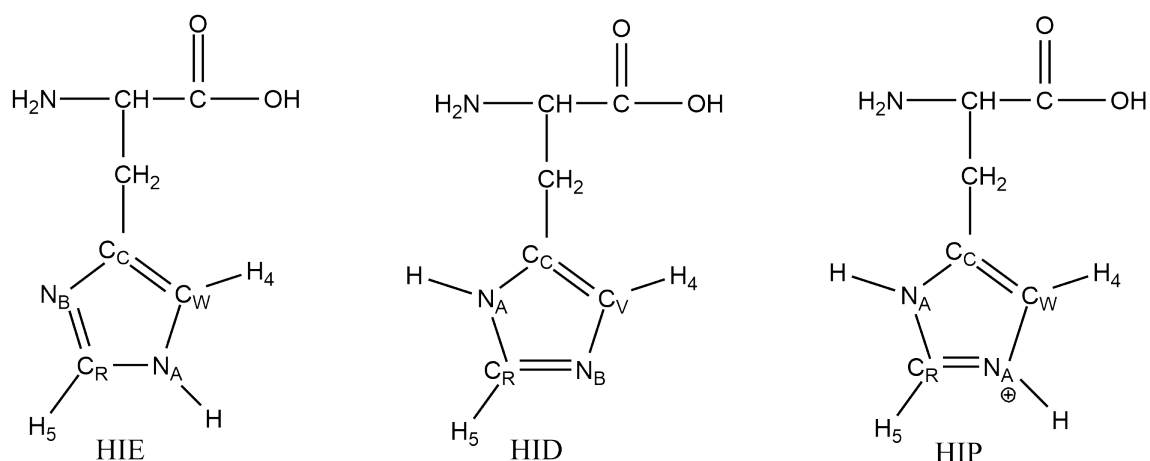


FIGURE 2.1 – Schéma représentant les 3 types d’histidines, Hie, His et Hip que l’on peut trouver dans un système biologique. Les lettres et chiffres situés en indices des symboles chimiques correspondent à la représentation de cet atome dans le champ de force ff99.

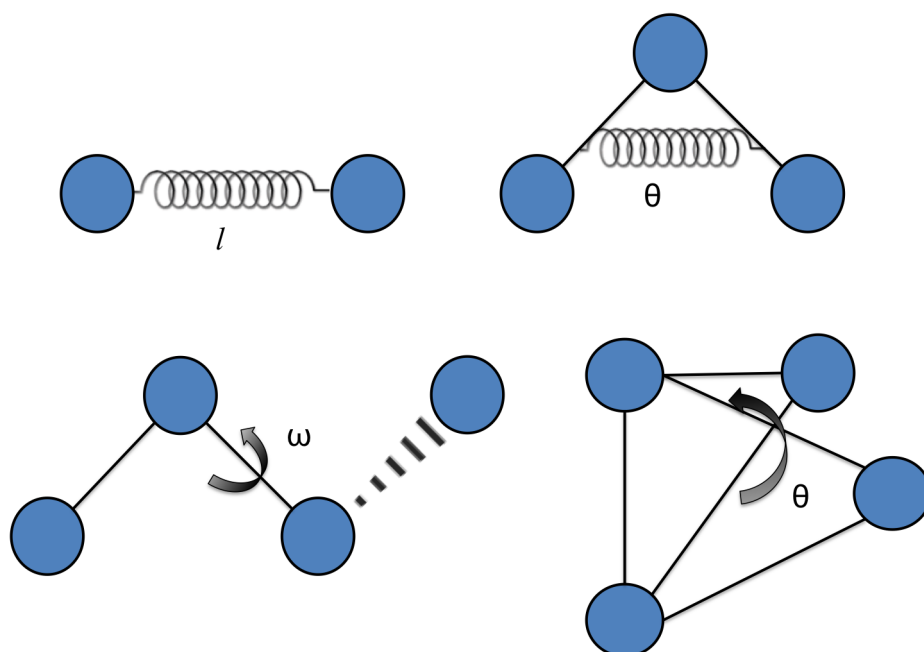


FIGURE 2.2 – Termes liés de la fonction d’énergie potentielle. En haut à gauche, terme de liaison de longueur  $l$ , en haut à droite, terme d’angle de valeur  $\theta$ , en bas à gauche terme d’angle dièdre de valeur  $\omega$  et à droite terme d’angle impropre  $\theta$ .

et

$$V_{angles} = \sum_{angles} \frac{1}{2} k_{\theta} (\theta - \theta_0)^2 \quad (2.3)$$

avec  $l_0(\theta_0)$  la position d'équilibre de la liaison (angle) considéré et  $k_l$  ( $k_\theta$ ) la constante de force associée au ressort.

L'approche harmonique est utilisée car elle reproduit bien le comportement d'une liaison covalente pour de faibles déformations. En cas de fortes variations il est nécessaire d'utiliser un potentiel de Morse plus coûteux en terme de temps de calcul. Pour les angles on peut ajouter des termes quadratiques afin d'améliorer la description.

Les angles impropres ou "hors du plan" sont décrits par :

$$V_{impropres} = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (2.4)$$

Ils permettent de contraindre les molécules à avoir une certaine géométrie. On peut citer par exemple l'utilisation de potentiel d'angles impropres afin d'obtenir des molécules planes (cycles aromatiques).

Enfin les angles dièdres sont représentés par :

$$V_{dièdres} = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \quad (2.5)$$

avec  $\omega$  l'angle de torsion,  $\gamma$  le facteur de phase,  $n$  la multiplicité et  $V_n$  "la hauteur de la barrière". Le nom de  $V_n$  est trompeur car les autres facteurs influent également sur l'énergie de la torsion. La valeur de  $n$  correspond au nombre de minimum de la fonction lors d'une rotation de  $360^\circ$ . Enfin  $\gamma$  est le facteur de phase.

Enfin, selon le champ de force utilisé, différents termes peuvent être ajoutés comme par exemple les termes croisés, qui permettent de coupler des termes liés en eux.

### 2.1.3 Les termes non-liés

On trouve parmi les termes non-liés les interactions électrostatiques et les interactions de type van der Waals (dispersion-répulsion).

L'interaction électrostatique est calculée selon un modèle de Coulomb :

$$V_{el} = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.6)$$

avec  $q$  la charge ponctuelle de l'atome/élément et  $\epsilon_0$  la permittivité du vide. Afin de pouvoir calculer cette interaction électrostatique il est nécessaire d'utiliser des charges ponctuelles situées au niveau du noyau atomique. Cette méthode reproduit bien les propriétés électrostatiques du système neutre ainsi que des espèces

chargées.

Les interactions de type van der Waals sont quand à elles décrites par le potentiel 12-6 de Lennard-Jones (LJ). Il peut être exprimé entre deux particules notées A et B comme :

$$V_{vdW} = 4\epsilon_{AB} \left[ \left( \frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left( \frac{\sigma_{AB}}{r_{AB}} \right)^6 \right] \quad (2.7)$$

avec  $\sigma_{AB}$  la distance pour laquelle le potentiel Lennard-Jones est nul et  $\epsilon_{AB}$  la profondeur du puits. Ces termes peuvent aussi s'exprimer entre deux particules identiques comme :

$$\sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \quad (2.8)$$

$$\epsilon_{AB} = (\epsilon_{AA}\epsilon_{BB})^{1/2} \quad (2.9)$$

C'est la relation de Lorentz-Berthelot. Le potentiel de LJ est caractérisé par une partie attractive en  $r^{-6}$  et une partie répulsive en  $r^{-12}$ . Il est intéressant de noter que la variation en  $r^{-6}$  est élevée à la même puissance que les énergies de dispersion des modèles théoriques (quantiques). Ce potentiel 12-6 de LJ est très largement utilisé pour les gros systèmes car rapidement calculable.

## 2.1.4 Des exemples de champ de force

### 2.1.4.1 Champ de force classique

Lors de cette thèse le CFC ff99 a été utilisé.

Ce CFC a été développé pour le logiciel AMBER (Assisted Model Building with Energy Refinement)<sup>55</sup> et s'est démocratisé dans plusieurs autres programmes dont TINKER<sup>56</sup> qui sera utilisé lors de calculs QM/MM. Il a été développé afin de traiter des molécules d'intérêt biologique comme des protéines ou l'ADN.

Il existe d'autres champs de force classique. On peut noter le champs de force charmm22 issu du logiciel CHARMM (Chemistry at HARvard Macromolecular Mechanics)<sup>57</sup> aussi utilisé pour la modélisation de protéines.

### 2.1.4.2 Champ de force polarisable

Les CFP sont de plus en plus communs et chaque logiciel en compte maintenant au moins un. Ils n'ont pas été utilisés durant cette thèse mais méritent d'être testés dans une futur proche. On peut citer AMOEBA (Atomic Multipole Optimised Energetics for Biomolecular Applications)<sup>58</sup> développé pour le logiciel TINKER mais également disponible dans la suite de logiciels AMBER.

## 2.1.5 Le traitement de l'eau

Lors de calculs basés sur des systèmes biologiques les molécules d'eau jouent un rôle important que l'on peut classer en deux effets principaux. L'eau a un rôle

de solvation dans les calculs de dynamique moléculaire, détaillé dans la partie dynamique moléculaire (2.2). Le deuxième rôle de l'eau est d'interagir avec le système étudié via par exemple des liaisons hydrogène.

Il existe différents moyens de traiter les molécules d'eau. Lors de cette thèse le modèle TIP3P<sup>59</sup> a été utilisé. En conséquence nous nous limiterons aux différents modèles issus de la famille TIPXP, mais de nombreuses autres paramétrisations sont possibles.

Le modèle pour l'eau appelé TIP3P est un modèle à 3 sites i.e. un site et une charge ponctuelle pour chaque atome. C'est le modèle le plus utilisé dans les systèmes protéiques pour sa simplicité et son faible coût computationnel. En outre de nombreux CF sont paramétrisés avec des molécules d'eau TIP3P.

Par la suite de nouveaux modèles on vu le jour comme les modèles TIP4P<sup>59</sup> et TIP5P<sup>60</sup> qui ajoutent un et deux sites, respectivement, au modèle précédent. Dans le modèle TIP4P, un pseudo-atome est ajouté au niveau de la bissectrice de l'angle  $\widehat{\text{H}-\text{O}-\text{H}}$ , ce pseudo-atome porte aussi la charge négative de la molécule à la place de l'oxygène. Cette paramétrisation permet d'améliorer la distribution électronique du modèle. Enfin dans le modèle TIP5P, deux pseudo-atomes sont ajoutés correspondant aux deux paires libres de l'oxygène. Les charges ponctuelles sont alors définies comme étant les deux pseudo-atomes et les hydrogènes. L'apport du modèle TIP5P est identifiable aux niveaux géométrique et thermodynamique. En effet, grâce à la présence des deux pseudo-atomes la molécule d'eau se rapproche d'une géométrie tétraédrique.

## 2.2 Dynamique moléculaire

### 2.2.1 Le principe

La dynamique moléculaire permet de suivre l'évolution d'un système dans le temps. On obtient ainsi une trajectoire, une collection de snapshots qui sont appelés *images* dans cette thèse. Chaque *image* représente le système à un temps  $t$ . Ainsi, contrairement à la mécanique moléculaire, on ne s'intéresse plus à un système statique mais à un système en mouvement.

Afin d'obtenir ces trajectoires on utilise l'équation de mouvement de Newton que l'on intègre pour chaque position des noyaux.

$$m_i \frac{d^2 x_i}{dt^2} = F_{x_i} \quad (2.10)$$

et

$$F_{x_i} = -\frac{d}{dx_i} V(x_i) \quad (2.11)$$

avec  $m_i$  la masse du noyau,  $x_i$  la coordonnée de déplacement,  $F_{x_i}$  la force appliquée sur le noyau dans la direction choisie et  $V(x_i)$  le potentiel dont est issu la force.

L'équation précédente ne s'applique que pour une seule coordonnée  $x_i$ , il est

bien sûr possible de l'appliquer à des systèmes en trois dimensions et de taille conséquente.

Afin de suivre l'évolution du système dans le temps, on a recours à différents algorithmes de récurrence. En pratique on définit un pas de temps  $\delta t$  correspondant à la fréquence d'intégration de l'équation du mouvement. La définition de ce pas de temps est très importante car il détermine quels phénomènes sont observés lors de la dynamique. Un pas de temps important (de l'ordre de la ps) permet de rendre compte du mouvement des boucles dans une protéine alors qu'un pas de temps petit (de l'ordre de la fs) permet d'étudier le mouvement des atomes ainsi que la vibration de leur liaison. La dynamique sera en conséquence plus ou moins longue et aura un temps de calcul plus ou moins élevé.

Pour chaque pas de temps l'intégration de l'équation du mouvement permet d'obtenir les nouvelles positions et vitesses des éléments du système qui seront utilisées dans le calcul pour le pas de temps suivant  $t+\delta t$ .

On commence par obtenir les conditions initiales du système c'est à dire la position et la vitesse des particules à un temps  $t_0$ . Les algorithmes les plus utilisés sont basés sur des développements de Taylor.

Par exemple l'algorithme de Verlet permet d'obtenir pour  $x(t+\delta t)$  :

$$x_i(t + \delta t) = 2x_i(t) - x_i(t - \delta t) + \frac{(\delta t^2)}{m_i} F_i(t) \quad (2.12)$$

On voit que dans cet algorithme, on ne fait pas intervenir la vitesse  $v_i$  (la valeur peut être accessible par un calcul supplémentaire).

D'autres algorithmes existent comme le Verlet aux vitesses :

$$x_i(t + \delta t) = x_i(t) + v_i(t)\delta t + \frac{(\delta t^2)}{2m_i} F_i(t) \quad (2.13)$$

$$v_i(t + \delta t) = v_i(t) + \frac{F(t) + F(t + \delta t)}{2m_i} \delta t \quad (2.14)$$

ou le Verlet "leap-frog" :

$$x_i(t + \delta t) = x_i(t) + v_i(t + \frac{\delta t}{2})\delta t \quad (2.15)$$

$$v_i(t + \frac{\delta t}{2}) = v_i(t - \frac{\delta t}{2}) + (\frac{\delta t}{m_i})F_i(t) \quad (2.16)$$

Dans ces deux algorithmes les vitesses sont explicitement utilisées.

## 2.2.2 Les paramètres de la simulation

### 2.2.2.1 Les ensembles thermodynamiques

Afin de pouvoir réaliser une trajectoire il est nécessaire de choisir un ensemble thermodynamique qui va fixer les conditions du calcul. Parmi ces ensembles on

retrouve l'ensemble isotherme-isobare (NPT) qui est le seul utilisé durant cette thèse. Dans cet ensemble, le nombre de particules, la pression et la température sont maintenus constants. Il existe d'autres ensembles comme par exemple :

- l'ensemble microcanonique (NVE) avec le nombre de particules, le volume et l'énergie constants.
- l'ensemble canonique (NVT) avec le nombre de particules, le volume et la pression constants.
- l'ensemble grand-canonique ( $TV\mu$ ) avec la température, le volume et le potentiel chimique constants.

Les paramètres définis comme constants dans les différents ensembles doivent le rester tout au long de la dynamique afin de garantir l'équilibre du système. Il existe des outils de contrôle qui ont pour rôle de s'acquitter de cette tâche.

Les thermostats permettent de conserver une température constante pendant la dynamique. Un des plus utilisés est le thermostat de Berendsen<sup>61</sup>. Dans ce thermostat, le système est couplé à un bain externe de température constante. A chaque pas de la trajectoire, une petite perturbation est appliquée sur les vitesses calculées afin de conserver une température constante. On peut citer les thermostats de Nosé-Hoover<sup>62</sup> et d'Andersen<sup>63</sup> comme autres exemples.

De la même manière les barostats garantissent une pression constante pendant la dynamique. Les dynamiques de Langevin font intervenir un thermostat de Berendsen couplée avec un piston. Le déplacement du piston permet de conserver une pression constante en faisant varier le volume du système.

### 2.2.2.2 Les conditions périodiques

Comme écrit précédemment le temps de la dynamique, le pas de temps choisi et la taille du système sont les principaux facteurs influant sur la longueur du calcul. En général, la taille du système reste le principal facteur limitant. En effet, afin de bien reproduire un système biologique par exemple il faudrait modéliser un système de taille macroscopique ce qui est pour l'instant impossible du point de vue calculatoire. La solution est alors de représenter notre système d'un point de vue microscopique et de répliquer ce système dans toutes les directions de l'espace. Ainsi afin de pouvoir reproduire un espace en trois dimensions, on représente les systèmes étudiés dans des boîtes polyédriques le plus souvent carrées. C'est ce qu'on appelle des conditions périodiques aux limites (PBC).

Ces conditions périodiques limitent également les effets de bords. Lors d'une dynamique rien n'empêche des particules de s'échapper de la boîte. On utilise une boîte dite initiale qui est ensuite reproduite périodiquement afin de conserver un nombre constant de particules dans chaque boîte. Un autre problème est le calcul des interactions électrostatiques entre particules. En effet, comme la boîte initiale est reproduite une infinité de fois, chaque particule interagit également avec les particules des

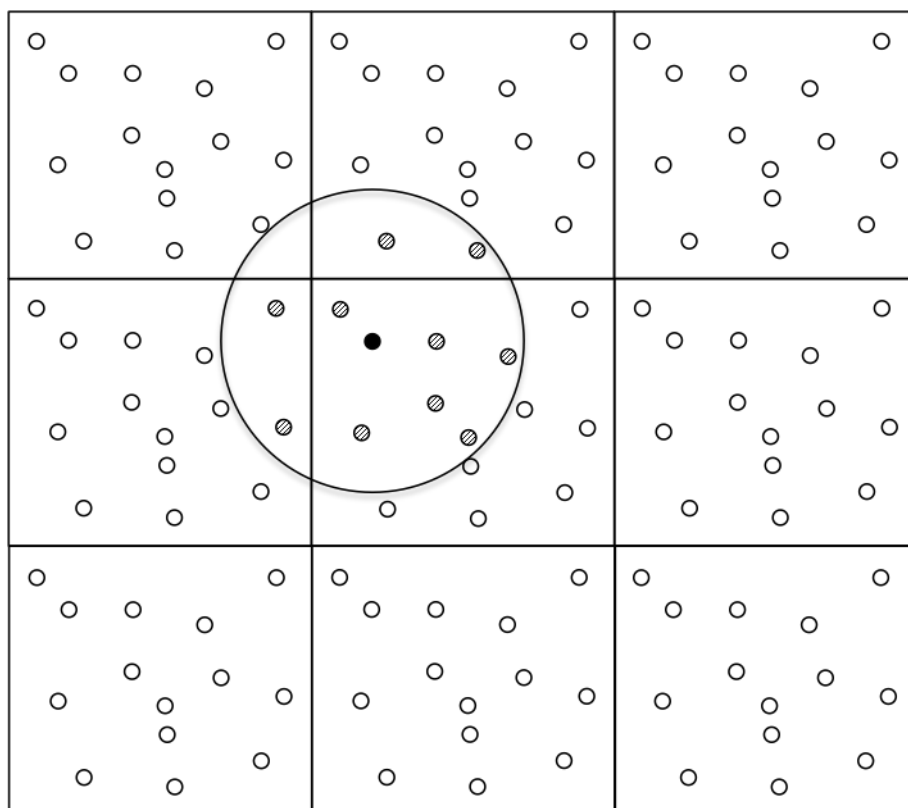


FIGURE 2.3 – Schéma d’une sphère avec un rayon de cutoff de la taille d’une demi-boîte et de centre la particule noire. Les particules grisées sont prises en compte dans le calcul des interactions.

boîtes voisines. Afin de limiter ces interactions chaque particule n’interagit avec les autres particules que lorsqu’elles sont incluses dans une sphère de rayon dit rayon cutoff. La convention est de prendre comme taille pour ce rayon la moitié de la taille de la boîte afin de ne pas compter plusieurs fois les mêmes interactions (voir Figure 2.3).





## Chapitre 3

# Méthode hybride QM/MM

L'étude des chapitres précédents a montré que les méthodes de mécanique quantique (QM) permettent une bonne représentation d'un système chimique et de ses propriétés comme par exemple les excitations électroniques. En revanche, l'utilisation des méthodes QM sont très consommatrices en temps de calcul et finalement les systèmes traités doivent être de taille limitée. Dans un deuxième temps, les méthodes de mécanique moléculaire (MM) peuvent traiter de larges systèmes possédant souvent plusieurs milliers de particules grâce à des paramètres listés dans des champs de force. En revanche, l'utilisation de champs de force ne permet pas d'obtenir des propriétés liées à la densité électronique.

Comme le nom de ce chapitre l'indique il est possible de combiner ces deux méthodes en une seule, hybride, appelée QM/MM (Quantum Mechanics/Molecular Mechanics). Cette méthode traite un système complexe en le séparant en plusieurs sous-espaces qui sont définis selon une approche QM ou une approche MM. Un exemple de cette méthode correspond au calcul des excitations énergétiques d'un substrat présent à l'intérieur d'une protéine ou de toutes molécules biologiques. Une autre application possible est la modélisation d'une substance liée à un catalyseur comme des billes de silice<sup>64</sup>.

Le premier couplage de type QM/MM a été effectué par deux chercheurs, Levitt et Warshel en 1976<sup>65</sup> sur des réactions enzymatiques. L'apport de cette méthode montre notamment l'importance des interactions électrostatiques dans la stabilisation de l'ion carbonium chez le lysozyme. Le développement de cette méthode multi-échelle a donné lieu à la remise d'un prix nobel de chimie en 2013 pour Levitt, Warshel et Karplus<sup>1</sup> La suite de ce chapitre regroupe les points essentiels d'un calcul QM/MM.

---

1. "The Nobel Prize in Chemistry 2013". Nobelprize.org. Nobel Media AB 2014. Web. 22 Apr 2017. <[http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/2013/](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/)>

### 3.1 Le principe

Le principe de la méthode QM/MM est de calculer l'énergie totale du système noté  $\mathbb{C}$  à partir des différents sous-systèmes présents dans  $\mathbb{C}$  que sont :

- $\mathbb{S}$ , la substance d'intérêt.
- $\mathbb{E}$ , l'environnement.

En outre, la dénomination du terme QM/MM à une importance. Dans le cas général,

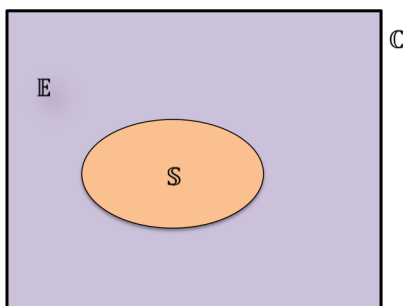


FIGURE 3.1 – Schéma d'une partition QM/MM.

on utilise la notation QM/MM avec un slash entre QM et MM. Si  $\mathbb{S}$  et  $\mathbb{E}$  sont liés de façon covalente on écrit QM-MM et QM:MM si ce sont deux sous-systèmes distincts<sup>2</sup>.

## 3.2 Les schémas QM/MM

### 3.2.1 Le schéma soustractif

Dans le schéma dit soustractif, l'énergie du système  $\mathbb{C}$  se fait par l'enchaînement de plusieurs calculs. Le premier détermine l'énergie de  $\mathbb{C}$  au niveau MM puis le second l'énergie de  $\mathbb{S}$  au niveau QM. Enfin comme l'énergie du sous-système  $\mathbb{S}$  est comptabilisée deux fois (une fois en QM, une fois en MM), on soustrait l'énergie de  $\mathbb{S}$  au niveau MM. Le schéma soustractif a été le premier utilisé de façon chronologique et se retrouve dans le programme ONIOM (Our own N-layered Integrated molecular Orbital and molecular Mechanics)<sup>66</sup> présent dans le logiciel Gaussian<sup>67</sup>. L'Hamiltonien total du système s'écrit alors :

$$H_{tot} = H_{MM}^{\mathbb{C}} + H_{QM}^{\mathbb{S}} - H_{MM}^{\mathbb{S}} \quad (3.1)$$

### 3.2.2 Le schéma additif

Dans le schéma additif, le système  $\mathbb{C}$  est divisé en deux sous-systèmes  $\mathbb{S}$  et  $\mathbb{E}$  qui sont traités à deux niveaux de théorie différents. Ainsi l'énergie de  $\mathbb{C}$  est la somme de  $\mathbb{S}$  traité au niveau QM, de  $\mathbb{E}$  traité au niveau MM mais également de l'interaction exercée entre les sous-systèmes  $\mathbb{S}$  et  $\mathbb{E}$ . Ce nouveau terme est essentiel et de doit pas

2. N. Ferré, X. Assfeld, l'actualité chimique - février-mars 2014 - n° 382-383

être négligé. Le schéma additif présente un Hamiltonien total du système pouvant être écrit comme :

$$H_{tot} = H_{MM}^E + H_{QM}^S + H_{QM-MM}^C \quad (3.2)$$

### 3.2.3 Les interactions QM/MM

Le terme  $H_{QM-MM}$ , défini dans l'équation 3.2 est un nouveau terme représentant les interactions ayant lieu entre la partie QM et la partie MM. Il peut être décomposé en plusieurs parties, chacune représentant un type d'interaction entre les deux sous-systèmes. Une représentation conventionnelle de ce terme peut-être exprimée comme :

$$H_{QM-MM} = H_{QM-MM}^{nuc} + H_{QM-MM}^{elec} + H_{QM-MM}^{vdW} \quad (3.3)$$

avec  $H_{QM-MM}^{nuc}$  l'interaction entre les noyaux de la partie QM et les charges ponctuelles de la partie MM,  $H_{QM-MM}^{vdW}$  un terme de van der Waals entre les deux sous-systèmes et  $H_{QM-MM}^{elec}$  l'interaction entre les électrons de la partie QM et les charges ponctuelles de la partie MM.

Il est possible de définir plusieurs modèles permettant de représenter les interactions QM/MM. Chaque modèle se différencie selon la prise en compte ou non des d'interactions précédemment citées mais également par l'expression mathématique utilisée pour représenter ces dernières.

#### 3.2.3.1 Le mechanical embedding (ME)

Le mechanical embedding (ME) est le modèle le plus simple afin de représenter les interactions entre la partie QM et la partie MM. En effet, ce modèle n'est défini que par les deux termes  $H_{QM-MM}^{nuc}$  et  $H_{QM-MM}^{vdW}$  et permet de représenter l'influence géométrique de la partie MM sur la partie QM. Dans ce modèle, les interactions électrostatiques entre la partie QM et la partie MM sont assimilées à des interactions électrostatiques de type MM-MM. Les atomes de la partie QM sont alors décrit par des charges ponctuelles et par des rayons de vdW extrait du champs de force utilisé.  $H_{QM-MM}^{vdW}$  peut être défini comme dans l'équation 2.7 entre deux atomes A et B.  $H_{QM-MM}^{nuc}$  est représenté par :

$$H_{QM-MM}^{nuc} = \sum_A^M \sum_{B \in MM} \frac{Z_A q_B}{|R_B - R_A|} \quad (3.4)$$

Il est important de vérifier quand on utilise ce modèle que les propriétés étudiées ne sont pas influencées par l'environnement électrostatique autour de la partie QM. En revanche, cet embedding est peu coûteux à incorporer dans le calcul QM/MM.

#### 3.2.3.2 L'electrostatical embedding (EE)

L'electrostatical embedding ajoute le terme  $H_{QM-MM}^{elec}$  dans l'Hamiltonien du QM-MM en plus des termes du mechanical embedding. On tient alors compte des interactions électrostatiques entre la densité de charges des atomes de la partie QM et

les charges ponctuelles des atomes MM. De cette façon le terme  $H_{QM-MM}^{elec}$  va modifier l'opérateur de Fock et polariser la fonction d'onde de la partie QM en ajoutant le terme :

$$H_{QM-MM}^{elec} = \sum_i^N \sum_{B \in MM} - \frac{q_B}{|r_i - R_B|} \quad (3.5)$$

avec N le nombre d'électrons du système.

L'electrostatical embedding améliore le traitement des interactions entre la partie QM et la partie MM. Ainsi, les charges de la partie QM deviennent polarisées par l'environnement proche ce qui influe par exemple sur la localisation de la HOMO et de la LUMO. Cet embedding est également indispensable afin de traiter les liaisons hydrogène entre  $\mathbb{S}$  et  $\mathbb{E}$ . L'electrostatical embedding est actuellement le plus utilisé mais présente néanmoins quelques désavantages. L'énergie électrostatique calculée est non symétrique, si on échange la méthode de calcul de la partie QM et de la partie MM l'énergie électrostatique obtenue est de l'ordre de plus ou moins 10% de la valeur initiale. De plus, le coût computationnel augmente avec la taille du sous-système MM, on utilise alors un cutoff qui délimite la taille du  $\mathbb{E}$  calculée par l'embedding. C'est dans cette optique qu'a été créée la méthode dite ESPF<sup>68</sup> (ElectroStatic Potential Fitted). Les charges ponctuelles de la partie MM sont modifiées afin d'être exprimées comme des multipôles disposées sur une grille autour de la partie QM. On peut alors calculer pour chaque point de cette grille l'influence et le potentiel exercé sur la densité électronique de la partie QM. L'énergie et la fonction d'onde de la partie QM sont alors optimisées en tenant compte du potentiel électrostatique exercé par chaque multipôle présent sur la grille.

### 3.2.3.3 Le polarisable embedding (PE)

L'apport d'un polarisable embedding permet de prendre en compte la réponse de la partie MM lors de l'étude des états excités. En effet, l'excitation d'un système QM provoque un réarrangement de la densité électronique du système. Cette modification de la densité électronique va affecter la partie MM qui en réponse modifie les charges ponctuelles des atomes de la partie MM. Ce polarisable embedding entraîne l'apparition de charges ponctuelles "flottantes" réagissant aux fluctuations de la densité électronique de la partie QM qui par la suite sont elles mêmes-modifiées suite à l'action de l'electrostatical embedding.

L'ajout de ce embedding permet d'obtenir une modélisation très complète du système. Il est en revanche très peu utilisé aujourd'hui car difficile à mettre en place et très couteux du point de vue calculatoire.

### 3.2.4 Définir la partie QM/MM

Une autre difficulté des calculs QM/MM est la définition de la coupure entre la partie MM et la partie QM. Lorsque les deux sous-systèmes ne sont pas liés par une liaison covalente on n'utilise qu'un embedding pour représenter le système.

En revanche, si on observe la présence d'une liaison covalente entre les deux sous-systèmes on doit utiliser d'autres algorithmes afin de bien représenter la coupure. Dans cette thèse le schéma link-atom est utilisé, une présentation rapide des autres schémas est également développé par la suite (voir Figure 3.2).

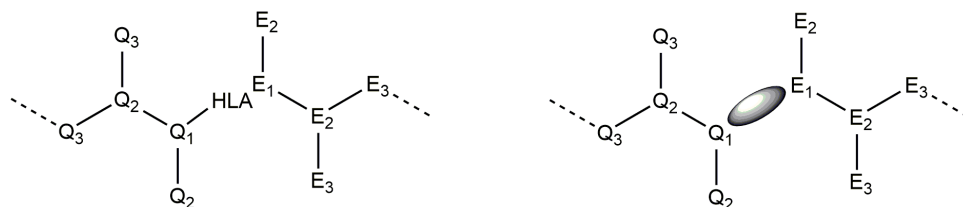


FIGURE 3.2 – Représentation des deux différentes coupures entre la partie QM et la partie MM. A gauche, l'utilisation d'un link-atom noté HLA. A droite l'utilisation des orbitales localisées. Les atomes notés  $Q_x$  correspondent aux atomes de la partie QM et les atomes notés  $E_x$  à ceux de la partie MM. La coupure QM/MM s'effectue entre l'atome  $Q_1$  et l'atome  $E_1$ .

### 3.2.4.1 Le link-atom (LA)

La méthode de l'atome de liaison ou link-atom (LA) est une méthode dite de saturation de liaison. C'est l'une des méthodes les plus anciennes et les plus utilisées. Elle est, en outre, relativement simple à visualiser. Le principe consiste à couper puis à saturer la liaison qui relie la partie QM à la partie MM. Ainsi, on va ajouter un atome, le plus souvent un hydrogène, afin de saturer l'atome de la partie QM affecté. L'utilisation du link-atom entraîne cependant quelques problèmes. En premier lieu, comme les atomes d'hydrogène ajoutés sont proches de la partie MM, le calcul de l'énergie électrostatique peut amener à une surpolarisation de la fonction d'onde. Une solution à ce problème est de fixer artificiellement la charge de l'atome MM de la liaison cassée à zéro. Le deuxième problème concerne la nature de la liaison à casser. En effet, le système est mal reproduit si on casse une liaison polaire de type C-O ou une liaison C-C doublement conjuguée. C'est pour cela que dans le cas du link-atom on casse le plus souvent des liaisons de type  $C_{sp^3} - C_{sp^3}$ .

### 3.2.4.2 Les orbitales localisées

Dans cette méthode, la liaison d'intérêt entre les atomes frontières de la partie QM et de la partie MM est représentée à partir d'orbitales hybrides localisées.

La méthode des orbitales liantes strictement localisée (LSBO) est une méthode développée à Nancy conjointement avec la méthode LSCF<sup>69</sup> qui permet d'utiliser une méthodologie de type champ autocohérent afin de traiter les systèmes comportant des orbitales liantes strictement localisées. Le principe des LSBO est d'introduire une orbitale entre les deux atomes frontières, représentant la liaison d'intérêt entre la partie QM et la partie MM. L'atome frontière de la partie MM fait alors partie du système QM et possède à la fois des paramètres de champ de force et une orbitale hybride. Enfin, en utilisant le principe de transférabilité on paramétrise

la nouvelle orbitale selon une orbitale possédant un comportement proche. Par exemple, une liaison  $C_{sp^3} - C_{sp^3}$  pourra être remplacée par une orbitale de la liaison C-C d'une molécule d'éthane. Cette méthode permet une meilleure description de la liaison frontière QM/MM notamment en cas de liaison polarisée.

Il existe d'autres méthodes utilisant des orbitales localisées comme la méthode des orbitales gelées<sup>70</sup> qui nécessite une paramétrisation plus importante de la zone frontière ou la méthode des orbitales hybrides généralisée<sup>71</sup> qui utilise des orbitales  $sp^3$  afin de représenter l'atome frontière de la partie MM. Parmi ces 4 orbitales, l'orbitale hybride participant à la liaison sera incluse dans le calcul de la partie QM et les autres resteront gelées.

# Chapitre 4

## Modéliser l'environnement

Plusieurs approximations permettent de décrire l'environnement. L'approximation la plus simple est de négliger l'environnement, on parle alors de calcul "in vacuum". Ces calculs sont difficiles à exploiter car les valeurs expérimentales issues de manipulations sous vide sont rares, difficiles et moins pertinentes que des résultats en solution. Il existe néanmoins des expériences réalisées dans ces conditions de presque vide et possédant un intérêt scientifique concret<sup>72</sup>. Les calculs in vacuum servent également à comparer les valeurs obtenues avec des résultats prenant en compte les perturbations extérieures afin de quantifier l'effet de l'environnement.

### 4.1 Le modèle implicite

Le modèle implicite décrit le solvant comme un milieu homogène défini par une constante diélectrique. Cette constante diélectrique possède une valeur fixe pour chaque solvant usuel. Dans ce modèle il n'y a donc aucune molécule de solvant présente dans le système.

Les modèles utilisant un milieu homogène sont regroupés dans la catégorie des champs de réaction auto-cohérent (SCRf), parmi différents modèles on ne présente ici que le modèle dit Polarisable Continuum Model (PCM)<sup>73,74</sup>.

Dans ce modèle l'énergie libre de solvation assimilable à l'ensemble des interactions moyennes entre le soluté et le solvant peut-être écrite comme :

$$\Delta G = \Delta G_{el} + \Delta G_{cav} + \Delta G_{rep} + \Delta G_{dis} \quad (4.1)$$

avec  $\Delta G_{el}$  la contribution électrostatique,  $\Delta G_{cav}$  l'énergie de cavitation,  $\Delta G_{rep}$  l'énergie de répulsion et  $\Delta G_{dis}$  l'énergie de dispersion.

Dans cette équation, l'énergie électrostatique correspond à l'interaction entre les charges du soluté et les charges créées en réponse par le continuum à la surface cavité-soluté. L'énergie de cavitation correspond à l'énergie nécessaire à la création d'une cavité dans le continuum. L'énergie de répulsion et l'énergie de dispersion

correspondent aux interactions de dispersion et de répulsion entre le soluté à l'interface de la cavité et plus particulièrement à des forces de dispersion de type London et de répulsion entre nuages électroniques.

L'utilisation de ce modèle nécessite de savoir quelle forme donner à la cavité abritant le soluté. Cette cavité doit épouser au mieux la surface du soluté et doit pouvoir être représentée par un modèle simple. Les choix courants sont de représenter les atomes du soluté par des sphères centrées sur le noyau des atomes et possédant un rayon de type van-der-Waals<sup>75</sup>. Le solvant est alors décrit comme une sphère et va "rouler" sur la surface créée par le soluté. On définit alors deux surfaces, la Surface Excluant le Solvant (SES) et la Surface Accessible au Solvant (SAS). Dans la première, la surface correspond au point de contact entre les sphères du soluté et la sphère du solvant. Dans la deuxième, la surface correspond à la trajectoire du centre de la sphère du solvant autour des sphères du soluté (voir Figure 4.1). Enfin il existe une autre méthode dite la Surface de van-der-Waals (vdWS) permettant d'obtenir une surface lisse et continue<sup>76,77</sup>. Cette méthode est aujourd'hui la méthode par défaut utilisé dans le logiciel gaussian.

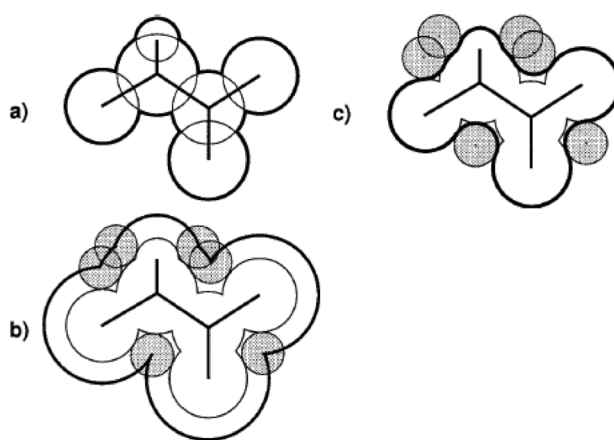


FIGURE 4.1 – Définition des différentes surfaces (en trait épais). a) Surface de van-der-Waals (vdWS) b) Surface Accessible au solvant (SAS) c) Surface Excluant le Solvant (SES). Appropriate credit for the requested material should be given as follows : "Adapted with permission from (Darrin M. York and Martin Karplus. A smooth solvation potential based on the conductor-like screening model. The Journal of Physical Chemistry A, 103(50) :11060-11079, 1999) Copyright (2017) American Chemical Society

".



## 4.2 Le modèle explicite

On peut aussi traiter l'environnement de façon explicite, c'est à dire que les molécules du solvant sont décrites de manière atomiques. Ce modèle permet une très bonne description des interactions soluté-solvant. Il permet notamment la prise en compte des effets liés à la présence de liaisons hydrogène avec le milieu. De plus comme les molécules de solvants doivent être explicitement positionnées, la polarité de la cavité autour du soluté est plus réaliste.

On a déjà parlé de la méthodologie QM:MM dans le chapitre 3. L'environnement est représenté par un champ de force plus ou moins détaillé et les interactions entre le soluté et le solvant, respectivement le QM et le MM, par un modèle mathématique. Cette méthodologie est souvent utilisée en présence de systèmes biologiques. Dans la méthode QM:MM l'environnement reste néanmoins décrit par un ensemble de paramètres plus ou moins précis issus de champ de force.

Des méthodes de type QM:QM' permettent la meilleure description du solvant car il alors décrit de façon quantique. Cependant seul un nombre réduit de molécules peuvent être considérées afin de garder des temps de calcul raisonnables. Une solution intermédiaire sont les méthodes de type QM:QM':MM ou QM:QM':PCM<sup>78,79</sup>. Les molécules situées autour du soluté sont traitées de façon quantique et les restantes de façon implicite ou par mécanique moléculaire. Ainsi on conserve la majorité des interactions soluté-solvant à faible distance tout en conservant des temps de calcul raisonnables.

## 4.3 Calculer l'influence de l'environnement

Dans cette thèse le terme  $T_e$  réfère à la transition électronique entre le 1<sup>er</sup> état singulet excité et l'état fondamental soit une émission de type fluorescence. On note  $T_e^{vac}$  la transition électronique entre S1 et S0 calculée dans le vide sur le substrat extrait dans la conformation optimisée dans l'état S1 la protéine. La comparaison de  $T_e$  et de  $T_e^{vac}$  permet de quantifier l'effet de l'interaction électrostatique de la protéine sur un substrat. Pour cela, il suffit d'enlever l'environnement extérieur au substrat, puis de réaliser un calcul "single point" du  $T_e$  S1->S0 en gelant la partie QM dans sa géométrie stable obtenue pour le 1<sup>er</sup> état excité.

## CHAPITRE 4 : MODÉLISER L'ENVIRONNEMENT

# Chapitre 5

## Modéliser des chemins réactionnels

L'étude des réactions chimiques en chimie computationnelle permet d'apporter des réponses aussi bien sur la faisabilité d'un mécanisme que sur la compréhension de ce dernier. Dans le domaine de la photochimie la prise en compte des états excités dans le mécanisme réactionnel apporte une difficulté supplémentaire compte tenu des croisements possibles entre ces différents états. On présentera dans un premier temps le déroulement d'un mécanisme réactionnel classique avant de s'intéresser aux mécanismes complexes de la réactivité dans les états excités.

### 5.1 Un mécanisme réactionnel simple

Une réaction chimique fait intervenir un ou plusieurs réactifs R pour donner un ou plusieurs produits P. Prenons par exemple une réaction simple de type :



Avec A, B et C trois molécules et X-Y représente une liaison covalente entre X et Y.

Une réaction chimique se caractérise par la variation de la distance B-C ( $R_{BC}$ ) et A-B ( $R_{AB}$ ) en fonction de l'énergie totale de l'ensemble du système. Ainsi, avant la collision entre A et B-C la distance  $R_{AB}$  est infinie et la distance  $R_{BC}$  est à sa position d'équilibre. A la fin de la réaction  $R_{AB}$  est à l'équilibre et  $R_{BC}$  infinie. La surface d'énergie potentielle (PES) de la réaction chimique permet de représenter l'énergie du système en fonction des coordonnées d'intérêts de la réaction. Au début de la réaction les réactifs correspondent à des minima sur la surface d'énergie potentielle. Ces minima représentent un site de stabilité pour la molécule. Lors de la réaction des liaisons vont se casser et d'autres se former conduisant à une augmentation de l'énergie du système. Celle-ci atteint son maximum en un point appelé état de transition (TS) avant de diminuer pour tomber dans un nouveau puits de potentiel de type minima, correspondant au produit de la réaction.

L'étude de l'état de transition est primordiale afin de décrire un mécanisme réactionnel. En effet, le calcul de l'énergie du TS permet de déterminer la barrière énergétique entre le réactif et le TS, rendant compte de la faisabilité de la réaction. La

géométrie du TS correspond à un point de jonction entre la géométrie du réactif et du produit. Cette dernière fournit des informations sur la stéréochimie de la réaction et l'influence des différents groupements de la molécule. Cet état est caractérisé par une fréquence de vibration imaginaire, représentant le vecteur de transition permettant de relier les réactifs aux produits. Le chemin réactionnel le plus court entre les réactifs et les produits et passant par le TS est calculé suivant le chemin minimum d'énergie (MEP) sur une surface d'énergie potentielle en  $3N-6$  dimension ( $N$  étant le nombre de noyaux dans le système).

## 5.2 La réactivité dans les états excités

Les mécanismes réactionnels faisant intervenir un ou plusieurs états excités sont communs dans la chimie luminescente. Les réactions de chimiluminescence et bioluminescence sont par définition des réactions faisant intervenir les états excités. En effet, l'état excité émetteur est obtenu suite à une réaction chimique. De nombreuses molécules photoluminescentes émettent suite à une cascade de réactions chimiques dans les états excités, le passage d'un état singulet à un état triplet étant un exemple concret (voir Figure 5.1).

La difficulté de la réactivité dans les états excités est double. Dans un premier temps, il faut considérer une surface d'énergie potentielle pour chaque état électronique inclus dans le mécanisme réactionnel. Il est alors beaucoup plus compliqué et long d'obtenir un scan détaillé de toutes ces surfaces, réduisant la précision quand au mécanisme réactionnel. Dans un second temps comme on tient compte de plusieurs états électroniques il existe des intersections permettant de passer de l'un à l'autre. Outre le fait que ces intersections sont très difficiles à représenter et à étudier, connaître les probabilités de passages d'un état à l'autre est un facteur clef afin de comprendre ces mécanismes. Certaines réactions photochimiques possèdent 4 ou 5 intersections donnant des produits finaux complètement différents; estimer le chemin réactionnel emprunté devient alors une tâche ardue.

Ainsi deux états électroniques peuvent être dégénérés à certains endroits de la surface d'énergie potentielle. On appelle une intersection conique ce croisement entre deux états de même spin. Parmi les mécanismes photoluminescents, la présence d'une intersection conique accroît la possibilité d'une transition non radiative et donc une perte d'émission de lumière. Cependant le passage d'un état excité par une intersection conique peut conduire à un nouveau produit de réaction.

Dans les réactions chimiluminescentes et bioluminescentes, la présence d'une intersection conique permet d'atteindre le premier état excité et permet au système d'accéder à l'état électronique responsable de l'émission de lumière.

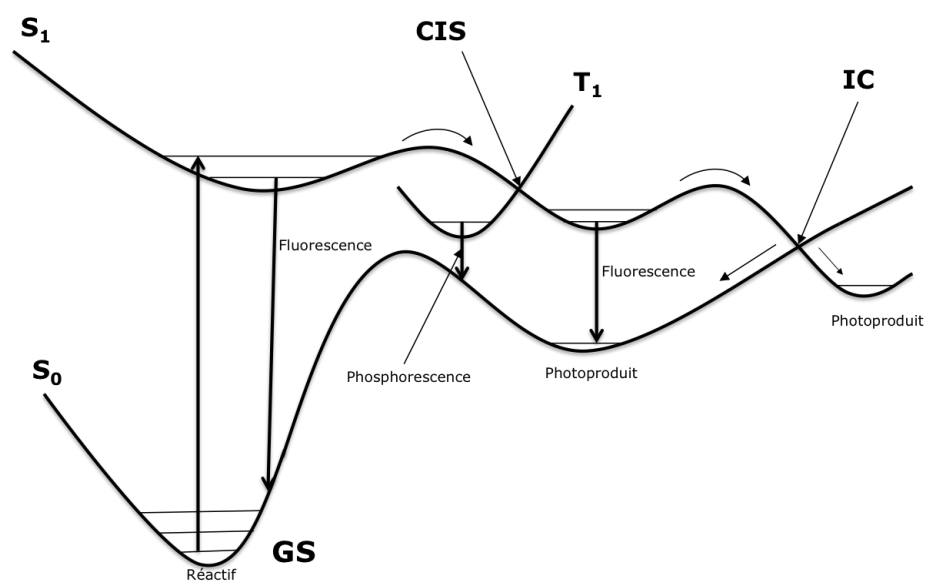


FIGURE 5.1 – Schéma des principales réactions photochimiques sur une surface d'énergie potentielle.



## **Deuxième partie**

### **La luminescence**





# Chapitre 6

## Les différentes sortes de luminescence

La luminescence est un mécanisme physique donnant lieu à une émission de lumière. Cette émission correspond à la perte d'énergie ou désexcitation d'un état électronique dit excité vers un état de plus basse énergie. L'énergie est libérée sous forme d'un photon lumineux avec une longueur d'onde d'émission correspondant à la différence d'énergie entre les deux états électroniques considérés.

La luminescence s'oppose à l'incandescence où l'émission de lumière provient du chauffage du système.

La première étape du mécanisme de la luminescence permet l'excitation du système c'est à dire que le système gagne suffisamment d'énergie afin d'atteindre un état dit excité. Il existe différentes façons d'exciter le système selon la source initiale d'énergie. Ainsi on peut parler de :

- Mécanoluminescence : l'état excité est obtenu par action mécanique sur un solide. On peut citer comme exemple la piézoluminescence ou la sonoluminescence. Ce type de luminescence peut également se retrouver chez des espèces vivantes comme la crevette pistolet. Cette crevette chasse on faisant imploser des bulles formées par la fermeture rapide de sa pince. L'effondrement de cette bulle, à des conditions de pression et de température extrêmes produit des flashes de lumière peu intense et invisible à l'œil nu. En conséquence ce type de mécanoluminescence à été nommé shrimpoluminescence<sup>80</sup>.

- Electroluminescence : l'état excité est obtenu suite au passage d'un courant électrique ou en la présence d'un champ électrique fort. L'électroluminescence est fortement implantée dans notre vie quotidienne, les LEDs (diodes électroluminescentes)<sup>81</sup> présentes dans les appareils électroniques utilisent ce type de mécanisme ainsi que certains types de laser ou de solides semi-conducteurs.

- Photoluminescence : l'état excité est obtenu suite à l'absorption d'un photon. Il existe deux types de photoluminescence, la fluorescence et la phosphorescence. La fluorescence est très utilisée dans le domaine médical comme dans le cas des fluorophores. La phosphorescence est quand à elle utilisée comme émission de

lumière retardée dans des peintures ou objets. Les OLED (diode électroluminescente organique)<sup>82</sup> émettent le plus souvent par phosphorescence. Ces deux types de photoluminescences seront étudiés en détails ci-après.

- Chimiluminescence : l'état excité est obtenu suite à une réaction chimique. Il existe peu de molécules dites chimiluminescentes, notons cependant comme exemple le luminol (dans les tubes lumineux) et les 2-coumaranones.

- Bioluminescence : l'état excité est obtenu suite à une réaction chimique dans un site catalytique d'une espèce vivante. De nombreux organismes vivants utilisent la bioluminescence comme moyen de défense, de communication, attraction... La bioluminescence est également utilisée dans le domaine biomédical. La bioluminescence sera étudiée plus en détails dans une section suivante.

Dans le reste de ce chapitre, on va approfondir les mécanismes de luminescence utilisés dans cette thèse. On détaillera aussi certains outils employés en chimie théorique afin de modéliser ces luminescences.

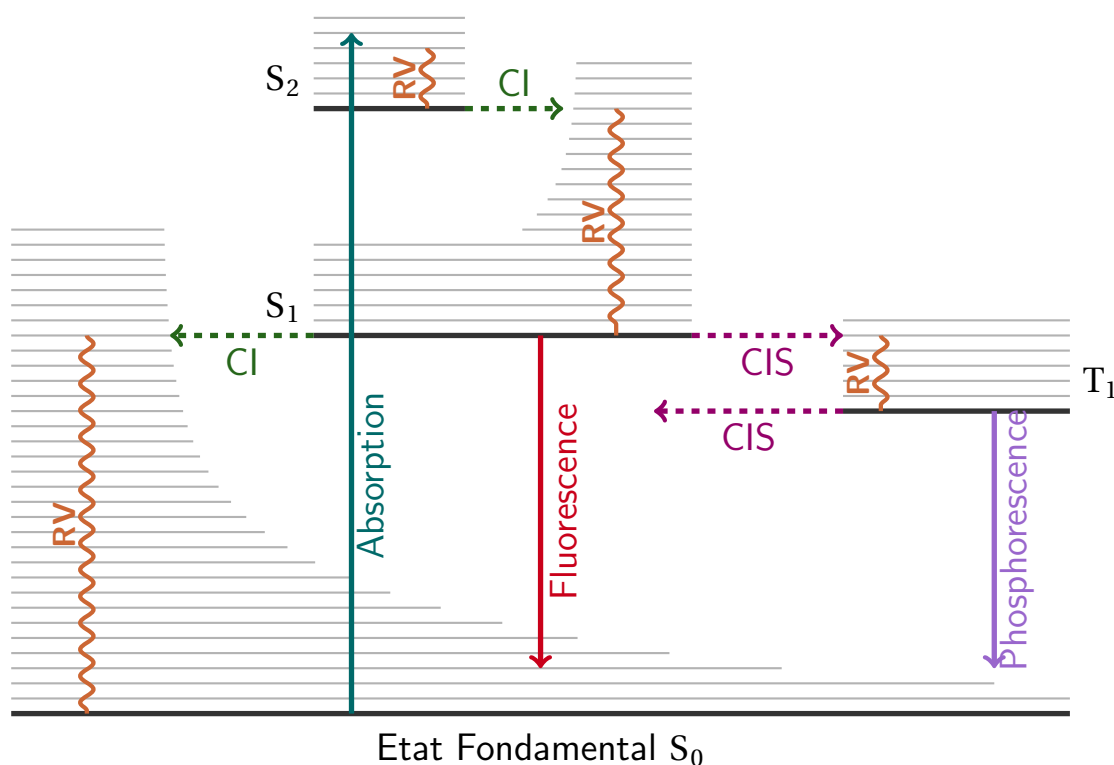
## 6.1 La photoluminescence

Comme indiqué dans la partie précédente la photoluminescence est la branche de la luminescence où l'obtention de l'état excité se fait par absorption d'un photon, c'est à dire d'un rayonnement lumineux possédant une énergie définie. Lors du processus d'émission, la perte d'énergie peut se faire selon deux processus différents, un processus radiatif et un processus non-radiatif. Le diagramme de Perrin-Jablonski permet d'avoir un point de vue global sur ces deux processus.

Dans ce diagramme (voir figure 6.1<sup>1</sup>), l'état fondamental du système est noté  $S_0$ , les états excités singulets  $S_n$  et les états excités triplet  $T_n$ . En outre, chaque état électronique du système est composé de niveaux vibrationnels qui participent aux processus radiatifs. Lors de l'absorption d'un photon d'énergie  $h\nu$ , un électron va être promu vers un des niveaux vibrationnels de l'un des états excités singulets. L'énergie du photon absorbée doit être au minimum égale à la différence d'énergie entre l'état fondamental et l'état excité. On dit d'une transition qu'elle est verticale si celle est suffisamment rapide pour qu'il n'y est pas de réorganisation des noyaux du système excité. On considère également que lors de cette transition verticale il n'y a pas de réorganisation de l'environnement lié au système.

Puis, dans un second temps, le système excité va perdre de l'énergie de telle sorte à repeupler l'état fondamental  $S_0$ . L'énergie peut être perdue sous forme de chaleur comme dans le cas de la relaxation vibrationnelle ou de la conversion interne (CI). Ces deux relaxations appartiennent aux processus non-radiatif. La relaxation vibrationnelle permet au système de se relaxer afin de retourner à l'état vibrationnel le

1. 1. Figure adaptée de celle créée par Germain Salvato-Vallverdu dont le code source est disponible à l'adresse <http://www.texample.net/tikz/examples/the-perrin-jablonski-diagram/>



### Légende




-  Conversion Interne,  $S_i \rightarrow S_j$  transition non radiative.
-  Croisement Inter-système,  $S_i \rightarrow T_j$  transition non radiative.
-  Relaxation Vibrationnelle.

Figure 6.1 – Le diagramme de Perrin-Jablonski.

plus bas en énergie de l'état électronique considéré. La conversion interne permet la transition entre deux états électroniques de même spin vers l'état le plus bas en énergie.

Le processus radiatif se traduit par l'émission d'un photon lors de la désexcitation. Si l'émission de ce photon a lieu suite à une perte d'énergie entre deux états singulets on parle de fluorescence. L'émission peut également avoir lieu entre un état triplet et l'état fondamental singulet, c'est la phosphorescence. Cette dernière est possible grâce au croisement inter-système (CIS) permettant la transition d'un électron entre deux états de multiplicité de spins différents (dans notre cas singulet et triplet). Comme dans le cas de l'absorption, la fluorescence et la phosphorescence sont assimilées à des transitions verticales.

Outre le spin de l'état électrique de départ, il est possible de discriminer la fluorescence de la phosphorescence selon le temps de vie de l'état excité. En effet, la fluorescence est un phénomène rapide de l'ordre de la nanoseconde alors que le temps de vie de la phosphorescence est de l'ordre de la seconde.

Dans la majorité des molécules organiques, la fluorescence et la phosphorescence ont lieu entre le 1<sup>er</sup> état excité, respectivement S1 ou T1 et l'état fondamental S0. Il est également important de noter que ces deux processus radiatifs sont en concurrence avec la conversion interne pour la transition entre S1 et S0. On définit alors la notion de rendement quantique, noté  $\Phi$ , correspondant au nombre de photons émis sur le nombre de photons absorbés par le système. Il est alors possible de rendre compte de l'efficacité des processus radiatifs en calculant ce rendement quantique.

Les molécules pouvant émettre des photons par photoluminescence sont appelées chromophores. Dans les chromophores on s'intéresse plus particulièrement à deux orbitales moléculaires, l'orbitale moléculaire la plus haute occupée (HOMO) et l'orbitale moléculaire la plus basse vacante (LUMO). En effet, la transition entre les états électroniques S1 et S0 est majoritairement issue d'une transition entre la HOMO et la LUMO. Dans le cas des molécules organiques, la présence d'une alternance de simple et doubles liaisons (molécule conjuguée) permet d'expliquer l'absorption puis l'émission de photons dans le visible. Cette conjugaison se traduit par une délocalisation des électrons  $\pi$  sur l'ensemble des atomes participant à cette conjugaison. La HOMO et la LUMO sont alors majoritairement constituées d'orbitales orthogonales au plan moyen de la molécule et localisées sur différentes parties du système. La notion de transfert de charge correspond au déplacement de la densité électronique entre la HOMO et la LUMO. Ce transfert de charge est nécessaire afin de pouvoir émettre dans le visible, plus celui-ci est important plus l'émission est déplacée vers les longueurs d'ondes rouges. Enfin, selon les atomes appartenant à la HOMO ou la LUMO et la nature de la transition électronique, la photoluminescence peut induire des transformations chimiques comme par exemple des isomérisations.

Plusieurs facteurs peuvent modifier l'écart énergétique et ainsi la longueur d'onde d'émission. Par exemple, la présence de groupements auxochromes permet de stabiliser ou déstabiliser les deux orbitales frontières HOMO et LUMO. Ces groupements sont dits électroaccepteurs si ils induisent un déplacement de la densité électronique dans leur direction comme pour le groupement carbonyle C=O. Les groupements électrodonneurs induisent quand à eux un déplacement de la densité électronique vers les autres atomes de la molécule participant à la conjugaison.

Le deuxième facteur prépondérant dans la modification de la longueur d'onde d'émission fait intervenir l'environnement extérieur de la molécule. La présence d'un solvant, plus ou moins polaire ou d'autres molécules (protéines, chimie hôte-invité) permet également le déplacement de la longueur d'onde d'émission. Il est alors primordial de décrire précisément ces effets environnementaux afin de bien reproduire le phénomène de photoluminescence. La prise en compte de

l'environnement dans le calcul des énergies des états électroniques a été développée dans le chapitre 4.

Enfin, on appelle un déplacement hypsochrome quand l'écart énergétique entre la HOMO et la LUMO augmente (la HOMO est stabilisée et/ou la LUMO est déstabilisée) et un déplacement bathochrome quand ce dernier diminue (la HOMO est déstabilisée et/ou la LUMO est stabilisée).

## 6.2 La chimiluminescence

### 6.2.1 Exemples de chimiluminescence

La chimiluminescence permet l'émission d'un photon suite à une réaction chimique en solution. Cette réaction fait souvent intervenir un catalyseur et une molécule de dioxygène  $O_2$  en plus de la molécule de départ. En effet, comme on le verra plus en détail dans la description des phénomènes bioluminescents, l'énergie d'excitation du substrat est apportée par la cassure d'une liaison peroxyde O-O. La chimiluminescence possède de nombreux avantages comme une simplicité d'utilisation ou des molécules faciles à synthétiser. Cependant le faible nombre de molécules possédant cette propriété limite le développement de cette partie de la luminescence. En outre, le rendement quantique des différentes molécules chimiluminescentes reste encore faible, souvent inférieur à 10%. Bien que la chimiluminescence soit un peu méconnue, elle est présente dans de nombreuses applications.

Le luminol est la molécule la plus connue en terme de chimiluminescence. La lumière émise par cette molécule interagit avec les ions ferriques  $Fe^{3+}$  du sang et dévoile la présence de sang dans une pièce. L'intérêt porté depuis peu à la police scientifique dans la culture populaire a mis en valeur le luminol. Un autre exemple de molécule chimiluminescente réside dans la création de bâtons lumineux. A l'intérieur de ce bâton les deux substances précurseurs de la chimiluminescence sont séparées et l'une d'elles enfermée dans une capsule de verre. L'émission de lumière est alors visible lorsque le verre est brisé et le mélange des deux réactifs possible. Plusieurs teintes de couleurs peuvent être obtenues par l'addition d'un colorant chimique dans le milieu. Ce colorant absorbe la lumière émise par chimiluminescence avant d'émettre dans une autre longueur d'onde. Outre un intérêt purement ludique, ces bâtons lumineux peuvent être utilisés lors d'explorations spéléologiques. Enfin pour finir, notons l'utilisation de la chimiluminescence comme outils de détection d'explosifs. En effet, de nombreux explosifs possèdent des groupements nitro ( $-NO_2$ ) ou nitrate ( $-NO_3$ ). Les molécules possédant ces groupements sont alors détruites par pyrolyse ce qui libère du monoxyde d'azote qui réagit ensuite avec une molécule d'ozone. Cette réaction entraîne la formation de dioxyde d'azote ( $NO_2$ ) excité produisant une lumière issue par chimiluminescence visible en infra-rouge<sup>83</sup>.

### 6.2.2 Les 2-coumaranones

Les 2-coumaranones sont une classe de molécule chimiluminescentes découverte en 1976 par Lofthouse *et al.*<sup>84</sup> (voir figure 6.2). Elles émettent de la lumière en solvant polaire et en présence d'une base forte. Cette classe de molécule a été remise à jour récemment par le groupe du Prof. Dr. Beckert et plus particulièrement par le Dr. Schramm<sup>85,86</sup>. Leurs travaux ont permis de faciliter la synthèse d'une grande variété de 2-coumaranones par un mécanisme en 1 étape avec un rendement compris entre 10 et 90%. En outre, le rendement quantique de luminescence maximum atteint par cette famille de molécule est de 5.9%. Un des intérêts d'étudier ces

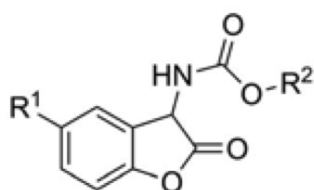


FIGURE 6.2 – A gauche, structure d'une 2-coumaranone. R<sup>1</sup> est un halogène et R<sup>2</sup> un aryle ou un alkyle. A droite, montage montrant l'émission de chimiluminescence des 2-coumaranones dans un pièce obscure (demi-image de gauche) et en plein jour (demi-image de droite).

molécules est de comprendre le déroulement du mécanisme réactionnel. En effet, celui-ci fait intervenir une molécule de dioxygène et entraîne l'excitation de la molécule via la formation d'un cycle dioxétanone. Il est alors possible de tracer de nombreux liens entre le mécanisme réactionnel d'émission chez les lucioles et dans les 2-coumaranones. Dans le chapitre 10 de cette thèse, on développe certains résultats obtenus sur la chimiluminescence des 2-coumaranones en collaboration avec le groupe de Prof. Dr. Beckert.

## 6.3 La bioluminescence

La bioluminescence est la propriété des espèces vivantes à produire de la lumière suite à une réaction chimique réalisée à l'intérieur d'une protéine. Ces réactions mettent souvent en jeu un substrat et une protéine, on parle ainsi de catalyse enzymatique. On retrouve la bioluminescence chez de nombreuses espèces aussi bien terrestres que maritimes, animales que végétales. Bien que la bioluminescence soit connue depuis l'antiquité avec Aristote, les découvertes importantes ont eu lieu ces cinquante dernières années grâce à l'obtention des premières structures

cristallographiques et à de nouveaux moyens de recherche. La découverte de nouvelles espèces bioluminescentes s'est accélérée, l'exploration des fonds marins montrant que plus de 90% espèces présentes sont bioluminescentes. Finalement, la compréhension de ce mécanisme chimique complexe a permis le développement de nouvelles technologies comme l'utilisation de la bioluminescence comme marqueur lumineux pour les cellules. Le prix Nobel de chimie de 2008 a récompensé Shimomura, Chalfie et Tsien pour la découverte et le développement de la Green Fluorescent Protein (GFP) présente dans la méduse *Æquorea victoria* et activée par la réaction de bioluminescence issue de cette dernière<sup>2</sup>.

Dans le suite de cette partie, on présente les principales espèces bioluminescentes avant de parler des applications développées chez les espèces vivantes puis par les scientifiques.

### 6.3.1 Les espèces bioluminescentes

Les espèces bioluminescentes peuvent être classées selon les espèces terrestres et les espèces maritimes (voir figure 6.3). Parmi les espèces terrestres, les lucioles sont parmi les plus connues et les plus étudiées. On retrouve également de la bioluminescence chez des coléoptères, des vers, des gastéropodes et des champignons.

Parmi les espèces maritimes on trouve beaucoup de méduses bioluminescentes, mais également des crevettes et des micro-organismes comme le plancton. La bioluminescence marine est également issue en majorité d'une collaboration symbiotique entre espèces maritimes et des bactéries.

Chez les espèces vivantes, la bioluminescence peut se manifester de différentes manières : les espèces peuvent luire, émettre des flashes lumineux plus ou moins long ou éjecter des nuages bioluminescents. Les applications issues de la bioluminescence sont alors très variées<sup>87</sup>

- Pour communiquer : les lucioles utilisent la bioluminescence pour communiquer entre elles mais également comme parade nuptiale. Pour cela elles peuvent réguler la fréquence des flashes lumineux en les accélérant ou les ralentissant et même en gardant une émission de lumière constante.
- Pour se défendre : Par des flashes lumineux ou en éjectant un nuage de fluides bioluminescents, certaines espèces peuvent distraire leur prédateur. D'une autre manière, des méduses peuvent se camoufler grâce à la bioluminescence. Des photophores situés sur le ventre de la méduse imitent la lumière venue de la surface et se faisant camoufle la méduse des prédateurs situés en profondeur.
- Pour attaquer : la bioluminescence est un outil remarquable en tant que leurre pour attirer les proies. De nombreux poissons utilisent ce mécanisme notamment les poissons pêcheurs qui possèdent une relation symbiotique avec des bactéries bioluminescentes. Elle peut également servir à aveugler les proies.

---

2. "The Nobel Prize in Chemistry 2008". Nobelprize.org. Nobel Media AB 2014. Web. 22 Apr 2017. <[http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/2008/](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/)>

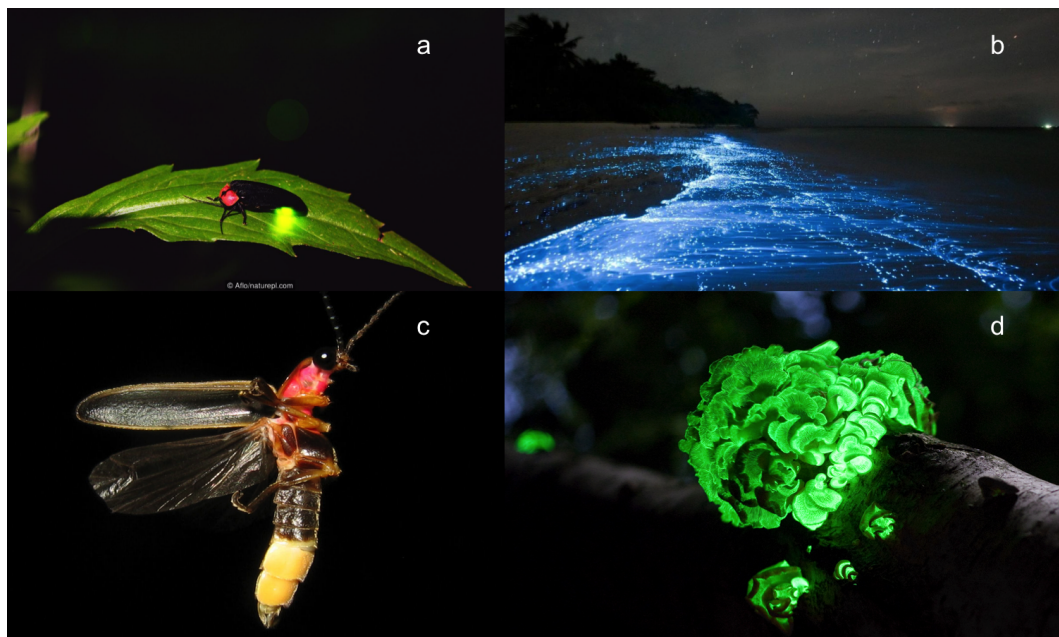


FIGURE 6.3 – Images de plusieurs espèces bioluminescentes : a) *Luciola cruciata*, Crédit : Aflo/naturepl.com. b) *Dinoflagellata*, Crédit : Doug Perrine/naturepl.com. c) *Photonus Pyralis*, Crédit : By art farmer from evansville indiana, usa - firefly 8823, CC BY-SA 2.0. d) *Panellus stipticus*, Crédit : By Ylem - Own work

### 6.3.2 Les applications dans le domaine scientifique

La communauté scientifique a rapidement entrevu les possibilités de la bioluminescence comme source de lumière (voir Figure 6.4). En effet, cette dernière est issue de systèmes biologiques et ne présente pas de risque de toxicité pour les cellules du vivant. En outre, l'émission de lumière a lieu suite à une réaction chimique in-vivo simplifiant le protocole expérimental contrairement à un système fluorescent classique où l'émission fait suite à une excitation lumineuse. Ainsi la BioLuminescence Imaging (BLI) ou imagerie bioluminescente est très efficace dans l'étude des systèmes in-vivo. Je vais présenter par la suite plusieurs types d'applications dans différents domaines scientifiques utilisant cette méthode.

L'utilisation principale de la BLI dans le domaine médical se traduit par le marquage de cellules cibles<sup>88,89</sup>. Cette technique non invasive permet de visualiser les processus cellulaires et biologiques chez les êtres vivants.

L'application la plus connue de la BLI permet le suivi et le traitement de cancers<sup>88,89</sup>. En effet, la bioluminescence est une façon simple de visualiser et de suivre l'évolution de cellules cancéreuses dans le temps. La versatilité et la facilité de cette méthode a permis son développement dans les laboratoires biologiques



et médicaux. Le principe est le suivant : des gènes codant pour la synthèse d'une protéine bioluminescente sont insérés dans les cellules cancéreuses d'un individu. Puis le substrat correspondant à la protéine choisie est inséré à l'intérieur du corps par injection intraveineuse ou sous-cutanée. Ainsi les cellules cancéreuses vont "s'illuminer" par bioluminescence permettant de tracer et suivre leurs développement. La lumière visible est de l'ordre de quelques centimètres de profondeur selon la couleur émise ce qui permet une visualisation précise et avec une bonne résolution. La longueur d'onde d'émission de la bioluminescence joue un rôle important. Plus celle-ci émet dans le rouge plus l'intensité du signal sera important. De cette façon, le choix du système émetteur n'est pas trivial. En outre chaque système bioluminescent possède ses qualités et ses défauts. Outre la longueur d'onde d'émission, le nombre de réactifs nécessaires à la réaction ainsi que leur taux de consommation sont des facteurs importants. Cette méthodologie est plus adaptée aux petites espèces où l'intensité lumineuse est maximale. Cette méthodologie peut également permettre de suivre l'évolution des cellules souches dans un corps<sup>88,89</sup>. Ainsi l'imagerie bioluminescente rend compte de la prolifération de ces cellules souches et de leur efficacité dans le traitement de cancers ou de régénération de cellules cardiaques. Citons également le suivi et la visualisation de l'expression du gène impliqué dans la maladie d'Alzheimer par BLI.

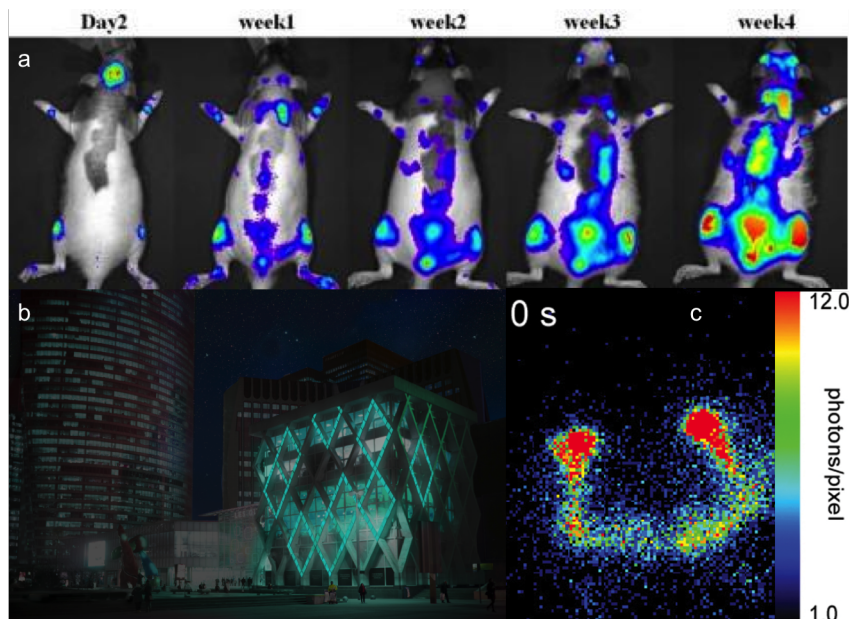


FIGURE 6.4 – Image de différentes applications utilisant la bioluminescentes : a) BLI de cellule cancéreuse du sein chez la souris<sup>90</sup>. b) Bâtiment éclairé par la bioluminescence issue de bactérie symbiotiques du calamar, Crédit : Glowee <http://www.glowee.fr/> c) BLI issue d'un cerveau de drosophile après un influx nerveux<sup>91</sup>.

Outre la visualisation de cellules, la BLI permet de suivre l'abondance et l'activité de systèmes protéiques<sup>88,89</sup>. Cette application fait intervenir le principe du BRET,

Bioluminescence Resonance Energy Transfert. Afin d'expliquer ce principe je vais développer le fonctionnement de la méduse *Æquorea victoria*. Dans cette dernière la bioluminescence est assurée par le substrat coelenterazine qui en association avec du dioxygène et dans l'enzyme aequorine activée par des ions  $\text{Ca}^{2+}$  produit une lumière bleue. Cependant la méduse émet une lumière verte issue de la GFP. Le mécanisme de la fluorescence se traduit par un transfert de l'énergie issue de la bioluminescence et permettant l'excitation du chromophore fluorescent. C'est la fluorescence du chromophore qui émet un photon dans le vert. Dans le domaine médical, le BRET permet de visualiser l'interaction entre deux protéines. Pour cela, l'une des protéines d'intérêt est liée au système bioluminescent et l'autre au système fluorescent. Quand les deux protéines sont éloignées, on ne visualise que la bioluminescence mais quand les deux protéines sont en interactions, on observe uniquement la fluorescence indiquant que les deux protéines sont suffisamment proches.

Les sous-facteurs présents dans la réaction de la bioluminescence jouent également un rôle important. Le cation  $\text{Ca}^{2+}$  permet la libération de neurotransmetteurs impliqués dans l'activité des neurones<sup>92</sup>. On peut alors suivre le déplacement de ces ions dans les cellules nerveuses en utilisant le principe du BRET de l'exemple décrit précédemment.

La molécule d'ATP présente dans la réaction de bioluminescence chez les lucioles est également étudiée et observée par BLI<sup>93</sup>. Dans la méthode dite d'ATP-métrie, on quantifie la concentration de la molécule d'ATP par bioluminescence. La propriété de la molécule d'ATP est de servir de carburant pour le fonctionnement des cellules chez le vivant. Ainsi si on observe de la bioluminescence émanant d'un échantillon test, on peut en déduire que ce dernier abrite des micro-organismes. Cette méthode d'ATP-métrie est souvent utilisée dans la détection micro-biologiques dans des aliments ou dans la vérification de la stérilisation de matériel.

La bioluminescence est également utilisée afin de détecter des polluants environnementaux comme des métaux lourds et autres molécules toxiques<sup>93</sup>. On peut citer l'utilisation de bactéries bioluminescentes du produit Microtox<sup>TM</sup> qui permet de détecter et quantifier la présence de polluants en ecotoxicologie de l'analyse de l'eau. On relève pour cela la quantité des produits toxiques nécessaire à diminuer l'émission de lumière de moitié.

Enfin, la bioluminescence est depuis peu utilisée comme source lumineuse dans l'éclairage public ou l'illumination de bâtiments. Une start-up parisienne nommée Glowee développe une technologie basée sur des bactéries symbiotiques de calamar, comme espèces bioluminescentes.

# **Troisième partie**

## **Les protéines**



La réaction de bioluminescence fait intervenir la transformation chimique d'un substrat catalysée par une protéine. Afin de comprendre ce phénomène il est également important de s'intéresser aux protéines en général ainsi qu'à leurs modélisations. Les protéines jouent un rôle fondamental dans les espèces vivantes. Elles sont issues de l'information présente dans les gènes et assurent de nombreuses fonctions dans le vivant. Dans les prochains chapitres on s'intéresse dans un premier temps à l'intérêt biologique de ces molécules. On étudie pour cela leurs structures et leurs principales applications. Puis, dans un second temps, on développe les différentes méthodes théoriques utilisées pour représenter et étudier ces protéines.



# Chapitre 7

## Généralités sur les protéines

### 7.1 La structure des protéines

#### 7.1.1 Les acides aminés, structure primaire

Les protéines sont un assemblage de un ou plusieurs polypeptides, des molécules issues de la polymérisation d'acides aminés. Les acides aminés sont des composés organiques pouvant être classés selon trois catégories : les acides aminés apolaires, polaires et chargés. Ces molécules sont composées d'un groupement amine et d'un groupement carboxylate séparés par un groupe CH-R, R étant différent pour chaque acide aminé. Les quatre fonctions chimiques (carboxylate, amine, le groupement R et l'hydrogène), sont portées par le carbone central, appelé également carbone

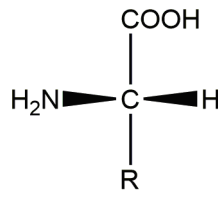


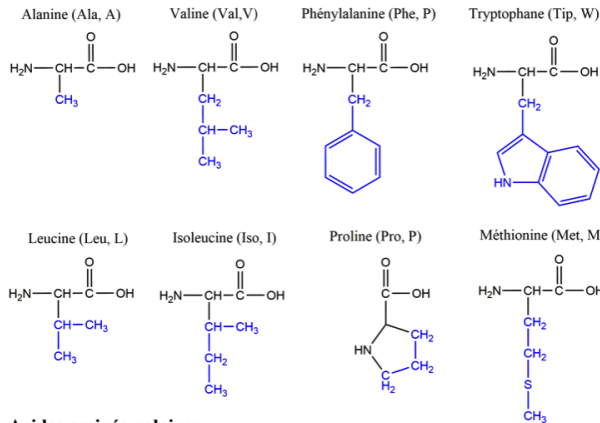
FIGURE 7.1 – L-acide aminé.

alpha ou  $C_\alpha$ . Notons également que les acides aminés sont des molécules chirales à cause du carbone alpha asymétrique. Dans le monde biologique la configuration du  $C_\alpha$  des acides aminés naturels est la forme levogyre, on parle de L-acide aminé comme décrit dans la figure 7.1.

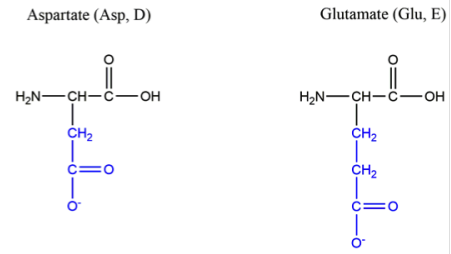
Il existe 20 acides aminés "standard" que l'on retrouve dans toutes les protéines. On peut discriminer les acides aminés selon plusieurs classifications : leur structure chimique, leur nom, par un code à trois lettres et par un code à une lettre. La structure des 20 acides aminés, leur nom et la signification des codes sont renseignés dans la figure 7.2.

Les acides aminés apolaires possèdent une chaîne latérale R qui interagit peu ou

**Acides aminés hydrophobes**



**Acides aminés Chargés**



**Acides aminés polaires**

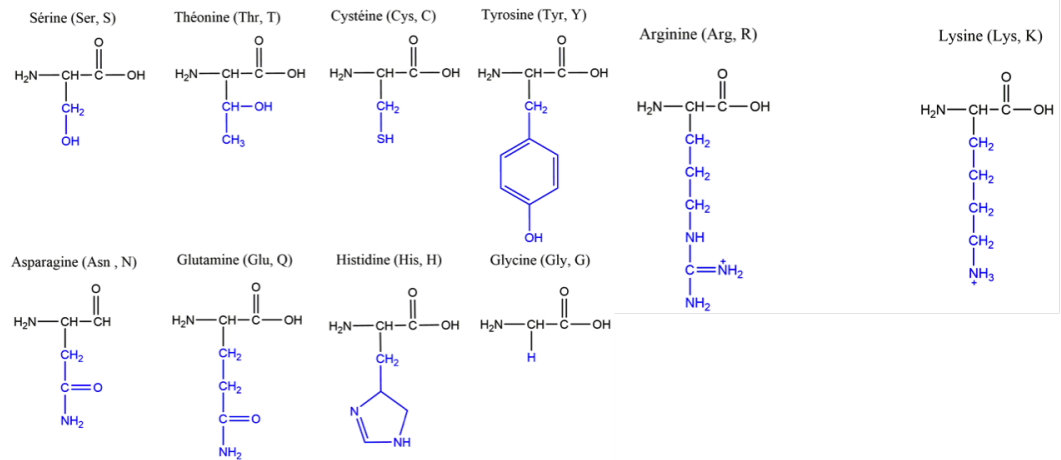


FIGURE 7.2 – Structures et abréviations des 20 acides aminés standard. La chaîne latérale (groupement R) est représentée en bleu.

pas avec les molécules d'eau environnantes. Ces acides aminés sont souvent situés à l'intérieur de la protéine et, à cause de leur apolarité, n'interagissent pas avec les substrats impliqués dans des réactions chimiques.

Les acides aminés polaires possèdent une chaîne latérale R qui interagit avec les molécules d'eau environnantes par des interactions faibles de type liaisons hydrogène. Ces acides aminés se situent aussi bien à la surface de la protéine où ils entrent en interaction avec le solvant qu'à l'intérieur de la protéine. Les liaisons hydrogène sont responsables de la stabilité des structures secondaires des protéines. De plus, ces acides aminés peuvent coordonner et stabiliser des substrats extérieurs à la protéine et impliqués dans des réactions chimiques grâce à des liaisons hydrogène. Selon la polarité de l'environnement autour de ces acides aminés, certains peuvent exister sous plusieurs formes, une forme protonée et une forme déprotonée. C'est le cas par exemple de l'histidine et de la cystéine. L'absence du proton sur la cystéine peut ainsi induire un repliement différent de la protéine, via la formation d'un pont disulfure, impliqué dans la stabilisation de la structure tertiaire. Enfin notons la présence de la glycine, l'acide aminé le plus simple (R=H) qui est inclus dans cette catégorie des acides aminés polaires. Sa présence y est par défaut, ce dernier n'étant



ni hydrophobe ni chargé.

Les acides aminés chargés possèdent une chaîne latérale R qui est chargée au pH physiologique de la protéine. Comme pour les acides aminés polaires, la charge portée par le groupement R permet à des molécules d'eau ou à d'autres molécules de former des liaisons hydrogène avec l'acide aminé considéré.

Les acides aminés sont les briques de la construction des protéines. Par une réaction de condensation ces derniers peuvent se lier afin de former un polypeptide. Cette réaction fait intervenir la fonction hydroxyde d'un acide aminé avec la fonction amine d'un autre acide aminé et conduit à la formation d'une liaison peptidique C-N liant de façon covalente les deux acides aminés. On parle également d'acide aminé résidu ou tout simplement de résidu lorsqu'un acide aminé est à l'intérieur d'une structure protéique. Enfin, par convention la succession de résidus du polypeptide est écrite de telle sorte à avoir à gauche la fonction amine du premier résidu libre, c'est le N-terminal et à droite la fonction acide carboxylate du dernier résidu libre, c'est le C-terminal.

Dans les protéines, l'organisation structurale se présente selon 4 niveaux qui seront détaillés par la suite. Le premier niveau dit structure primaire de la protéine correspond à la séquence des acides aminés d'un polypeptide comme présenté dans la figure 7.3. On appelle le backbone ou chaîne principale la succession des atomes carbone, oxygène, azote et hydrogène que constitue la liaison peptidique et le C $\alpha$  protoné.

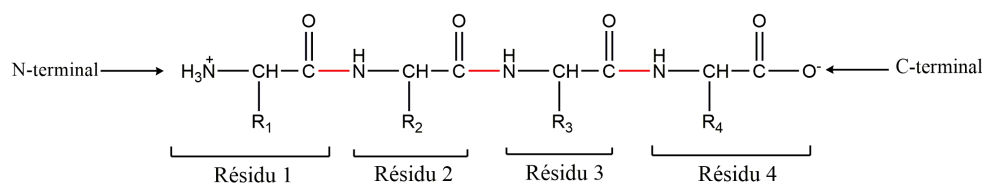


FIGURE 7.3 – Structure d'un polypeptide d'une protéine selon la convention standard. Le résidu 1 est appelé le résidu N-terminal et le résidu 4 est appelé le résidu C-terminal.

### 7.1.2 Structure secondaire

La structure secondaire d'une protéine rend compte de la conformation adoptée par le polypeptide. Dans ce dernier, la liaison peptidique entre chaque résidu possède deux formes résonnantes dues à la délocalisation électronique. Cette partie de la molécule est alors considérée comme une structure plane. Les torsions observées dans les protéines sont alors majoritairement issues des rotations entre les atomes C $\alpha$ -N et C $\alpha$ -C.

Le polypeptide va alors se replier afin de former une structure secondaire satisfaisant deux conditions : elle doit générer un minimum de contraintes stériques et posséder le plus grand nombre possible de liaisons hydrogène entre les groupements polaires du polypeptide. Deux différents types de structures secondaires sont souvent rencontrés dans les protéines : l'hélice  $\alpha$  et le feuillet  $\beta$ .

L'hélice  $\alpha$  correspond à une conformation torsadée du backbone. Celui-ci se tord en formant une hélice droite de l'ordre d'une dizaine de résidus. Dans cette structure secondaire les chaînes latérales sont pointées vers l'extérieur de l'hélice (voir Figure 7.4).

Le feuillet  $\beta$  consiste quand à lui a une succession de rangées de résidus alignées et

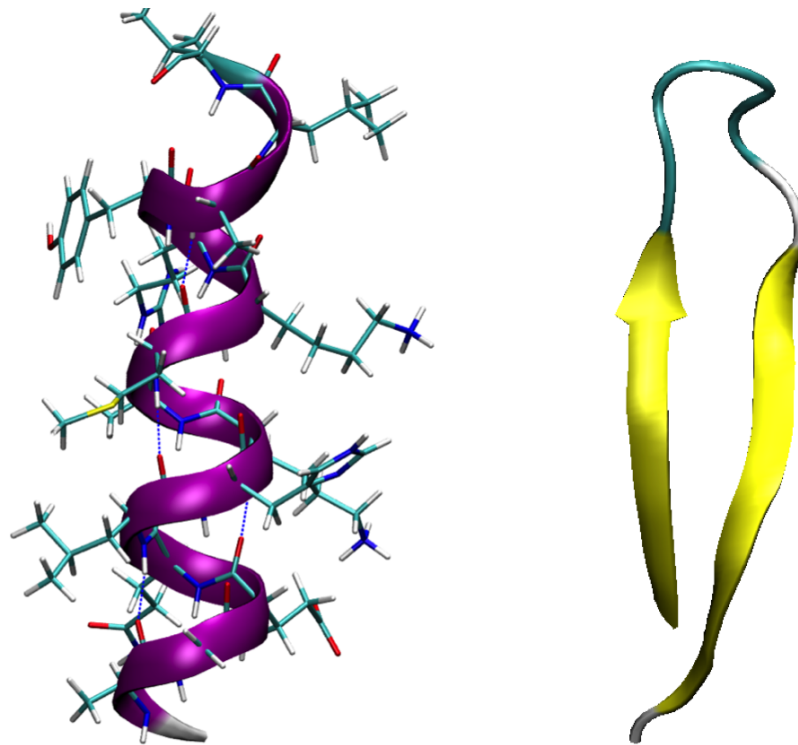


FIGURE 7.4 – A gauche, structure en ruban d'une hélice  $\alpha$  et représentation en tout atome des résidus la composant. A droite, structure en ruban de deux feuillets  $\beta$  antiparallèles en jaune ainsi que d'une boucle "épingle a cheveu" reliant ces deux feuillets.

liées par des liaisons hydrogène. Ces feuillets  $\beta$  peuvent être alignés de deux façons : les feuillets  $\beta$  parallèles et antiparallèles. Dans les feuillets parallèles, les rangées de résidus sont alignées dans le même sens alors que dans les feuillets antiparallèles elles sont alignées de façons opposées. En détails, cela signifie que pour les feuillets parallèles les séquences de résidus du N-terminal au C-terminal sont orientées dans le même sens. Finalement, les chaînes latérales sont disposées de part et d'autres du

plan formé par ce feuillet (voir Figure 7.4). Ces deux structures secondaires sont dites régulières car chaque résidu appartenant à cette structure possède une conformation du backbone similaire à son voisin proche. Des structures secondaires irrégulières existent et permettent de joindre deux structures secondaires régulières entre elles. Des boucles peuvent ainsi connecter des structures secondaires régulières de différentes façons ce qui entraîne l'apparition de motifs structuraux. Un exemple standard dit d'épingle à cheveu est visible dans la Figure 7.4. Finalement en terme de présence on considère en moyenne que 31% des résidus d'une protéine appartiennent à une hélice  $\alpha$ , 28% à un feuillet  $\beta$  et le reste à des boucles irrégulières.

### 7.1.3 Structure tertiaire

La structure tertiaire d'une protéine rend compte de la forme tri-dimensionnelle de celle-ci ainsi que de la stabilité de cette dernière. Après la formation des structures secondaires, les protéines se replient selon plusieurs types d'interactions.

Comme vu précédemment les résidus des protéines peuvent être classés selon leur nature hydrophobe (apolaire) ou hydrophile (polaire). Ainsi les résidus hydrophiles vont se retrouver à la surface de la protéine où la création de liaisons hydrogène avec le solvant est possible. En revanche, les résidus hydrophobes vont s'orienter vers l'intérieur de la protéine où le contact avec les molécules d'eau du solvant est minimal. On se retrouve alors avec un cœur de la protéine hydrophobe et une surface hydrophile. Le domaine d'une protéine représente une unité structurale d'une protéine comportant un cœur hydrophobe. Les protéines peuvent posséder plusieurs domaines comme par exemple la luciférase, protéine responsable de la bioluminescence chez les lucioles qui en possède deux (voir Figure 7.5).

Le repliement des protéines conduit le plus souvent à un seul arrangement stable en trois dimensions. Étonnamment, la structure repliée des protéines n'est pas excessivement plus stable que la structure dépliée. On considère en général une stabilisation de l'ordre de 0.1 kcal/mol par résidu pour la structure repliée. La principale force de stabilisation est l'effet hydrophobe. En effet, quand les résidus hydrophobes sont en contact direct avec les molécules d'eau cela perturbe le réseau de liaisons hydrogène entre la protéine et le solvant. Ces résidus hydrophobes s'agrègent alors entre eux afin de minimiser au maximum le contact avec les molécules d'eau et diminuer l'énergie totale de la protéine.

D'autres effets permettent également de stabiliser la structure de la protéine. On peut citer l'interaction entre une paire d'ions, comme par exemple entre le groupement ammonium d'une lysine et le groupement carboxylate d'un glutamate. La création d'un pont disulfure joue également un rôle dans le repliement de la protéine et influence la forme de cette dernière. Ces deux précédentes interactions n'apportent cependant qu'une légère stabilisation de la protéine repliée.

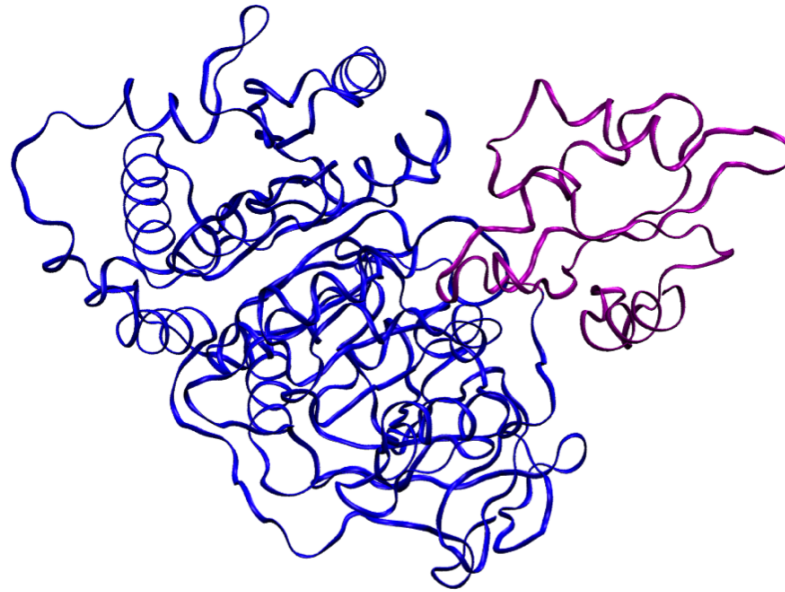


FIGURE 7.5 – Structure des domaines de la luciférase. En bleu, le domaine N-terminal et en violet le domaine C-terminal.

#### 7.1.4 Structure quaternaire

La structure quaternaire d'une protéine intervient sur les grosses protéines quand celles-ci possèdent plusieurs chaînes polypeptidiques. Chaque chaîne est alors appelée sous-unité. Ces différentes sous-unités peuvent être similaires, on utilise le préfixe -homo ou bien différentes, on utilise le préfixe -hétéro. Dans tous les cas ces sous-unités s'organisent de façon à avoir le plus haut degré de symétrie possible. L'intérêt de ces sous-unités est que celles-ci vont interagir les unes par rapports aux autres par effet de coopération ou par allostérie, le principal mécanisme à l'œuvre dans l'hémoglobine.

## 7.2 L'utilité des protéines

Les protéines possèdent de nombreuses applications chez les espèces vivantes. On ne parlera ici que d'un petit nombre d'applications parmi les plus connues.

Certaines protéines ont un rôle de transport de ligands comme par exemple l'hémoglobine qui transporte les molécules de dioxygène des poumons vers le reste du corps via les voies sanguines. L'intérêt de cette protéine réside dans son fonctionnement de captation du dioxygène, selon un mécanisme coopératif. En effet l'hémoglobine est une protéine hétérotétramérique formée de quatre chaînes peptidiques identiques deux à deux (deux chaînes  $\alpha$  et deux chaînes  $\beta$ ) (voir Figure 7.6). Chaque chaîne peut fixer une molécule de dioxygène soit quatre au total. Si on trace la courbe de la saturation de la protéine en fonction de la concentration de

dioxygène du milieu on obtient une courbe de type sigmoïde contrairement à une hyperbole observée dans les autres protéines fixatrice de dioxygène. A ce comportement on en déduit qu'à chaque fois qu'un dioxygène est capté par une chaîne, cela facilite la fixation des suivants. Le mécanisme à l'oeuvre est de type structural. Dans chaque sous-unité, le dioxygène se fixe à un hème, une métallo-porphyrine composé d'un ion fer(II). Le dioxygène va se coordiner sur ce fer(II) ainsi que sur une histidine voisine ce qui modifie la conformation de cette sous-unité et induit par la suite un changement important de la forme des sous-unités non-pourvue en dioxygène. Ainsi la quatrième molécule de dioxygène possède une affinité de fixation à l'hémoglobine cent fois supérieur à celle de la première molécule.

Les protéines sont également à l'oeuvre dans la structure des cellules du vivant. On

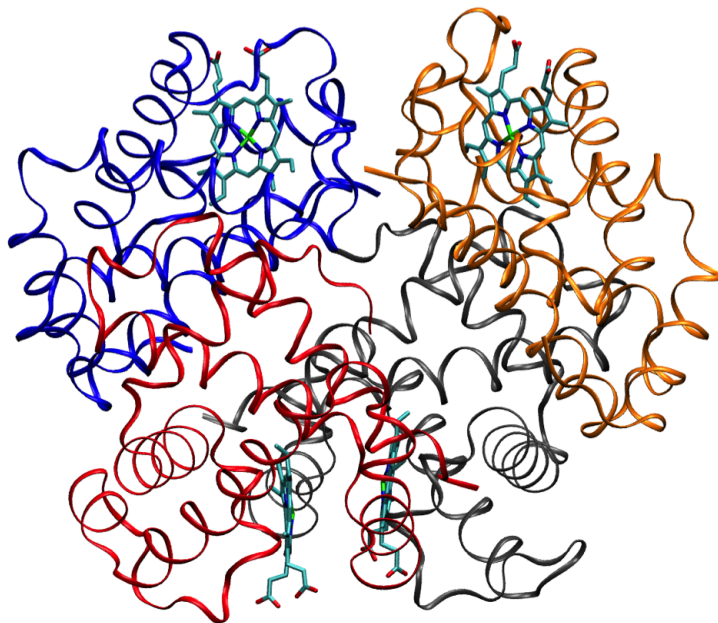


FIGURE 7.6 – Structure cristallographique déprotonée de l'hémoglobine  $\alpha$ . Les différentes sous-unités sont représentées en ruban et d'une couleur différente. Les quatre hèmes sont représentés en tout atome. La structure est issue du PDB 3hhb.pdb<sup>94</sup>.

appelle ces protéines des protéines fibreuses. On peut citer comme exemple le collagène et la kératine présents respectivement dans le cartilage, dans les poils et dans les ongles. Les protéines structurales peuvent être classées en trois catégories selon la taille des fibres. Les microtubules possèdent un diamètre supérieur à 240 Å comme le collagène ; les filaments intermédiaires un diamètre supérieur à 100 Å comme la kératine. Enfin les micro-filaments ont un diamètre de l'ordre de 70 Å. Les micro-filaments présents dans le cytosquelette des cellules eucaryotes sont formés d'actine.

L'actine est une protéine importante dans le phénomène de contraction musculaire. Cela conduit à un troisième type de fonction importante, les protéines motrices. L'actine va se lier avec la myosine, une protéine composée de deux poly-

peptides enroulés ensemble. Chaque polypeptide possède une tête, qui peut se lier à l'actine. La myosine fait alors coulisser le filament d'actine grâce à un changement de conformation due à une réaction avec l'ATP. En parallèle, le même comportement est observé sur un autre filament d'actine, ces deux filaments se rapprochent alors et induisent la contraction musculaire (voir Figure 7.7).

Finalement on va aborder les réactions enzymatiques. Les enzymes sont une

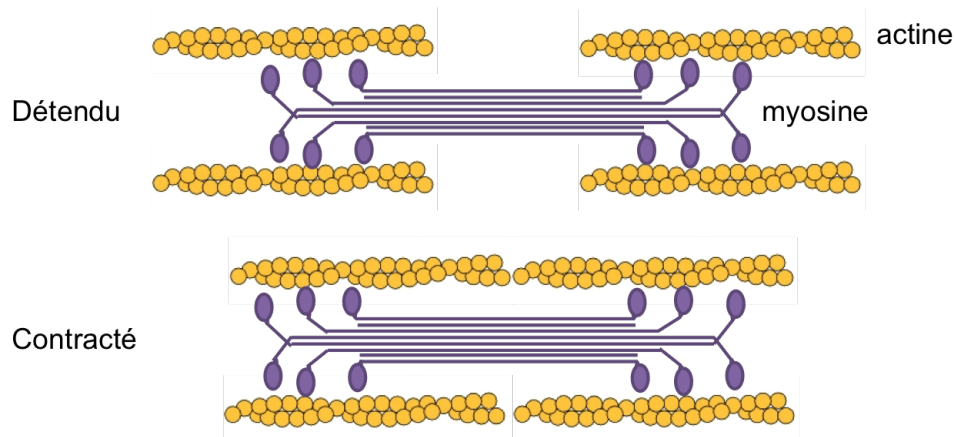


FIGURE 7.7 – Mouvement des micro-filaments d'actines lors de la contraction musculaire.

sous-catégorie de protéine permettant la catalyse de réactions chimiques. Chaque enzyme permet le plus souvent de catalyser une réaction, c'est à dire que seul un substrat, ou des molécules chimiquement similaires peuvent interagir avec une seule enzyme. De même un substrat peut généralement subir l'influence d'une protéine, mais certains sont identifiables dans plusieurs réactions catalytiques différentes. On assimile souvent les réactions enzymatiques à une image de clé (substrat) avec une serrure (enzyme). On classe les enzymes selon le type de réaction qu'elles catalysent avec le suffixe -ase représentant l'enzyme. Ainsi les hydrolases représentent les enzymes catalysant les réactions d'hydrolyse. Le rôle des enzymes est primordial, car elles réalisent la majorité des réactions chimiques dans les systèmes biologiques. Il existe de nombreuses études sur les enzymes, permettant une meilleure compréhension de ces systèmes complexes. Les réactions de bioluminescence font notamment intervenir des enzymes, ces molécules sont donc au cœur de ma recherche.

# Chapitre 8

## Comment modéliser une protéine

### 8.1 Modéliser et représenter un système protéique

Les systèmes protéiques sont modélisés à partir de fichier contenant les coordonnées des atomes. Ces coordonnées sont issues de la résolution de cristaux par diffraction des rayons X ou par diffraction de neutrons. Selon la résolution et la précision de la carte de densité électronique obtenue, le modèle construit sera plus au moins complet et ne contiendra pas d'hydrogène dans le cas d'une structure obtenue par RX.

Une fois ce modèle construit et publié, il devient disponible en libre accès sur le site <http://www.rcsb.org>, formaté selon le type PDB pour Protéin Data Bank. De nombreux logiciels existent ensuite afin de protoner la protéine, calculer le pKa des résidus et construire les boucles manquantes. Dans cette thèse, le pKa des résidus a été calculé avec le site H++<sup>95</sup>, les résidus protonés avec le programme leap de la suite logicielle Amber<sup>55</sup> et les boucles manquantes construites avec le logiciel Disgro<sup>96</sup>.

Pour les protéines utilisées dans cette thèse, un certain nombre d'approximations ont également été considérées. Ainsi, nous n'incluons pas d'ion magnésium dans le milieu car il n'y a pas de preuves expérimentales de sa présence après la réaction de bioluminescence. De même les systèmes modélisés sont équilibrés afin d'avoir une charge totale neutre. On protone pour cela certaines histidines (charge +1) ou on ajoute des contre-ions au milieu. Il est alors important de noter que le pH du milieu protéique n'est pas pris en compte et que celui-ci peut varier d'un modèle à l'autre.

Lors de cette thèse, le logiciel VMD (Visual Molecular Dynamics) est utilisé afin de visualiser et représenter les structures des systèmes protéiques. Dans les figures présentes dans cette thèse, les protéines sont dessinées sous forme de ruban. Ce ruban fait également apparaître les structures secondaires de la protéine. Les substrats, résidus et molécules d'eau à l'intérieur de la protéine sont représentés par la méthode CPK ou Licorice (appelé aussi bâton dans cette thèse). Dans la représentation CPK, les atomes sont dessinés comme des sphères et les liaisons comme des cylindres. Pour la représentation Licorice, les liaisons sont dessinées par des cylindres et les

atomes par des points. Dans ce logiciel le code couleur utilisé afin de représenter les résidus et les substrats est le suivant : O rouge, H hydrogène, C cyan, N bleu, S jaune. Pour la structure secondaire de la protéine, les hélices alpha sont dessinées en violet et les feuillets bêta en jaune.

## 8.2 Les modèles disponibles

Les protéines sont des molécules comportant plusieurs milliers d'atomes, les méthodes de chimie quantiques sont alors peu adaptées. Une solution consiste à représenter ces protéines par des champs de force en mécanique moléculaire (voir Chapitre 2). On ne reviendra pas sur la définition ces champs de force mais nous allons définir les différents modèles utilisés.

Le premier modèle possible est le tout-atome. Chaque atome de la protéine est explicitement représenté. Ce modèle permet une bonne représentation des propriétés intrinsèques de la protéine, notamment les propriétés locales comme le déplacement d'une boucle. Ce modèle est aussi utilisé quand on cherche à obtenir une représentation précise du site actif des enzymes et de l'interaction entre la protéine et le substrat.

D'autres modèles existent également quand on ne s'intéresse qu'à des propriétés globales ou à de très grandes protéines. Dans les méthodes gros-grain<sup>97</sup>, les résidus sont remplacés par des pseudo atomes afin de réduire le nombre de degrés de liberté de la protéine (voir Figure 8.1). Il existe plusieurs façons de construire ces pseudo atomes selon la simplification voulue. En partant de l'approche tout atome, une première approximation consiste à négliger les atomes d'hydrogène non-polaires et les associer à l'atome auquel ils sont liés. On peut ensuite simplifier les chaînes latérales, passant d'une représentation explicite vers un à quatre pseudo atomes. Le backbone peut-être également simplifié, et finalement l'approximation extrême est d'avoir un pseudo-atome pour un résidu. Afin de représenter ces différents modèles gros-grain, on peut aussi bien utiliser des champs de force basés sur des équations physiques ou issues de données statistiques sur les structures cristallographiques de protéines.

## 8.3 Applications

### 8.3.1 Umbrella sampling

L'umbrella sampling<sup>98</sup> est une méthode utilisée en dynamique moléculaire classique afin d'échantillonner précisément la surface énergétique d'une protéine en fonction d'une coordonnée de réaction. Le principe est de séparer le chemin réactionnel en plusieurs fenêtres que l'on va explorer séparément lors d'une dynamique moléculaire. Pour cela on introduit un potentiel pour chaque fenêtre afin de forcer la coordonnée de réaction à rester proche de la valeur de la coordonnée de réaction imposée. Ce potentiel est le plus souvent représenté par un potentiel quadratique totalement paramétrable (voir Figure 8.2). L'avantage de cette méthode est de pouvoir échantillonner un grand nombre de conformations possibles dont



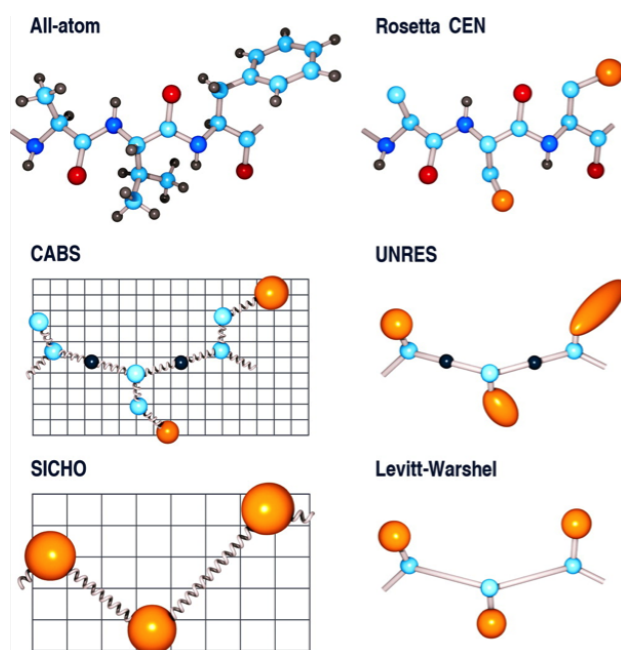


FIGURE 8.1 – Représentation en tout-atome d'un tripeptide ainsi que les modèles gros-grain correspondant<sup>97</sup>. Les modèles présentés sont le Rosetta centroid mode (CEN), le CABS, le UNRES, le SICHO et le modèle Levitt et Warshel. Le code couleur est le suivant : O rouge, H noir, C cyan, N bleu, pseudo-atomes orange et bleu foncé. Les pseudo-liaisons de distances variables sont représentées sous forme de ressort.

certaines peu ou pas accessibles lors d'une dynamique moléculaire. Les valeurs échantillonnées sont alors récupérées et, assumant des recouvrements entre plusieurs fenêtres, une méthode d'analyse d'histogrammes pondérés<sup>99,100</sup> est utilisée afin de tracer le potentiel d'énergie libre.

La méthode d'umbrella sampling est souvent utilisée pour comparer la stabilité de deux conformations différentes, aussi bien lors de la rotation d'un angle, ou lors du déplacement d'une boucle.

### 8.3.2 Docking

Le docking<sup>101</sup> est une approche computationnelle permettant de connaître l'affinité d'un substrat vis à vis d'un site actif d'une enzyme. Le plus souvent, cette méthode se déroule selon deux étapes. La première consiste à lister le nombre de conformations possibles que peut adopter un substrat dans un site actif, c'est à dire la position et l'orientation de ce substrat à l'intérieur de la protéine. Puis la fonction de score est calculée pour chaque conformation, cette fonction rendant compte de la stabilité et de l'affinité de la conformation testée dans le site actif. Ces deux étapes donnent lieu à des développements théoriques différents. Pour les méthodes d'échantillonnage, on peut chercher à maximiser les interactions chimiques créées entre le substrat et la protéine et à raisonner en séparant le substrat en fragment

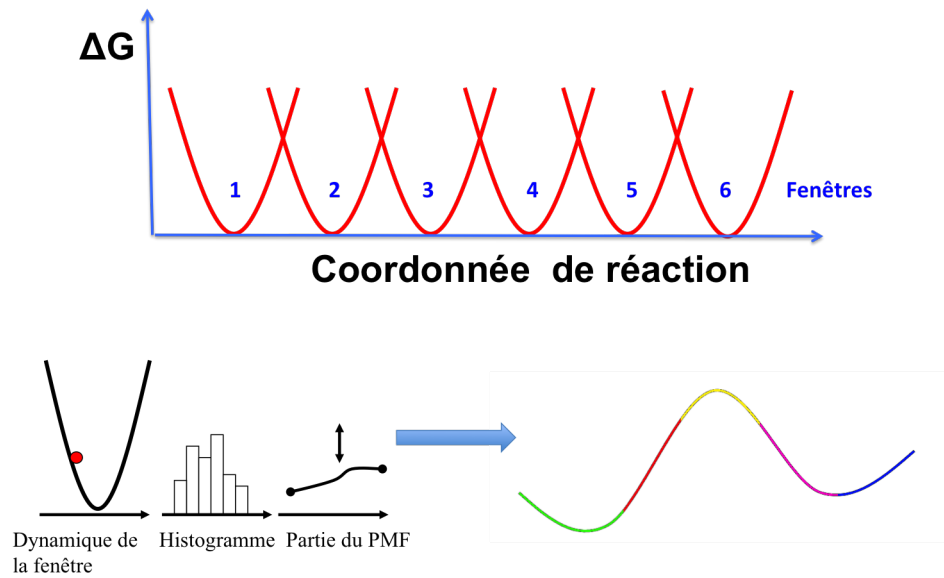


FIGURE 8.2 – Schéma des différentes étapes d'un umbrella sampling. La coordonnée de réaction est séparée en plusieurs fenêtres, les valeurs obtenues sont alors regroupées pour former un histogramme pondéré permettant d'obtenir le PMF.

ou en générant aléatoirement des conformations. Le calcul des fonctions de score peut également être réalisé de différentes façons. La méthode la plus simple consiste à utiliser des fonctions de score basées sur le calcul d'interactions non-covalentes issues de champ de force classique. D'autres algorithmes ajoutent également les contributions issues des liaisons hydrogène, de la solvation ou de l'entropie. On peut aussi tenir compte de résultats statistiques issus de structures de protéines comportant une relation complexe-substrat.

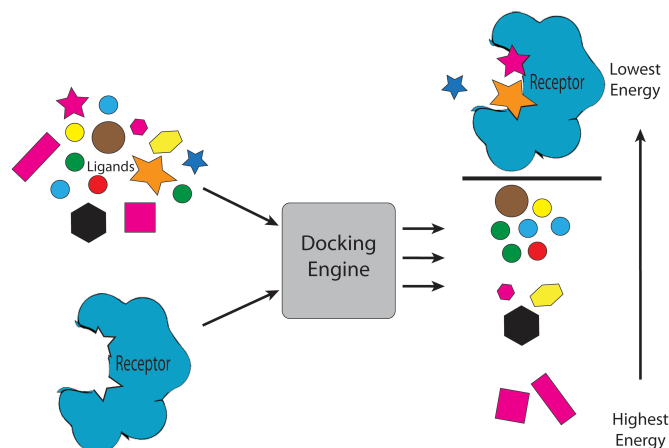


FIGURE 8.3 – Schéma du fonctionnement d'un screening virtuel grâce au docking<sup>102</sup>.

Le docking est l'une des approches les plus utilisées dans l'étude des protéines. Les structures cristallographiques des protéines sont rarement obtenues avec le

substrat d'intérêt à l'intérieur de la protéine. L'apport du docking permet d'avoir une première idée de la position et de l'orientation du substrat pour un faible coût computationnel. Une autre utilisation du docking consiste à effectuer un screening virtuel sur des molécules tests afin par exemple, de trouver de nouveaux inhibiteurs ou substrats actifs d'une enzyme (voir Figure 8.3). Une autre application est de modéliser l'influence que peut avoir la mutation d'un résidu présent dans le site actif, ou une mutation du substrat lui-même. Le docking reste une méthode peu chère et relativement efficace afin de modéliser les interactions protéines-substrats.

### 8.3.3 Processus biologiques

L'un des buts final des sciences computationnelles est de pouvoir modéliser entièrement le fonctionnement du métabolisme biologique des êtres vivants. L'évolution spectaculaire des moyens de calcul permet aujourd'hui de visualiser certains mécanismes complexes.

La modélisation et l'étude des membranes cellulaires sont particulièrement représentées dans ce domaine<sup>103,104</sup>. Les membranes cellulaires possèdent en effet différents canaux permettant le transport de molécules entre le cytoplasme et l'environnement extérieur. Il est ainsi possible de simuler le passage d'ion potassium à travers le canal dédié par dynamique moléculaire classique. Ces systèmes sont très difficiles à modéliser car ils possèdent un nombre important d'atomes, certaines simulations dépassant le million d'atomes. En outre, le temps de dynamique nécessaire au passage d'ion est relativement long souvent supérieur à 100 ns. L'utilisation des modèles de type gros-grains couplés avec des dynamiques moléculaire classiques guidées (steered MD) permettent de réduire les temps de calculs. Dans cette dernière on impose une force à un certain nombre d'atomes d'intérêts. Par exemple on peut grâce à ce type de dynamique "tirer" une molécule d'eau à travers l'aquaporin, la protéine servant de canal à l'eau dans les cellules.

D'autres types de canaux ont également été modélisés comme des transporteurs membranaires, ou le complexe de translocation, permettant la sortie de polypeptide hors de la cellule.

### 8.3.4 Prédiction et repliement

Une autre avancée permise par les méthodes computationnelles est de prédire la structure repliée des protéines<sup>105,106</sup>. La multiplication du nombre de structures cristallographiques de protéines couplée avec une augmentation exponentielle des capacités de calculs est majoritairement responsable du développement de cette science. Grâce à cette méthode, il est possible de prédire la conformation stable et repliée d'une protéine en ne connaissant que sa structure primaire.

La prédiction des structures de protéines donne lieu à un concours bisannuel où de nombreuses équipes testent de nouveaux algorithmes afin de reproduire au mieux les structures. Le CASP (Critical Assessment of protein Structure Prediction, <http://predictioncenter.org/>) permet ainsi de quantifier les efforts et améliorations liés à ce domaine. De nos jours, les algorithmes permettent, pour environ 10% de protéines, de prédire une structure repliée avec une résolution proche de celle obtenue par cristallographie et donc assez précise afin de poursuivre les études biologiques sur ce système (mécanisme biochimique, screening médicinal ...). Les outils de prédictions font intervenir deux processus différents. Le premier consiste à utiliser la base de données issue de structures cristallographiques et à faire l'hypothèse que deux séquences peptidiques similaires donnent lieu à des structures secondaires et tertiaires similaires. La deuxième méthode consiste à découper la chaîne polypeptidique en plusieurs morceaux puis à identifier ces morceaux dans d'autres protéines. La nouvelle protéine est donc construite à partir de pièces issues de plusieurs protéines. L'assemblage est rendu possible par le calcul de fonctions de scores utilisées dans le docking.

L'étude du repliement des protéines a également été un moteur dans la création d'échange entre le monde scientifique et le grand public. Les sites Folding@home (<http://folding.stanford.edu/>) et Rosetta@home (<https://boinc.bakerlab.org/>) permettent aux utilisateurs de donner du temps de calcul de leur ordinateur afin de réaliser des repliements de protéines. Le serious game Foldit (<https://fold.it/portal/>) incite les joueurs à replier une protéine comme ils le souhaitent mais en respectant quelques contraintes. Ces initiatives permettent d'explorer de nouvelles conformations et repliements possibles et incongrues pour des protéines.

De ces différentes recherches, la notion de protéinopathie ou maladies issues d'un mauvais repliement d'une protéine est apparue. La prédiction de structures de protéines a permis de mieux comprendre et traiter cette classe de maladie, incluant la maladie de Parkinson, d'Alzheimer et du diabète de type II.

## **Quatrième partie**

# **Comment modifier la couleur de la bioluminescence**



Dans cette partie de la thèse, on s'intéresse aux mécanismes permettant de modifier et moduler la longueur d'onde de la bioluminescence. Il existe deux différentes voies permettant d'altérer la couleur d'émission, en modifiant le substrat ou la protéine.

Dans le chapitre 9, on discute de l'effet apporté par un analogue de l'émetteur de la bioluminescence chez les lucioles. Ce substrat modifié, l'infraoxyluciférine donne lieu à une émission dans le rouge et peut adopter différentes géométries. Le but est alors de reproduire puis de comprendre l'impact induit par la modification du substrat mais également de l'environnement sur la longueur d'onde d'émission.

Dans le chapitre 10, de nouvelles structures de luciférases ont été obtenues par rayon X. Les espèces d'où proviennent ces protéines possèdent une longueur d'onde d'émission de la bioluminescence différente de celle obtenue chez les lucioles. On étudie deux systèmes bioluminescents, l'un émettant dans le rouge et l'autre dans le vert, à comparer à la couleur jaune-verte des lucioles. Une étude complète du système émettant dans le vert est réalisée afin de trouver les effets permettant d'expliquer la modification de la couleur de la bioluminescence.





# Chapitre 9

## Modulation grâce à un analogue de la luciférine

### 9.1 Contexte

L'imagerie bioluminescente consiste à mettre en valeur certaines cellules cibles en les "éclairant". Le but est alors de suivre l'évolution du nombre et de la position de ces cellules dans le temps. Cette technique est notamment utilisée lors du traitement des cancers<sup>88</sup>, voir section 6.3.2. Cependant, une difficulté actuelle reste de détecter cette bioluminescence, la couleur et l'intensité impactant fortement la visibilité. Les lucioles produisent naturellement une lumière intense jaune-verte émise dans le milieu du spectre visible à environ 550 nm. La recherche actuelle s'oriente alors vers la mise en place de systèmes ayant une émission vers la couleur rouge du spectre visible ; moins énergétique mais ayant une meilleure pénétration des tissus biologiques.

C'est dans ce contexte que les groupes du Prof. Anderson et du Dr. Pule ont synthétisé un analogue de la luciférine appelé infraluciférine ou iLH<sub>2</sub><sup>9</sup> (voir figure 9.1). Cet analogue a été obtenu en ajoutant une double liaison entre les deux cycles de la luciférine. En conséquence, cette molécule émet une bioluminescence dans le rouge à 1.85 eV (670 nm) avec la luciférase naturelle *Photonus Pyralis*<sup>9</sup>. Dans un

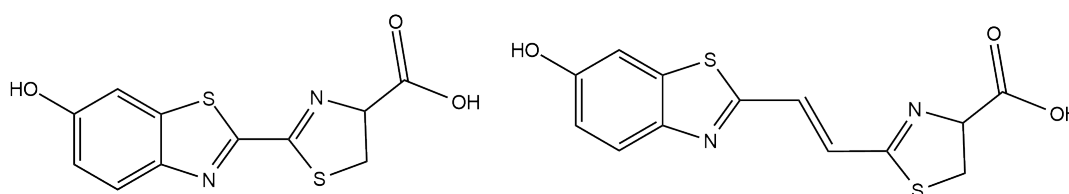


FIGURE 9.1 – Représentation graphique de la luciférine LH<sub>2</sub> à gauche et de l'infraluciférine iLH<sub>2</sub> à droite.

deuxième temps, les équipes d'expérimentateurs cités ci-dessus ont également pu moduler la longueur d'onde d'émission en mutant des résidus précis de la luciférase *Photonus Pyralis*. Ainsi la mutation du résidu sérine 284 en thréonine est également

responsable d'un déplacement vers le rouge de la longueur d'onde d'émission de la bioluminescence. Dans ce chapitre, le système modélisé est issu du fichier PDB 4G37 représentant une conformation particulière de la luciole *Photonus Pyralis* (voir section 9.5 et Annexes).

Lors de ce chapitre nous allons montrer les résultats théoriques qui permettent de discuter des différentes modifications entraînant un changement de la valeur de la longueur d'onde d'émission de la bioluminescence.

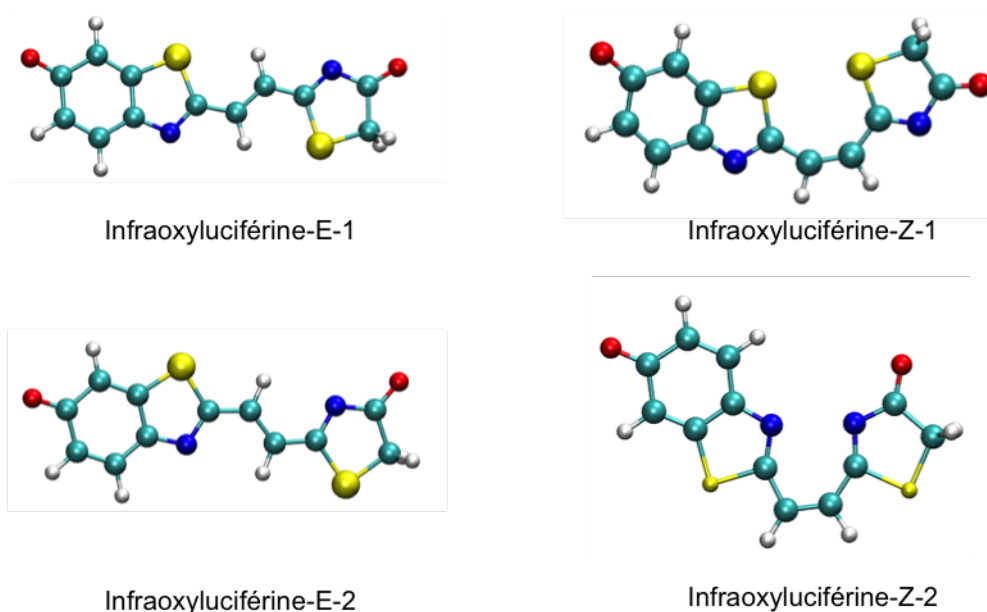


FIGURE 9.2 – Représentation graphique des différents substrats issus de OxyiLH<sub>2</sub> étudiés dans ce chapitre.

## 9.2 Problématiques

La première priorité de l'étude menée au cours de cette thèse est de réussir à modéliser le système et à reproduire les longueurs d'onde d'émission obtenues par les expériences menées par le groupe les groupes d'Anderson et Pule.

Le deuxième objectif est de détailler et interpréter quels sont les effets qui s'exercent entre la protéine et le substrat. Les principales interactions sont de nature géométrique (gêne stérique) ou de nature électrostatique (cavité polaire, liaisons hydrogène). On regarde également s'il existe un lien entre le résidu 284 et le substrat, pouvant expliquer la modification de la longueur d'onde d'émission.

Le troisième but est de réaliser une étude détaillée du substrat modifié infraoxy-luciférine (OxyiLH<sub>2</sub>). En effet celui-ci peut exister sous différentes conformations chimiques, pouvant être responsables d'un changement de la longueur d'onde

de la bioluminescence. Dans la cavité de la luciférase, OxyLH<sub>2</sub>, ainsi que OxyiLH<sub>2</sub> subissent de la gêne stérique des résidus proches, bloquant le substrat dans une seule conformation possible. Il est alors impossible d'avoir un changement de conformation du substrat pendant la réaction de bioluminescence. En outre, cet aspect conformationnel n'a pas été abordé du point de vue expérimental, les structures cristallographiques de la protéine comportant iLH<sub>2</sub> n'étant pas disponibles. Par la suite on étudie quatre conformères différents listés dans la figure 9.2, deux avec la double liaison sous forme trans (E) et deux sous forme cis (Z).

Finalement, la compréhension de ce système nous permet d'émettre des hypothèses sur le mécanisme possible de la modulation de longueur d'onde de la bioluminescence chez les lucioles (voir Figure 9.3).

### 9.3 Résultats

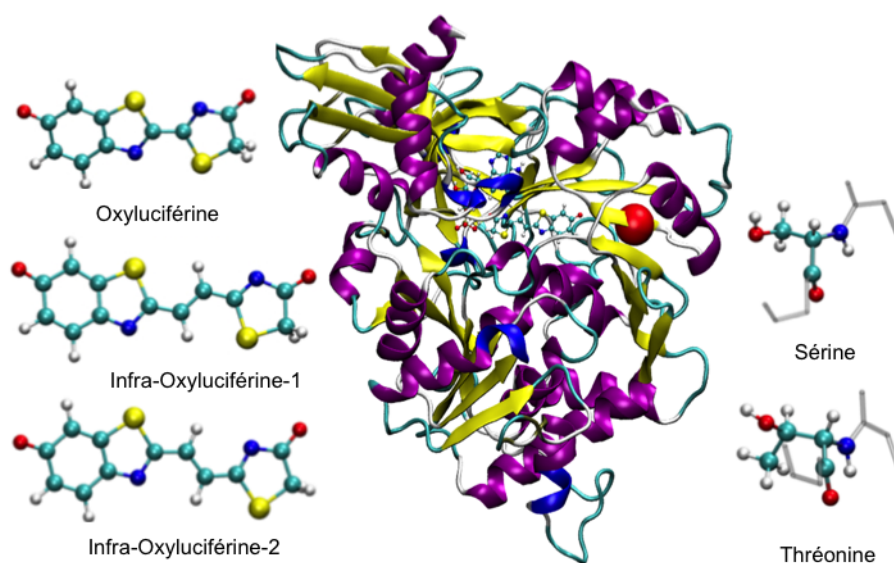


FIGURE 9.3 – Représentation graphique des différents modèles principalement utilisés pour cette étude. Sur la gauche, les structures et noms des 3 principaux substrats étudiés. Sur la droite, la structure et nom du résidu 284 dans la protéine naturelle (Sérine) et dans la protéine mutée (Thrénine). Au milieu, la représentation de la protéine luciférase avec OxyiLH<sub>2</sub>-E-1. La protéine est représentée sous forme de ruban avec les hélices alpha en violet et les feuillets bêta en jaune. La position du résidu 284 est indiquée en rouge.

Tableau 9.1 – Description des 10 modèles utilisés dans ce chapitre.

Protéine étudiée	Résidu 284	Substrat	Nom du modèle choisi
4G37	Sérine	oxyluciferine	Ser_OxyLH <sub>2</sub>
4G37	Sérine	infraoxyluciférine-E-1	Ser_OxyiLH <sub>2</sub> -E-1
4G37	Sérine	infraoxyluciférine-E-2	Ser_OxyiLH <sub>2</sub> -E-2
4G37	Sérine	infraoxyluciférine-Z-1	Ser_OxyiLH <sub>2</sub> -Z-1
4G37	Sérine	infraoxyluciférine-Z-2	Ser_OxyiLH <sub>2</sub> -Z-2
4G37-S284T	Thréonine	oxyluciférine	Thr_OxyLH <sub>2</sub>
4G37-S284T	Thréonine	infraoxyluciférine-E-1	Thr_OxyiLH <sub>2</sub> -E-1
4G37-S284T	Thréonine	infraoxyluciférine-E-2	Thr_OxyiLH <sub>2</sub> -E-2
4G37-S284T	Thréonine	infraoxyluciférine-Z-1	Thr_OxyiLH <sub>2</sub> -Z-1
4G37-S284T	Thréonine	infraoxyluciférine-Z-2	Thr_OxyiLH <sub>2</sub> -Z-2

### 9.3.1 Benchmark des méthodes

Pour chaque modèle décrit dans le tableau 9.1, on a choisi une *image* (snapshot) correspondant à la structure d'énergie minimale dans la MD associée réalisée, voir la section 9.5 pour les détails de la MD. Cette structure a ensuite été minimisée au niveau MM et optimisée géométriquement dans l'état fondamental puis dans le 1<sup>er</sup> état singulet excité avec la méthode hybride QM/MM afin d'obtenir le  $T_e$  dans la protéine.

Afin de reproduire la valeur du  $T_e$  obtenue expérimentalement<sup>9</sup> une 1<sup>ère</sup> approche a consisté à réaliser des calculs de type TD-DFT/MM, rapides et relativement performants. L'approche standard dans notre groupe et collaborateurs est l'utilisation de la fonctionnelle CAM-B3LYP couplée avec la base 6-31G(d,p)<sup>107,108</sup>. Les résultats obtenus sont cependant très en dessous des valeurs espérées avec une différence d'énergie supérieure à 0.25 eV entre les énergies expérimentales et celles calculées théoriquement ( $\Delta T_e$ ), donnant lieu à une valeur de l'émission calculée déplacée vers le bleu par rapport aux valeurs expérimentales. Un benchmark réalisé sur 4 autres fonctionnelles, B3LYP, M06, M06-2X et  $\omega$ B97-XD en gardant la même base a quand à lui montré un problème de convergence lors des calculs, la transition HOMO-LUMO recherchée devenant interdite avec une force d'oscillateur égale à 0 ( $f=0$ ), notamment pour les fonctionnelles B3LYP et M06. Le choix d'une autre base, à savoir 6-311G(2d,p) est issu d'un article<sup>109</sup> où la base 6-311G+(2d,p) représente le choix le plus raisonnable. La fonction diffuse a été retirée à cause de l'incompatibilité de ces fonctions diffuses avec l'electrostatic embedding que nous avons utilisé, voir la section 9.5.

Les calculs effectués sur la protéine 4G37 et avec plusieurs substrats montrent des résultats très différents selon la fonctionnelle choisie (voir tableau 9.2). Pour les 3 fonctionnelles CAM-B3LYP, M06-2X et  $\omega$ B97-XD les résultats sont déplacés vers le bleu avec une valeur du  $T_e$  plus faible et comprise entre +0.25 eV et +0.5 eV par rapport au résultat expérimental. Cependant pour B3LYP et M06, les résultats sont compris dans l'incertitude liée à la DFT avec un écart entre les  $T_e$  calculés et

Tableau 9.2 – Transitions électroniques entre S1 and S0 ( $T_e$ ) (et force d’oscillateur associée) obtenues en QM/MM sur la structure 4G37 pour différentes fonctionnelles. TD-DFT/MM avec une base 6-311G(2d,p) sur une structure optimisée dans l’état S1 avec la fonctionnelle indiquée.

Modèle	Ser_OxyLH <sub>2</sub>	Ser_OxyLH <sub>2</sub> -E-1	Ser_OxyLH <sub>2</sub> -E-2
B3LYP	2.20 (564) f=0.34	2.00 (616) f= 0.53	1.87 (663) f=0.53
M06	2.28 (544) f=0.37	2.04 (607) f=0.64	1.91 (651) f=0.52
CAM-B3LYP	2.54 (488) f=0.59	2.23 (554) f=1.13	2.10 (590) f=1.13
M062X	2.53 (490) f=0.57	2.21 (559) f=1.11	2.07 (597) f=0.83
$\omega$ B97XD	2.54 (487) f=0.57	2.2 (563) f=1.11	2.13 (581) f=0.81
Exp. eV (nm)	2.22 (558)	1.85 (670)	

ceux obtenus expérimentalement inférieur à 0.2 eV. Le choix retenu pour la suite est l’utilisation de la fonctionnelle B3LYP couplée avec la base 6-311G(2d,p).

Par la suite des calculs MS-CASPT2/MM ont également été réalisés. Les  $T_e$  calculées sont issues de calculs *single point* MS-CASPT2/MM avec la géométrie du système dans le premier état excité optimisé lors d’un calcul TD-DFT/MM avec la fonctionnelle CAM-B3LYP et la base 6-31G(d,p). Plusieurs tests ont été réalisés afin d’obtenir le meilleur ratio en terme de résultats et de temps de calcul. Pour le substrat OxyLH<sub>2</sub> trois différents espaces actifs ont été pris en compte alors que pour OxyLH<sub>2</sub> seul un espace actif a pu être testé à cause de limitations techniques (cf 9.5.3 et Annexes A1, A2 et A3). D’autres facteurs comme le nombre d’états électroniques pris en compte dans le calcul MS-CASPT2 et la valeur du shift IPEA ont également été testés, voir la section 9.5.

Le shift IPEA a une valeur par défaut de 0.25 dans la suite de programme MOLCAS. Cette valeur avait été choisie car elle réduit la valeur moyenne des erreurs systématiques<sup>33</sup>. Cependant le choix de cette valeur du shift a été prise il y a plusieurs années et les tests effectués sur des molécules à couches ouvertes, des molécules diatomiques ou des métaux de transition. Ainsi plusieurs études<sup>110</sup> sont en cours sur de nouvelles molécules afin de vérifier la robustesse de cette valeur. La conclusion de l’article montre que pour les molécules organiques, un shift IPEA à 0 semble le plus adapté. Les résultats n’étaient pas encore publiés lors de cette étude. Nous avons alors choisi de comparer le shift IPEA standard (0.25) avec un shift à 0.1. Les résultats obtenus avec un shift de 0.1 sont corrects comme nous allons le voir dans le paragraphe suivant. Nous n’avons cependant pas testé les conclusions de l’article<sup>110</sup>, un benchmark complet combinant shift IPEA, taille de l’espace actif et taille de la base serait nécessaire afin d’avoir un point de vue global sur la question.

Si on prend maintenant en compte tous les modèles, la comparaison entre le shift IPEA = 0.25 et le shift IPEA = 0.1 montre que ce dernier donne des résultats plus proches des valeurs obtenues expérimentalement avec un  $\Delta T_e$  plus faible (voir tableau 9.3). La taille de l’espace actif a un impact visible pour un shift de 0.25, mais donne des résultats plus chaotiques pour un shift de 0.1. Ainsi, pour un shift de 0.1, l’augmentation de la taille de l’espace actif entre 14/13 et 16/14 dans le modèle

Tableau 9.3 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en MS-CASPT2/MM entre différents espaces actifs et 2 valeurs différentes du shift IPEA. Les calculs sont réalisés avec une base ANO-RCC-VTZP, en tenant compte de 2 états électroniques et avec un level shift de 0.1.

Modèle	Espace MS-CASPT2 eV (nm)		MS-CASPT2 eV (nm)	Exp. <sup>9</sup> eV (nm)
	actif	IPEA= 0.25		
Ser_OxyLH <sub>2</sub>	14/13	2.41 (513)	2.24 (554)	2.22 (558)
	16/14	2.30 (531)	2.14 (579)	
	18/15	2.37 (523)	2.17 (570)	
Ser_OxyLH <sub>2</sub> -E-1	16/15	2.35 (527)	2.19 (566)	1.85 (670)
Ser_OxyLH <sub>2</sub> -E-2	16/15	2.06 (601)	1.92 (645)	
Thr_OxyLH <sub>2</sub>	14/13	2.39 (518)	2.07 (598)	2.04 (605)
	16/14	2.32 (534)	2.15 (577)	
	18/15	2.32 (532)	2.15 (577)	
Thr_OxyLH <sub>2</sub> -E-1	16/15	2.35 (526)	2.15 (578)	1.75 (706)
Thr_OxyLH <sub>2</sub> -E-2	16/15	2.09 (592)	1.93 (643)	

Ser\_OxyLH<sub>2</sub> diminue la valeur du  $T_e$  alors que dans Thr\_OxyLH<sub>2</sub> la valeur du  $T_e$  augmente.

Les résultats ainsi obtenus en utilisant les méthodes TD-DFT/MM et MS-CASPT2/MM montrent un bon accord théorie-expérience (voir tableau 9.4). L'augmentation du système  $\pi$  conjugué entre OxyLH<sub>2</sub> et OxyiLH<sub>2</sub> induit une diminution du  $T_e$ , donnant lieu à un déplacement de la longueur d'onde d'émission dans le rouge, comme observé expérimentalement<sup>9</sup>. Au niveau orbitalaire cette émission est liée à un transfert de charge entre l'orbitale moléculaire la plus basse vacante (LUMO) située sur le cycle thiazolone vers l'orbitale moléculaire la plus haute occupée (HOMO) située sur le cycle benzothiazole (voir figure 9.4).

On en conclue que les calculs TD-DFT/MM avec B3LYP/6-311G(2d,p) donnent les résultats les plus fiables avec un temps de calcul raisonnable. Les calculs réalisés avec MS-CASPT2/MM mettent en valeur l'importance non négligeable du shift IPEA sur les résultats obtenus. Il reste cependant difficile d'établir une tendance vis à vis de ce shift IPEA sans des benchmarks plus complets.

### 9.3.2 Emission sans protéine

Une étude a également été menée sur l'impact du solvant sur la valeur du  $T_e$  en l'absence de protéine. Pour cela des calculs TD-DFT B3LYP/6-311G(2d,p) ont été réalisés dans le vide, dans un solvant polaire (l'eau) et dans un solvant apolaire (le chloroforme). Ces solvants sont modélisés de façon implicite par une approche de type PCM<sup>73,74</sup>. La structure optimisée obtenue au niveau S1 est similaire entre ces 3 modèles avec un RMSD inférieur à 0.02 Å. Si on compare les valeurs du  $T_e$  obtenues pour

Tableau 9.4 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. MS-CASPT2/MM obtenu avec une base ANO-RCC-VTZP, un espace actif 16-in-14 avec 2 états électroniques, un level shift de 0.1 et un IPEA shift de 0.1. TD-DFT/MM obtenu avec une base 6-311G(2d,p) sur une structure optimisée dans l'état S1 avec la fonctionnelle B3LYP.

Modèle	MS-CASPT2 eV (nm)	B3LYP eV (nm)	Exp. <sup>9</sup> eV (nm)
Ser_OxyLH <sub>2</sub>	2.14 (579)	2.20 (564)	2.22 (558)
Ser_OxyiLH <sub>2</sub> -E-1	2.19 (566)	2.00 (616)	1.85 (670)
Ser_OxyiLH <sub>2</sub> -E-2	1.92 (645)	1.87 (663)	
Thr_OxyLH <sub>2</sub>	2.15 (577)	2.15 (577)	2.04 (605)
Thr_OxyiLH <sub>2</sub> -E-1	2.15 (578)	1.98 (625)	1.75 (706)
Thr_OxyiLH <sub>2</sub> -E-2	1.93 (643)	1.87 (661)	
vacuum OxyLH <sub>2</sub>	2.25 (550)	2.27 (547)	
vacuum OxyiLH <sub>2</sub> -E-1	2.02 (614)	2.13 (580)	
vacuum OxyiLH <sub>2</sub> -E-2	1.93 (642)	2.03 (608)	

les différents solvants on remarque que l'augmentation de la polarité du solvant diminue la valeur du  $T_e$  (voir tableau 9.5). Cependant c'est avec le chloroforme que l'on obtient les valeurs les plus proches de celles obtenues avec le modèle QM/MM Ser\_OxyLH<sub>2</sub> qui tient compte de l'environnement protéique.

### 9.3.3 L'influence de la dynamique moléculaire

Pour chaque modèle une dynamique moléculaire MD de durée comprise entre 5 ns et 10 ns a été réalisée. Après équilibration de la MD, observée après environ 1 ns, on observe l'arrivée de plusieurs molécules d'eau absentes de la structure cristallographique dans la cavité où se trouve le substrat. L'ajout de ces molécules d'eau conduit à la formation d'un réseau de liaisons hydrogène entre le substrat et l'envi-

Tableau 9.5 – Transitions électroniques entre S1 and S0  $T_e$  (et force d'oscillateur associée) obtenues en TD-DFT/MM avec la fonctionnelle B3LYP et la base 6-311G(2d,p) pour différents environnements et sur une structure optimisée dans l'état S1 avec la même fonctionnelle dans les mêmes conditions.

Modèle	Dans l'eau (PCM)	Dans 4G37 (QM/MM)	Dans le chloroforme (PCM)	Dans le vide
OxyLH <sub>2</sub>	2.09 (592) f=0.89	2.20 (564) f=0.33	2.23 (554) f=1.15	2.27 (547) f=0.41
OxyiLH <sub>2</sub> -E-1	1.84 (672) f=1.38	2.00 (616) f=0.53	1.92 (643) f=1.21	2.13 (580) f=0.76
OxyiLH <sub>2</sub> -E-2	1.77 (700) f=1.23	1.87 (663) f=0.53	1.85 (670) f=1.08	2.03 (608) f=0.65

ronnement proche, comme certains résidus de la luciférase ou des molécules d'eau (voir figure 9.5). Lors de la MD, ce réseau évolue mais on note une certaine récurrence dans la position des molécules d'eau et des résidus visés. Ainsi la majorité de ce réseau est situé entre l'atome d'oxygène du cycle benzothiazole, noté O1 du substrat (voir figure 9.6) et les résidus polaires environnants, se délocalisant jusqu'à l'interface solvant/protéine. Parmi les résidus les plus présents on note la glutamine 311, l'asparagine 229, l'arginine 337 et le résidu 284 dont on étudie la mutation. Ce réseau de liaisons hydrogène a également une influence non négligeable sur la valeur du  $T_e$ . Par la suite le terme "avant MD" correspond à la structure cristallographique minimisée au niveau MM et le terme "après MD" à la structure d'énergie minimale sélectionnée après MD, et appelée *image* par la suite. Si on compare les valeurs des  $T_e$  sur le modèle Ser\_OxyLH<sub>2</sub> on remarque que les valeurs obtenues "après MD" sont toutes déplacées vers le bleu par rapport à celle "avant MD" (voir tableau 9.6). Ce résultat est en accord avec les précédents résultats expérimentaux et théoriques<sup>72,111</sup>. C'est l'interaction d'une ou plusieurs molécules d'eau avec le substrat qui conduit à ce déplacement de la longueur d'onde d'émission vers le bleu. Ainsi, les résultats obtenus "après MD" possèdent un meilleur accord avec l'expérience que ceux "avant MD", démontrant la nécessité de recourir à des simulations de MD afin de bien reproduire l'environnement protéique. Finalement on remarque que les calculs TD-DFT/MM sont plus sensibles à ces interactions que les calculs MS-CASPT2/MM.

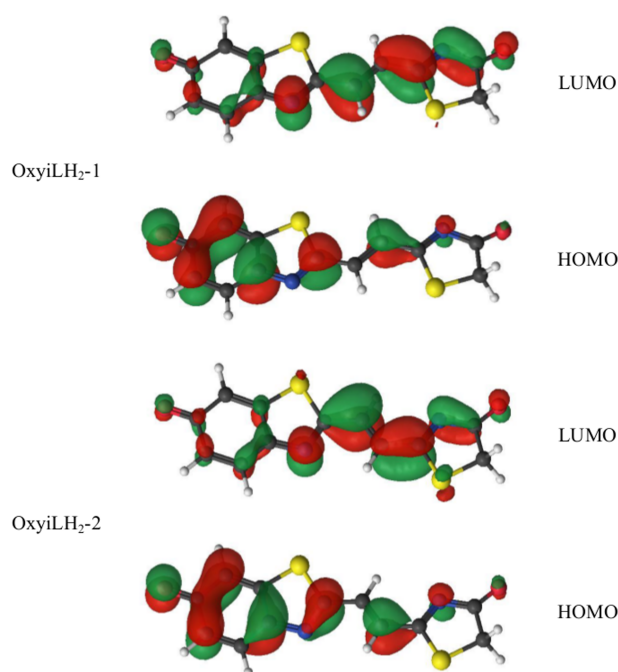


FIGURE 9.4 – Orbitales frontières de OxyiLH<sub>2</sub>-E-1 (en haut) and OxyiLH<sub>2</sub>-E-2 (en bas) obtenues par un calcul CASSCF/MM dans la protéine et dessinées avec le logiciel Luscus<sup>112</sup>.



Tableau 9.6 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur une structure issue avant et durant une dynamique moléculaire classique. MS-CASPT2/MM obtenu avec une base ANO-RCC-VTZP, un espace actif 16-in-14 avec 2 états électroniques, un level shift de 0.1 et un IPEA shift de 0.1. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la fonctionnelle B3LYP.

Modèle	Ser_OxyLH <sub>2</sub> Avant MD	Ser_OxyLH <sub>2</sub> Pendant MD	Exp.
B3LYP	1.98 (623)	2.20 (564)	2.22 (558)
MS-CASPT2 16-in-14	2.09 (594)	2.14 (579)	

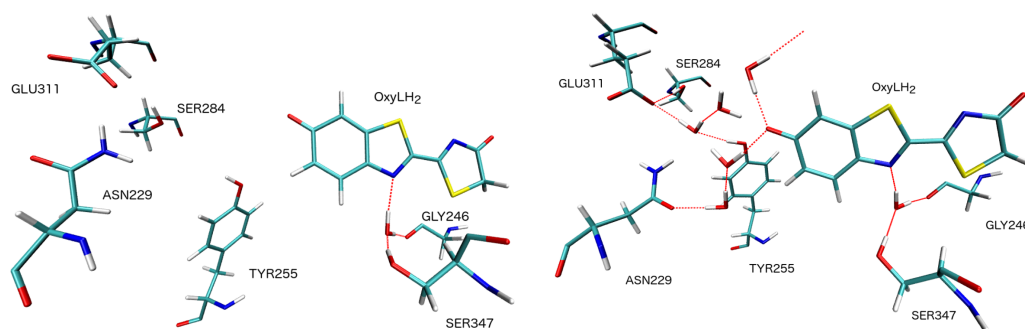


FIGURE 9.5 – Représentation du réseau de liaisons hydrogène dans le modèle Ser\_OxyLH<sub>2</sub> avant la dynamique moléculaire classique (MD) (à gauche) et "après MD" (à droite).

### 9.3.4 Le résidu 284 et le $T_e$

Parmi les résidus appartenant au réseau de liaisons hydrogène dans la cavité, le numéro 284 est souvent choisi pour être muté par les expérimentateurs. Cette mutation permet un déplacement bathochrome de la longueur d'onde d'émission dans le cas des mutations S284T ou S284N (sérine en thréonine ou asparagine)<sup>9</sup>. Dans le cas de S284T, la fonction alcool reliée au réseau est conservé mais l'hydrogène situé sur le carbone adjacent est remplacé par un groupement méthyl. Pour S284N le groupement alcool est remplacé par une fonction amide. On obtient une valeur expérimentale du  $T_e$  de 2.05 eV (605 nm) dans les protéines mutées S284T et S284N contre 2.22 eV (558 nm) pour la protéine 4G37 contenant la sérine. Au niveau théorique on a comparé les valeurs des  $T_e$  issues de structures avec en position 284 trois résidus différents, la sérine, présente dans la protéine naturelle, la thréonine et l'alanine qui est un résidu apolaire. Dans le cas des deux résidus polaires, la sérine et la thréonine on observe peu de différences au niveau du réseau de liaisons hydrogène. Les transitions électroniques  $T_e$  obtenues sont également peu différentes aussi bien au niveau TD-DFT/MM que MS-CASPT2/MM. Par exemple, pour le substrat oxyluciférine, la valeur du  $T_e$  obtenue dans la protéine 4G37(avec la sérine 284) est de 2.20 eV (564 nm) contre 2.15 eV (577 nm) dans la protéine avec S284T. Au niveau expérimental cette différence est plus importante, 2.22 (558 nm) eV contre 2.04 (608 nm) eV. Cependant, la tendance observée dans les calculs suit la logique des résultats expérimentaux. Lorsqu'on mute la sérine en thréonine on observe au mieux un déplacement dans le rouge de l'émission (pour OxyLH<sub>2</sub>) et au pire la même valeur en terme d'émission (OxyLH<sub>2</sub>-E-2). Ces résultats sont cependant peu concluants, il reste difficile d'extrapoler sur les causes possibles de ces changements d'émission.

Dans le cas de l'alanine, le réseau de liaisons hydrogène disparaît au niveau de ce résidu car il est de nature apolaire. De plus, on note que le nombre de molécules d'eau en bordure de la protéine est réduit car ces molécules ne bénéficient plus du réseau de liaisons hydrogène relayé par le résidu 284. On observe une légère diminution en terme de  $T_e$  entre les calculs dans la protéine contenant la sérine et la protéine avec l'alanine, passant de 1.87 eV (663 nm) à 1.89 eV (655 nm) pour OxyLH<sub>2</sub>-E-2. Pour le substrat OxyLH<sub>2</sub> la mutation S284N entraîne un déplacement vers le rouge de l'émission, celle-ci évoluant de 2.17 eV (573 nm) pour la protéine mutée S284N à 2.22 eV (558 nm) pour la protéine non mutée (sérine). Ces résultats ne suivent pas une tendance précise mais montrent que la modification du réseau de liaisons hydrogène a une influence sur la géométrie de la cavité et sur la transition électronique  $T_e$ .

Il reste encore difficile de quantifier l'influence de la mutation d'un résidu sur la valeur de l'émission de la bioluminescence. Plusieurs problèmes dans notre protocole peuvent expliquer ces difficultés. Dans un 1<sup>er</sup> temps, la mutation de résidu par analogie reste critiquable malgré une MD réalisée par la suite. En effet, la mutation du résidu peut entraîner des changements structuraux de la protéine comme le déplacement d'un boucle ou d'une hélice alpha qu'il est difficile de prédire au niveau calculatoire lors de simulations courtes. L'obtention de la structure cristallographique de la protéine reste le meilleur moyen de prédire le comportement d'une mutation. Malheureusement, il est difficile d'obtenir cette structure RX. Un autre moyen

Tableau 9.7 – Contribution de l'effet électrostatique sur les substrats OxyiLH<sub>2</sub>-E-1 et OxyiLH<sub>2</sub>-E-2 dans la protéine 4G37 et la protéine 4G37-S284T. *Single point* TD-DFT obtenu avec une base 6-311G(2d,p) dans le vide sur un structure optimisée géométriquement en TD-DFT/MM avec la même fonctionnelle et base.

	OxyiLH <sub>2</sub> -E-1 optimisé dans 4G37	OxyiLH <sub>2</sub> -E-1 optimisé dans 4G37-S284T	OxyiLH <sub>2</sub> -E-2 optimisé dans 4G37	OxyiLH <sub>2</sub> -E-2 optimisé dans 4G37-S284T
$T_e^{vac}$ eV	2.16	2.13	2.00	1.98
$T_e$ eV	2.00	1.98	1.87	1.87
$T_e^{vac}-T_e$ (eV)	+0.16	+0.15	+0.13	+0.11
$T_e^{vac}-T_e$ (nm)	-42	-44	-43	-37

d'améliorer nos résultats est de réaliser une longue dynamique de l'ordre de 100 ns et de sélectionner plusieurs *images* issues de la MD et de calculer la valeur du  $T_e$ . Cependant l'obtention au niveau théorique du  $T_e$  demande un temps de calcul important, il est donc difficile de traiter de nombreuses *images*. En outre, malgré le fait de sélectionner plusieurs structures, le risque d'étudier un artefact de la MD reste grand.

### 9.3.5 L'impact de l'environnement protéique

Les calculs QM/MM permettent de connaître et d'identifier les interactions de la protéine influençant la transition électronique  $T_e$  du substrat. Ces interactions peuvent être classées sous deux types : les interactions électrostatiques et les contraintes géométriques. On a vu précédemment que la mise en place d'un réseau de liaisons hydrogène entre le substrat et la protéine entraîne un déplacement vers le bleu du  $T_e$ . Cette propriété se classe parmi les interactions électrostatiques. Il est également possible de quantifier l'effet global de la protéine sur le  $T_e$ . Pour cela le substrat est gelé avec sa conformation obtenue en QM/MM dans son 1<sup>er</sup> état excité. On retire ensuite la partie MM du modèle : la protéine, l'AMPH et les molécules d'eau. Enfin un calcul TD-DFT sans optimisation est réalisé permettant d'obtenir le  $T_e^{vac}$ , la transition énergétique entre S1 et S0 dans le vide. Finalement la différence entre  $T_e$  et  $T_e^{vac}$  permet de quantifier l'effet de l'interaction électrostatique de la protéine.

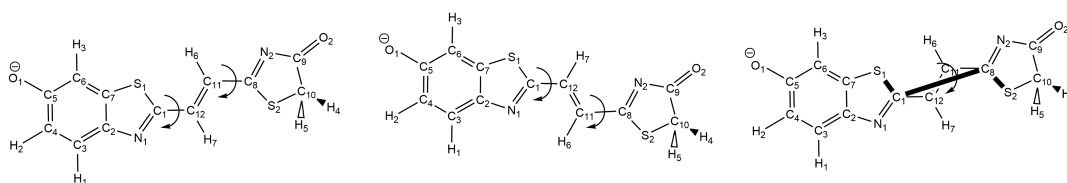


FIGURE 9.6 – Structure de OxyiLH<sub>2</sub>-E-1 (à gauche) définissant les dièdres  $\beta_{1-12}$  et  $\beta_{11-8}$ ; Structure de OxyiLH<sub>2</sub>-E-2 (au milieu) définissant les dièdres  $\beta_{1-12}$  et  $\beta_{11-8}$ ; Structure de OxyiLH<sub>2</sub>-E-1 (à droite) définissant le dièdre  $\beta_{2cycles}$ .

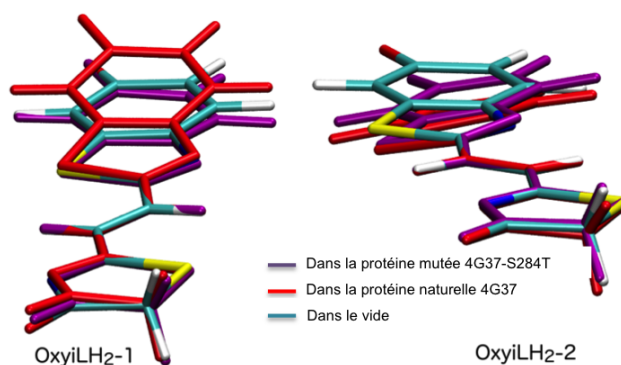


FIGURE 9.7 – Superposition géométrique des structures OxyiLH<sub>2</sub>-E-1 (à gauche) et OxyiLH<sub>2</sub>-E-2 (à droite) dans le vide et dans les protéines 4G37 et 4G37-S284T. La superposition est réalisée le long des atomes C11 et C12 du pont entre les deux cycles. Le code couleur est le suivant : géométrie obtenue dans le vide cyan ; dans la protéine 4G37 rouge ; dans la protéine 4G37-S284T violet.

Comme on peut le voir dans le tableau 9.7 l'impact de la protéine sur le  $T_e$  se traduit par un déplacement vers le rouge de l'émission compris entre 37 nm à 44 nm.

Concernant la géométrie des substrats, on observe peu de différence en terme de distance entre les liaisons de type C-O, aussi bien dans le vide que dans la protéine. En outre, les substrats optimisés dans le vide ont une géométrie plane alors que ceux obtenus dans la protéine possèdent plusieurs angles de torsion. Ainsi pour OxyLH<sub>2</sub>, on observe une torsion de l'ordre de 3° entre les deux cycles dans la protéine. Pour OxyiLH<sub>2</sub>-E-1 et OxyiLH<sub>2</sub>-E-2, 3 angles dièdres sont étudiés et représentés dans la figure 9.6. L'angle  $\beta_{1-12}$  correspond à la torsion entre le cycle benzothiazole et la double liaison C=C du pont central alors que l'angle  $\beta_{11-8}$  correspondent à la torsion entre l'autre double liaison C=C du pont et le cycle thiazole. Le dernier dièdre  $\beta_{2cycles}$  mesure la valeur de la torsion entre les deux cycles.

Pour OxyiLH<sub>2</sub>-E-1, les deux angles  $\beta_{1-12}$  et  $\beta_{11-8}$  sont d'intensités comparables mais avec des valeurs faibles comprises entre 1,5° et 5°. Pour le dièdre  $\beta_{2cycles}$ , l'angle observé est plus faible que pour OxyLH<sub>2</sub>.

Cependant pour OxyiLH<sub>2</sub>-E-2, les déformations observées sont plus importantes. Ainsi  $\beta_{1-12}$  est de l'ordre de 9°. Quant au dièdre  $\beta_{2cycles}$  il atteint une valeur de 19°.

Sur la figure 9.7 on a représenté pour chaque substrat la conformation adoptée dans le vide, dans la protéine naturelle et dans la protéine mutée S284T. Les deux substrats OxyiLH<sub>2</sub>-E-1 et OxyiLH<sub>2</sub>-E-2 adoptent ainsi un comportement différent dans la protéine. Pour OxyiLH<sub>2</sub>-E-1, on observe une distorsion du plan contenant le cycle benzothiazole alors que OxyiLH<sub>2</sub>-E-2, on observe une rotation autour de la liaison C1-C12.

Malgré ces contraintes géométriques, la valeur du  $T_e$  se trouve peu influencée. Pour OxyiLH<sub>2</sub>-E-1, la différence entre le  $T_e^{vac}$  et le  $T_e$  du substrat optimisé dans le vide est de -8 nm (+0.03 eV) pour la protéine naturelle et de 0 nm pour la protéine mutée. On peut expliquer ces valeurs par le fait que les torsions faibles affectant le substrat ne modifient pas la densité électronique des orbitales HOMO et LUMO. La valeur du

$T_e^{vac}$  est en conséquence inchangée par rapport à  $T_e$ . Pour OxyiLH<sub>2</sub>-E-2, la présence de plus fortes contraintes modifient légèrement la position énergétique des orbitales HOMO et LUMO. En effet, la torsions autour de C1-C12 déstabilise l'orbitale LUMO. Cela se traduit par un déplacement vers le rouge de +9 nm (-0.03 eV) entre le  $T_e^{vac}$  et le  $T_e$  du substrat optimisé dans le vide pour la protéine naturelle. Pour la protéine mutée cet écart augmente jusqu'à +15 nm (-0.05 eV). L'ensemble de ces valeurs sont rassemblées dans le tableau 9.8.

En étudiant le substrat présent dans la cavité, on observe deux types de torsions pouvant aussi bien engendrer un déplacement vers le bleu qu'un déplacement vers le rouge du  $T_e$ . Malheureusement il est difficile de rationaliser comment la cavité peut imposer des contraintes différentes sur des substrats ayant une géométrie proche.

L'effet de la géométrie sur le  $T_e$  du substrat reste faible, au minimum 3 fois moins importante que l'effet électrostatique de la protéine. Chercher à construire des substrats jouant sur l'influence de la géométrie semble une stratégie peu efficace.

Tableau 9.8 – Paramètres géométriques des deux substrats OxyiLH<sub>2</sub>-E-1 et OxyiLH<sub>2</sub>-E-2 optimisés dans la protéine ou dans le vide (optimisation réalisée avec la fonctionnelle B3LYP et la base 6-311G(2d,p)).  $T_e^{vac}$  réalisée sur ces structures (fonctionnelle B3LYP et la base 6-311G(2d,p)).

	OxyiLH <sub>2</sub> -E-1 optimisé			OxyiLH <sub>2</sub> -E-2 optimisé		
	dans 4G37	dans 4G37-S284T	dans le vide	dans 4G37	dans 4G37-S284T	dans le vide
C <sub>5</sub> -O <sub>1</sub> <sup>a</sup>	1.25	1.26	1.25	1.25	1.26	1.25
C <sub>9</sub> -O <sub>2</sub> <sup>a</sup>	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> Longueur de liaison en Å ; <sup>b</sup> Mesure d'angle en degré

<sup>c</sup> Angle dièdre entre S1-C1-C12-C11 <sup>d</sup> Angle dièdre entre N1-C1-C12-C11

<sup>e</sup> Angle dièdre entre C12-C11-C8-S2 <sup>f</sup> Angle dièdre entre C12-C11-C8-N2

<sup>g</sup> Angle dièdre entre S1-C1-C8-S2 <sup>h</sup> Single point dans le vide  $T_e$  en eV

### 9.3.6 L'étude conformationnelle

Les calculs présentés jusqu'à maintenant concernent deux conformères (E) du substrat OxyiLH<sub>2</sub>. Ces deux conformères présentent des comportements différents aussi bien dans le vide qu'à l'intérieur de la protéine. Dans les calculs effectués dans le vide ou dans un solvant modélisé de façon implicite, le substrat OxyiLH<sub>2</sub>-E-2 possède toujours une valeur du  $T_e$  plus faible que OxyiLH<sub>2</sub>-E-1. Cette différence est autour de 0.1 eV. Ainsi la différence géométrique entre ces deux conformères a une influence importante sur le  $T_e$  en solution ou dans le vide.

A l'intérieur de la protéine, les deux conformères adoptent une position spatiale dans la cavité relativement proche. Cependant comme évoqué dans le point précé-

Tableau 9.9 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. TD-DFT/MM avec une base 6-311G(2d,p) sur une structure optimisée dans l'état S1 avec la fonctionnelle indiquée.

Modèle	CAM-B3LYP	B3LYP	Exp.
Ser_OxyiLH <sub>2</sub> -E-1	2.23 (554) f=1.13	2.00 (616) f= 0.53	1.85 (670)
Ser_OxyiLH <sub>2</sub> -E-2	2.10 (590) f=1.13	1.87 (663) f=0.53	
Ser_OxyiLH <sub>2</sub> -Z-1	2.13 (579) f=0.67	1.88 (657) f=0.33	
Ser_OxyiLH <sub>2</sub> -Z-2	2.18 (567) f=0.72	1.89 (654) f=0.36	
Thr_OxyiLH <sub>2</sub> -E-1	2.24 (552) f=1.12	1.98 (625) f=0.47	1.75 (706)
Thr_OxyiLH <sub>2</sub> -E-2	2.16 (574) f=1.12	1.87 (661) f=0.47	
Thr_OxyiLH <sub>2</sub> -Z-1	2.15 (575) f=0.67	1.94 (640) f=0.31	
Thr_OxyiLH <sub>2</sub> -Z-2	2.20 (564) f=0.72	1.97 (627) f=0.41	

dent, ces deux molécules subissent des contraintes géométriques très différentes venant de la protéine. Le conformère OxyiLH<sub>2</sub>-E-2 se retrouve beaucoup plus affecté, entraînant une torsion importante autour de la liaison C1-C12. Au niveau de l'émission, la différence d'énergie entre les  $T_e$  s'accroît un peu, à 0.13 eV pour la protéine naturelle et 0.11 eV pour la protéine mutée. Cette faible différence s'explique par des effets électroniques et géométriques en opposition l'un par rapport à l'autre. Par exemple, dans la protéine mutée, OxyiLH<sub>2</sub>-E-1 subit une interaction électrostatique de la protéine plus forte que pour OxyiLH<sub>2</sub>-E-2. Cependant, OxyiLH<sub>2</sub>-E-2 est plus impacté au niveau de son  $T_e$  par les contraintes géométriques.

La différence de  $T_e$  s'explique néanmoins majoritairement par la nature du conformère.

La conformation du substrat est importante dans le calcul de la longueur d'onde de bioluminescence. Par des calculs QM/MM, l'un des conformères montre un meilleur accord avec les résultats expérimentaux pour le  $T_e$ . OxyiLH<sub>2</sub>-E-2 semble être le seul conformère participant à la réaction de bioluminescence pour l'infraoxyluciférine. Les conformères (Z) OxyiLH<sub>2</sub>-Z-1 et OxyiLH<sub>2</sub>-Z-2 ont également été testés et les résultats présentés dans le tableau 9.9. L'hypothèse que le substrat émetteur soit un conformère (Z) semble plus difficile à rationaliser que pour un conformère de type (E). En effet, la présence d'un conformère (Z) dans la cavité de la protéine induit une importante déformation de celle-ci. Au niveau expérimental, il est compliqué de synthétiser uniquement un seul conformère. Il n'y a alors pas de données expérimentales concernant la comparaison entre ces deux conformères (E) et (Z). Les conformères (Z) ne semblent alors pas être de bons candidats pour expliquer la bioluminescence chez OxyiLH<sub>2</sub>. Il est important de noter qu'on n'exclue toutefois pas la possibilité aux deux conformères (Z) d'être des précurseurs de la bioluminescence. Les faibles preuves expérimentales ainsi que les calculs réalisés pendant cette thèse ne permettent pas de donner une réponse claire sur ces molécules.

## 9.4 Conclusion

L'étude de la modulation de la longueur d'onde de la bioluminescence chez les lucioles est un sujet riche et vaste. Par l'étude d'un analogue rouge de l'oxyluciférine on a pu proposer de nouvelles hypothèses qui gouvernent la modulation de couleur. Du point de vue du substrat lui-même, l'augmentation de la conjugaison du système  $\pi$  induit un déplacement vers le rouge de la valeur de l'émission. Ce résultat est souvent utilisé par les expérimentateurs afin de concevoir de nouveaux substrats. Par cette étude on a également montré que la conformation du substrat est très importante. Chez l'émetteur infraoxyluciférine seul l'un des deux conformères (E) semble être le précurseur de la bioluminescence.

On peut classer l'effet de l'environnement sur la longueur d'onde d'émission du substrat selon trois points différents. Le premier représente l'effet de la formation du réseau de liaisons hydrogène dans la cavité de la protéine. Des calculs de dynamique moléculaire classique ont permis de montrer que des molécules d'eau se positionnent à l'intérieur de la cavité de la luciférase créant un réseau de liaisons hydrogène entre le substrat et certains résidus de la protéine. La création de ce réseau induit un déplacement vers le bleu de l'émission. Le deuxième et troisième effet proviennent de l'interaction de la protéine avec le substrat. La protéine exerce ainsi une interaction électrostatique entraînant un déplacement vers le rouge de l'émission et une interaction géométrique, influant sur la géométrie du substrat mais ne modifiant pas son émission. Finalement si on tient compte de ces trois interactions, l'environnement protéique permet un déplacement vers le rouge de l'émission de bioluminescence.

Il reste néanmoins des effets encore inexplicables du point de vue de la chimie théorique. La mutation du résidu S284T a une influence notable sur l'émission au niveau expérimental. Cependant, nos calculs n'arrivent pas à reproduire ces changements. La position du résidu muté, plutôt éloignée du substrat rend la modélisation de son effet compliquée. L'utilisation de MD afin de prendre en compte l'impact de la mutation sur la structure protéique demande des ressources énormes pour un résultat pas toujours pertinent.

Ce chapitre a donné lieu à la publication d'un article dans un journal scientifique. Cet article se situe en fin de thèse, dans le chapitre Publication des Annexes.

## 9.5 Détails de calculs

Les calculs QM/MM sur les différents modèles permettent de calculer l'énergie d'émission  $T_e$ , soit l'énergie verticale de transition entre l'état excité S1 et l'état fondamental S0. Ces calculs sont précédés de dynamique moléculaire classique (MD). Nous présentons ci-dessous les détails techniques qui ont permis d'obtenir les résultats présentés dans ce chapitre.

### 9.5.1 Partie MM

Afin d'obtenir les modèles pour l'étude envisagée, on commence par préparer la partie MM. Nos modèles sont construits à partir de la structure cristallographique PDB 4G37 obtenue par le groupe du Prof. Branchini<sup>2</sup>. Cette structure représente le complexe luciférase-luciférine au moment de la coordination du dioxygène. Pour cela, la structure a été modifiée chimiquement en ajoutant un pont disulfure entre les résidus Isoleucine 108 et Tyrosine 447 afin de geler la protéine dans cette conformation. Le fichier PDB 4G37.pdb a été téléchargé à partir du site internet RSCB PDB. Les boucles manquantes de la protéine ont été rajoutées grâce au logiciel Disgro<sup>96</sup>. La molécule DLSA ou 5'-O-[N-(dehydroluciferyl)-sulfamoyl]adénosine a été remplacée par de l'adénosine monophosphate (AMP) et par OxyLH<sub>2</sub> ou OxyiLH<sub>2</sub>. Les deux molécules SO<sub>4</sub><sup>-</sup> ainsi que le résidu pontant noté XLX ont également été retirés. Le retrait de ce dernier a peu d'influence sur les calculs. En effet les dynamiques réalisées n'excèdent pas 10 ns, un temps suffisamment court pour éviter le déplacement de la boucle bloquée par le résidu XLX.

Les résidus ont été protonnés par le programme Leap provenant de la suite logiciel Amber 14<sup>55</sup>. Les cas discutables, notamment pour les résidus histidines ont été élucidés en calculant leur pKa avec le program H++<sup>95</sup>. Les histidines suivantes ont été doublement protonées : résidus 76, 171, 221, 310, 332, 419, 461 and 489. L'AMP (chargé -2) a été protoné en AMPH (chargé -1) et un ion chlorure a été rajouté afin d'obtenir une charge neutre pour le modèle.

Un premier modèle ainsi créé correspond alors à la 4G37 couplée avec OxyLH<sub>2</sub>, le substrat naturel. Il est nommé Ser\_OxyLH<sub>2</sub> (Substrat OxyLH<sub>2</sub> dans la protéine avec le résidu 284 correspondant à une sérine.) Les autres modèles explicités ci-après sont dérivés de celui-ci.

Les modèles comportant la protéine mutée (la mutation S284T), sont notés avec le préfixe Thr pour Thréonine. La mutation est effectuée à la main dans le fichier PDB. Il a été choisi d'orienter l'oxygène de la chaîne aliphatique vers le substrat. Il en découle que Thr\_OxyLH<sub>2</sub> correspond au substrat naturel dans la 4G37-S284T.

On s'intéresse maintenant aux modèles comportant OxyiLH<sub>2</sub>. Pour ce substrat nous avons étudié plusieurs conformères, tous présentés dans la figure 9.2. Les conformères possédant une double liaison en trans (E) sont à la fois privilégiés par les expérimentateurs et par des études de docking préliminaires. En effet, les conformères (E) s'intègrent mieux à l'intérieur de la cavité de la luciférase. Pour les deux conformères (E), l'un est nommé infra-oxyluciférine-1 (OxyiLH<sub>2</sub>-E-1) et l'autre infra-oxyluciférine-2 (OxyiLH<sub>2</sub>-E-2). Le changement de conformation s'effectue à la main en gelant le cycle thiazolone de OxyLH<sub>2</sub> puis en faisant les modifications nécessaires. Les deux conformères en cis (Z) sont créés à partir des deux conformères (E) correspondants. OxyiLH<sub>2</sub>-Z-1 est créé à partir de OxyiLH<sub>2</sub>-E-1 et OxyiLH<sub>2</sub>-Z-2 à partir de OxyiLH<sub>2</sub>-E-2. Pour cela on modifie la double liaison trans (E) des conformères en cis (Z) en gelant le cycle thiazolone.

Finalement les modèles comportant les substrats modifiés dans la protéine naturelle sont nommés Ser\_OxyiLH<sub>2</sub>-E-1, Ser\_OxyiLH<sub>2</sub>-E-2, Ser\_OxyiLH<sub>2</sub>-Z-1 et



Ser\_OxyiLH<sub>2</sub>-Z-2 et pour la protéine mutée Thr\_OxyiLH<sub>2</sub>-E-1, Thr\_OxyiLH<sub>2</sub>-E-2, Thr\_OxyiLH<sub>2</sub>-Z-1, Thr\_OxyiLH<sub>2</sub>-Z-2.

### 9.5.2 Partie MD

Les simulations de dynamiques moléculaires classiques ont été réalisées avec la suite logicielle Amber14<sup>55</sup> afin d'obtenir une représentation du système pour une future optimisation QM/MM. Les modèles ont été solvatés par des molécules d'eau de type TIP3P dans une boîte de forme cuboïde. Le système résultant contient environ 90000 atomes au total, dont 28000 molécules d'eau formant une sphère de solvation d'au moins 15 Å autour du substrat. Un champ de force de type AMBER99ff a été utilisé pour modéliser les résidus de la protéine. Les paramètres de OxyLH<sub>2</sub> ainsi que ceux de l'AMPH sont développés par notre groupe<sup>111</sup>. Ceux-ci n'étant pas optimisés pour des états excités, la structure et les charges des substrats au niveau excité S1 ont été obtenues par calcul QM/MM préliminaire puis réinjectés dans les fichiers paramètres. Le substrat reste gelé dans cette géométrie pendant toute la dynamique. Le déroulement d'une dynamique est le suivant : 3 minimisations sont d'abord réalisées, la 1<sup>ère</sup> minimise seulement la protéine en gelant les molécules d'eau. La deuxième minimisation ne gèle que les hydrogènes alors que rien n'est gelé (à part le substrat) durant la 3<sup>ème</sup>. Le système est ensuite chauffé de 100K à 300K en 20 ps. Puis sous condition NPT avec T=300K, P=1 atm une dynamique de 10 ns ou plus si indiqué est conduite sous condition périodique aux bords avec comme pas de temps 2 fs. La température et la pression sont maintenues en utilisant l'algorithme de Berendsen avec une constante de couplage de 5 ps et par des contraintes de type SHAKE appliquées sur toutes les liaisons impliquant des atomes d'hydrogène.

### 9.5.3 Partie QM/MM

Les calculs QM/MM sont réalisés en utilisant un schéma de couplage ESPF entre Gaussian<sup>67</sup> et Tinker<sup>56</sup> ou Molcas<sup>113</sup> et Tinker (Gaussian 09d/Tinker ou Molcas 8.0/Tinker). La méthode ESPF (electrostatical potentiel fitted)<sup>68</sup> permet de prendre en compte le potentiel électrostat externe créé par les points de charge du MM dans le calcul de la fonction d'onde de la partie QM. Un cutoff de 9 Å est pris pour éviter une sommation infinie. Au niveau MM, les charges de type Mulliken<sup>114</sup> de la partie QM sont prises en compte dans le calcul électrostatique. La technique dite de microitérations est utilisée afin de faire converger la géométrie de l'espace MM choisi pour chaque pas de minimisation du QM. L'émetteur (ou substrat) représente le QM et le reste du système (AMPH + protéine + molécules d'eau) constitue la partie traitée au niveau MM. Les calculs QM/MM suivent toujours la même procédure. Le premier calcul permet d'atteindre le minimum énergétique du système dans l'état fondamental (S0). Puis, on minimise le système dans son 1<sup>er</sup> état excité singulet (S1) avant le calcul de la transition énergétique vertical (T<sub>e</sub>) entre ce même état S1 et l'état fondamental. Cette transition correspond à une émission de fluorescence, cependant chez les lucioles l'énergie de fluorescence est identique à l'énergie de la transition après la réaction de bioluminescence<sup>19</sup>, le T<sub>e</sub> peut alors être assimilé à

l'énergie d'émission de la bioluminescence.

Le niveau de théorie utilisé pour traiter la partie QM correspond à des calculs TD-DFT ou MS-CASPT2.

On détaille dans un 1<sup>er</sup> temps les calculs QM/MM utilisant une méthode TD-DFT/MM. Les calculs TD-DFT ont recours à deux types de bases : une double dzéta 6-31G(d,p) et une triple dzéta 6-311G(2d,p). Ces dernières ne possèdent pas de fonctions diffuses associées car celles-ci peuvent interagir avec la partie MM lors de l'utilisation d'un électrostatique embedding ESPF<sup>68</sup>. Plusieurs fonctionnelles sont également étudiées. On teste 5 fonctionnelles B3LYP<sup>44,45,47</sup>, CAM-B3LYP<sup>48</sup>, M06<sup>42</sup>, M06-2X<sup>42</sup> et  $\omega$ B97XD<sup>115</sup>. Des calculs TD-DFT ont également été testés dans le vide ou en présence d'un solvant. Ces effets de solvants sont pris en compte en utilisant le Polarizable Continuum Model ou PCM<sup>73,74</sup>.

Dans un deuxième temps des calculs de type multiconfigurationnal reference second-order perturbation theory ou MS-CASPT2<sup>34</sup> sont également mis en œuvre afin d'obtenir la  $T_e$ . Pour cela des calculs *single points* MS-CASPT2/MM sont réalisés à l'aide d'une base triple dzéta ANO-RCC-VTZP<sup>53</sup> sur les géométries optimisées de l'état excité S1 obtenues avec des calculs CAM-B3LYP 6-31G(d,p)/MM. La géométrie du substrat évolue peu entre les différentes fonctionnelles et bases testées. Ainsi le Root-Mean-Square Deviation of atomic positions (RMSD) calculé pour toutes les géométries est inférieur à 0.02 Å.

Pour le substrat OxyLH<sub>2</sub>, 3 espaces actifs sont testés : le 18 électrons dans 15 orbitales (18-in-15) correspond à toutes les orbitales conjuguées de type  $\pi$  dans OxyLH<sub>2</sub> (voir Annexe A1). Puis, deux espaces actifs sont créés, un 16-in-14 et un 14-in-13 construits à partir du précédent en enlevant une ou deux orbitales  $\pi$ . Dans le cas du 16-in-14, on enlève l'orbitale centrée sur le soufre du cycle thiazolone. Pour le 14-in-13 on enlève les deux orbitales centrées sur les soufres des cycles thiazolone et benzothiazole. Dans le cas des substrats correspondant à OxyiLH<sub>2</sub> un espace actif 16-in-15 a été choisi (voir Annexe A2 et A3). Celui-ci comporte toutes les orbitales conjuguées  $\pi$  à l'exception des deux centrées sur les soufres, comme dans le cas du 14-in-13 pour OxyLH<sub>2</sub>. On étudie plusieurs paramètres relatifs au MS-CASPT2 comme le nombre d'états électroniques pris en compte dans le calcul, la valeur du level shift<sup>32</sup> ou la valeur du shift IPEA choisi.

# Chapitre 10

## Modélisation et compréhension de nouvelles luciférases

### 10.1 Contexte

Grâce à une équipe de collaborateurs de l'université de New York d'Abu Dhabi, deux nouvelles structures cristallographiques de luciférase ont été cristallisées. La première est issue de l'espèce *Amydetes viviani*, un lamproye proche des vers luisants et est nommée GB<sub>Av</sub> (Green blue shift luciferase from *Amydetes viviani*) (voir Figure 10.1). Cette espèce produit une bioluminescence légèrement déplacée vers le bleu de l'ordre de 20 nm (538 nm) par rapport à celle dans la luciférase des lucioles (558 nm). La deuxième structure est issue du *Phrixothrix hirtus*, une larve de la famille des coléoptères et donnant lieu à une émission dans le rouge à 1.98 eV (625 nm). Elle est nommée RE<sub>Ph</sub> (Red-emitting luciferase from *Phrixothrix hirtus*) (voir Figure 10.1).

Ces deux nouvelles structures possèdent également des propriétés inédites dans le



FIGURE 10.1 – Photos de l'espèce bioluminescente *Amydetes viviani* à gauche et de l'espèce bioluminescente *Phrixothrix hirtus* à droite. Crédits photos : Prof. Vadim Viviani.

domaine de la bioluminescence. La  $GB_{AV}$  a été cristallisée dans une conformation dite ouverte, le domaine C-terminal est tourné de telle façon à ce que les réactifs ont un accès direct au site actif. Quand à la  $RE_{Ph}$ , la structure cristallographique montre la mise en place d'une structure quaternaire de type homo-octamérique.

## 10.2 Etude de la $GB_{AV}$ , l'impact de la conformation du site actif

### 10.2.1 Présentation du système

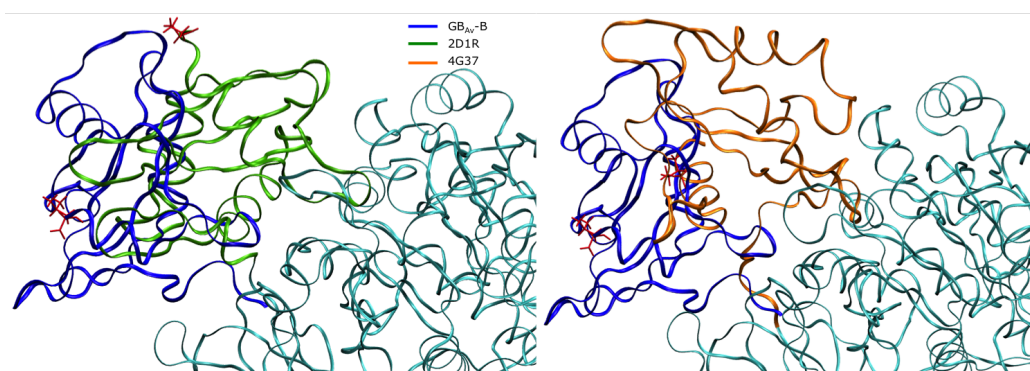


FIGURE 10.2 – Représentation de la conformation des domaines C-terminaux de la structure  $GB_{AV}$ -B et 2D1R à gauche et de la structure  $GB_{AV}$ -B et 4G37 à droite. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines. La protéine  $GB_{AV}$ -B est dessinée en ruban bleu, la 2D1R en vert et la 4G37 en orange. Les résidus colorés en rouge représentent les derniers résidus de la chaîne polypeptidique.

La découverte de la bioluminescence chez l'espèce *Amydetes viviani* date de quelques années<sup>17</sup>. Les expérimentateurs ont réussi très récemment à obtenir des cristaux de la luciférase de ce système et on fait appel à notre expertise avant de publier les structures. Ainsi, ils ont réussi à résoudre deux chaînes d'un même cristal par RX donnant donc lieu à deux conformations différentes. Cependant comme précisé dans la section contexte, le domaine C-terminal de ces cristaux possède une conformation distincte des autres structures cristallographiques obtenues jusqu'à maintenant chez les luciférases des lucioles (voir figure 10.2). Dans  $GB_{AV}$  pour Green Blue shift luciferase from *Amydetes viviani*, ce domaine est tourné de l'ordre de  $90^\circ$  par rapport aux structures naturelles de luciférases de lucioles que sont *Photonus Pyralis* (PDB 4G36) et *Luciola cruciata* (PDB 2D1R). Cette conformation induit plusieurs changements dans le comportement de la protéine. Le site actif de la luciférase est alors en interaction directe avec le solvant autour de la protéine. De plus, les deux substrats, l'oxyluciférine et l'AMPH ont la possibilité de s'échapper du site actif si ce déplacement est énergiquement favorable.

L'apport des structures cristallographiques se traduit également par le déchiffrement de la chaîne polypeptidique de la protéine. Une comparaison avec la structure naturelle 4G36 montre plusieurs substitutions pouvant être observées dans la structure primaire. Les trois substitutions A236S, F250S et I349C ont ainsi été mises en valeur par les expérimentateurs. Seule la deuxième est située à proximité du site actif et est alors étudiée en détail dans cette étude. Finalement les deux chaînes issues du cristal sont nommées GB<sub>AV</sub>-A pour la première et GB<sub>AV</sub>-B pour la deuxième. Ces deux structures ne contiennent ni le substrat luciférine ni la molécule d'AMP. Ces molécules sont alors dockées en vue de calculer la longueur d'onde d'émission de la bioluminescence.

Plusieurs questions nous ont été posées par nos collègues expérimentateurs : quel

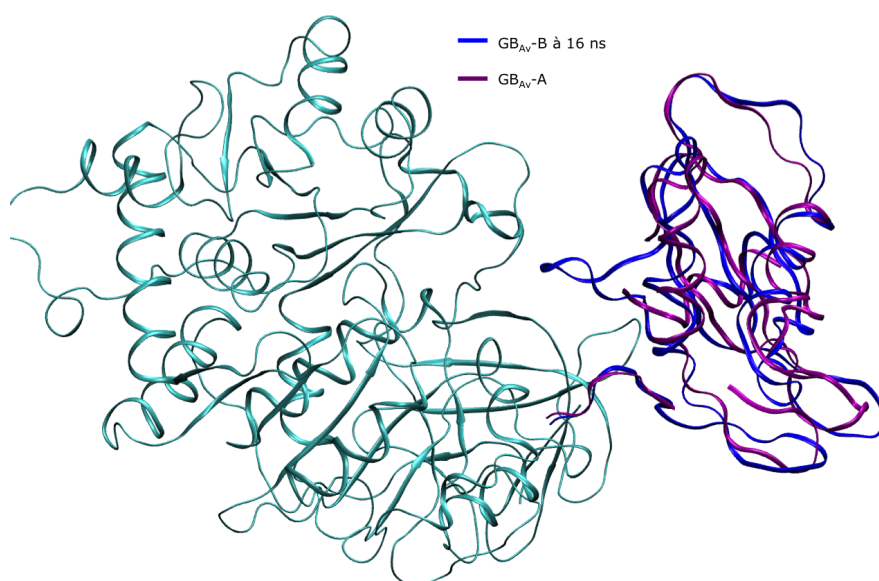


FIGURE 10.3 – Représentation des conformations des domaines C-terminaux de l’image obtenue à 16 ns de la MD sur GB<sub>AV</sub>-B et la structure cristallographique GB<sub>AV</sub>-A partielle. La protéine GB<sub>AV</sub>-B est dessinée en ruban bleu, et la GB<sub>AV</sub>-A partielle en violet. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines.

est le lien entre les deux structures cristallographiques ? Quelle est la conformation de la protéine lors de l’émission de lumière ? Peut-on modéliser correctement ce système afin de reproduire les valeurs d’émission expérimentales ?

### 10.2.2 Est-il possible de passer de GB<sub>AV</sub>-A à GB<sub>AV</sub>-B ?

Au début de la collaboration, les collaborateurs n’avaient pas totalement résolu la structure complète de GB<sub>AV</sub>-A. En effet, certaines boucles du domaine C-terminal étaient difficiles à définir. Les deux structures GB<sub>AV</sub>-A et GB<sub>AV</sub>-B sont issues du même cristal mais possèdent des conformations différentes. On se demande alors s’il est possible de passer d’une structure à l’autre par MD et si on peut reconstruire les boucles manquantes de la chaîne A grâce aux données cristallographiques de la

chaîne B.

Une dynamique moléculaire classique (MD) sur  $GB_{AV}$ -B sans substrat a été effectuée pour une durée de 80 ns. Dans cette dynamique on observe essentiellement un mouvement erratique du domaine C-terminal alors que le domaine N-terminal semble peu affecté. Cependant autour de 16 ns  $GB_{AV}$ -B semble se superposer de façon cohérente avec la structure cristallographique de  $GB_{AV}$ -A (voir figure 10.3). Une *image* est donc extraite de la dynamique moléculaire de  $GB_{AV}$ -B, minimisée de façon classique, et les deux boucles manquantes de  $GB_{AV}$ -A construites à la main. Pour cela, on superpose la structure  $GB_{AV}$ -B minimisée avec la structure cristallographique de  $GB_{AV}$ -A afin d'obtenir la position des carbones  $\alpha$ . Ces deux boucles protéiques correspondent aux résidus 435 à 436 et 519 à 524. La structure résultante finalement minimisée donne accès à une structure construite de  $GB_{AV}$ -A, notée  $GB_{AV}$ -A-construite. Cette dernière est comparée à la structure cristallographique de la chaîne A dans la figure 10.4.

Lorsque la structure RX de  $GB_{AV}$ -A a été totalement résolue par nos collègues

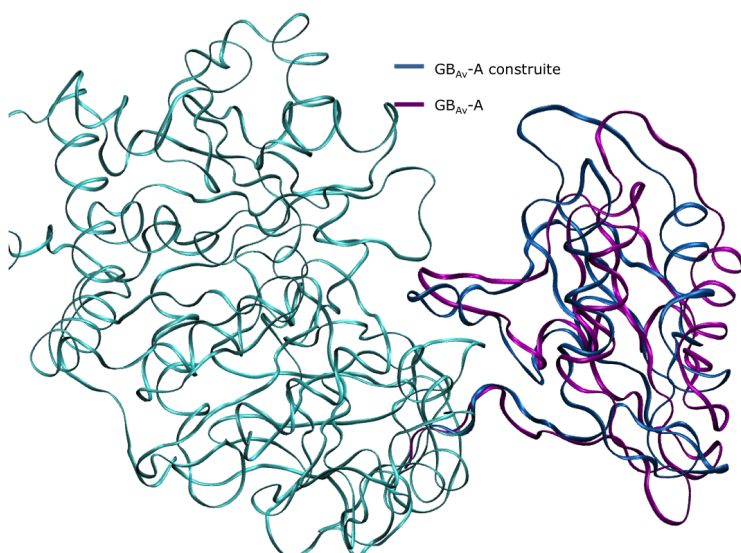


FIGURE 10.4 – Représentation des conformations des domaines C-terminaux de la structure construite de la  $GB_{AV}$ -A-construite et la structure cristallographique  $GB_{AV}$ -A complète. La structure construite de la  $GB_{AV}$ -A-construite est dessinée en ruban bleu de prusse, et la  $GB_{AV}$ -A en violet. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines.

expérimentateurs, nous avons pu la comparer à  $GB_{AV}$ -A-construite. Le RMSD est de 3.7 Å entre le domaine C-terminal construit et celui de la structure cristallographique ce qui est acceptable pour une construction de chaîne flexible.

Dans un second temps, une MD est été réalisée sur la structure  $GB_{AV}$ -A complète. Cette fois-ci,  $GB_{AV}$ -A se superpose à  $GB_{AV}$ -B après environ un temps de MD de 5 ns.

En conclusion, une première approche de l'étude de cette protéine montre que les deux chaînes cristallisées sont étroitement liées et qu'il est possible d'obtenir une modélisation correcte de l'une en partant de l'autre.

## 10.2.3 Calcul de la longueur d'onde de bioluminescence

### 10.2.3.1 Calcul de $T_e$ à partir des structures cristallographiques

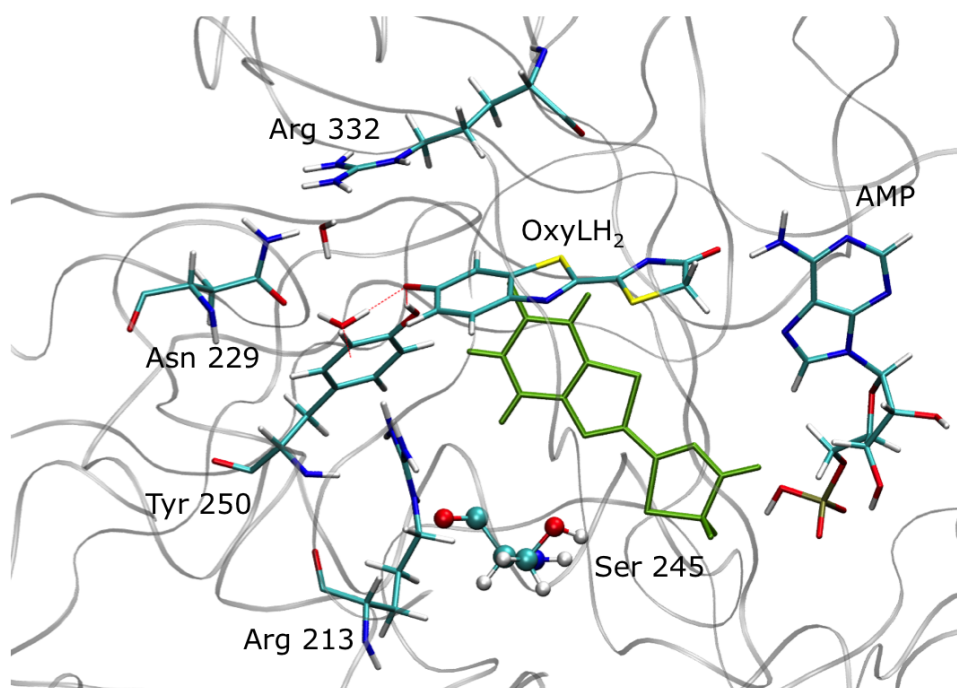


FIGURE 10.5 – Structure du site actif de la GB<sub>Av</sub>-B-OxyLH<sub>2</sub> de l'*image* (2) du tableau 10.1, superposée avec la structure du système de la luciole 2D1R. Les structures sont comparées en superposant leur domaines N-terminaux. L'OxyLH<sub>2</sub> dans l'*image* (2) du tableau 10.1 est représentée en tout atome. De la structure modèle 2D1R, seule OxyLH<sub>2</sub> est représentée en vert pour plus de clarté. Le résidu Ser 245, représenté en CPK correspond à une substitution entre 2D1R (Phe) et GB<sub>Av</sub> (Ser).

Dans un second temps on cherche à reproduire les résultats expérimentaux ainsi qu'à expliquer le déplacement vers le bleu de l'émission de GB<sub>Av</sub> par rapport aux systèmes des lucioles. Comme les deux structures GB<sub>Av</sub>-A et GB<sub>Av</sub>-B sont finalement très proches on ne s'intéresse plus particulièrement qu'à une seule des deux, GB<sub>Av</sub>-B. La première étape consiste à docker l'oxyluciférine et l'AMPH à l'intérieur du site actif. Un docking manuel est utilisé, et les réactifs sont disposés au même endroit que dans la luciférase naturelle 4G36. Le système est alors minimisé au niveau MM puis une MD de 10 ns est réalisée. Dans cette dernière, 5 *images* sont extraites, minimisées au niveau MM et le calcul de la longueur d'onde d'émission de fluorescence  $T_e$  effectué selon la méthodologie décrite dans 10.5.

Tableau 10.1 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B-open. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé dans 4G37 est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que les valeurs expérimentales issues des références<sup>172</sup>.

Modèle	$T_e$ avec B3LYP eV (nm)	SP in vacuo $T_e^{vac}$ eV (nm)	Exp. eV (nm)
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> (1)	2.24 (552)	2.24 (551)	
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> (2)	2.3 (538)	2.25 (550)	
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> (3)	2.24 (553)	2.25 (549)	2.3 (538)
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> (4)	2.29 (541)	2.24 (554)	
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> (5)	2.25 (551)	2.24 (554)	
4G37-OxyLH <sub>2</sub>	2.23 (555)	2.24 (553)	2.22 (558)

Plusieurs enseignements peuvent être tirés de la MD et des valeurs du  $T_e$ . L'étude de la MD montre que l'oxyluciférine se déplace à l'intérieur de la cavité lors de la simulation. Le cycle thiazolone pivote et se rapproche de la partie adénosine de l'AMPH alors que l'oxygène O1 du cycle benzothiazole bouge peu et sert de point de pivot (voir figure 10.5). Ce mouvement de l'oxyluciférine n'est pas un artéfact et a été observée dans 5 MD avec des conditions initiales différentes. L'oxyluciférine adopte alors une position différente dans la cavité de la luciférase GB<sub>Av</sub> en comparaison avec les luciférases issues des lucioles. Durant les MD, l'AMPH quand à lui ne bouge pas. Malgré la conformation ouverte de la protéine, le nombre de liaisons hydrogène formées entre l'oxyluciférine et les molécules environnantes (eau+résidus) est similaire au nombre observé dans les luciférases avec le domaine C-term fermées (autour de 3). Cependant, cette même conformation ouverte induit un nombre de molécules d'eau plus importante près du site actif, il y également plus de molécules d'eau dans le site actif. On peut alors même considérer que l'AMPH est à la limite entre la protéine et le solvant.

Le calcul des  $T_e$  au niveau TD-DFT/MM donne des résultats mitigés (voir tableau 10.1). Dans certains cas, la valeur est proche de la valeur expérimentale chez GB<sub>Av</sub> autour de 538 nm (*images* 2 et 4). Pour les autres *images*, les valeurs calculées se rapprochent de la valeur expérimentale observée chez les luciférases des lucioles *Photonus Pyralis* et *Luciola cruciata* autour de 558 nm (*images* 1, 3 et 5). Le calcul du  $T_e^{vac}$  montre que la géométrie de l'oxyluciférine ne varie pas ou peu lorsqu'elle est située à l'intérieure de la luciférase 4G37 ou la luciférase GB<sub>Av</sub>-B malgré une position différente dans la cavité et donc une interaction géométrique issue des résidus environnants modifiée (voir figure 10.6). Les différences les plus importantes se situent au niveau de l'angle de torsion entre les deux cycles, décrit ici selon le dièdre S-C-C-S. Cet angle est de l'ordre de 4° dans le modèle 4G37-OxyLH<sub>2</sub>. Dans les 4 *images* cet angle de OxyLH<sub>2</sub> est compris entre 2° et 5°. Cependant pour l'*image* 5,



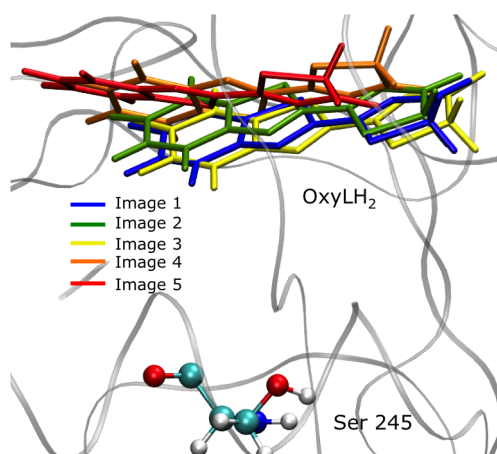


FIGURE 10.6 – Structure du site actif de la GB<sub>Av</sub>-B-OxyLH<sub>2</sub> dans les différents *images* (1 à 5). Les domaines N-terminaux ont été superposés. Le substrat OxyLH<sub>2</sub> est représenté dans une couleur dépendant de l'*image* référencée dans le tableau 10.1. En bleu, le substrat de l'*image* (1), en vert le (2), en jaune le (3), en orange le (4) et en rouge le (5). Le résidu Ser245 est mis en valeur par une représentation tout atome.

cet angle est de 10°. Cet angle n'a cependant pas d'influence sur le  $T_e$  puisque toutes les valeurs des  $T_e^{vac}$  calculées sont similaires.

### 10.2.3.2 Comprendre le déplacement vers le bleu

Une explication du déplacement vers le bleu de l'émission serait due à la présence de la substitution du résidu numéro 250 qui est une phénylalanine dans la protéine 4G37 à une sérine 245 dans GB<sub>Av</sub>. Le décalage de numérotation vient du fait que les deux chaînes n'ont pas la même séquence mais la position de Phe 250 de 4G37 correspond à celle de Ser 245 de GB<sub>Av</sub> dans la structure 3D des deux protéines superposées. Cette substitution est la seule située au niveau du site actif et transforme un résidu apolaire en résidu polaire. L'hypothèse avancée par nos collaborateurs est que la présence de la sérine induit la formation d'un réseau de liaisons hydrogène entre l'atome N1 de l'oxyluciférine, une molécule d'eau et le résidu Ser 245. L'effet électrostatique issu de ce réseau et modifiant la densité électrostatique de l'oxyluciférine serait responsable d'un déplacement vers le bleu de la valeur de l'émission.

Nous avons analysé et comparé les MD réalisées sur GB<sub>Av</sub>-B-OxyLH<sub>2</sub> afin d'observer la formation du réseau de liaisons hydrogène. Malheureusement la formation de ce réseau n'est pas observée lors des différentes MD. Il alors été convenu d'ajouter artificiellement une molécule d'eau entre l'oxyluciférine et le résidu 250. Ce modèle se nomme GB<sub>Av</sub>-B-OxyLH<sub>2</sub>+H<sub>2</sub>O. La MD réalisée par la suite sur ce modèle montre que la molécule d'eau ajoutée n'est pas stable et quitte rapidement la position centrale entre le substrat et la sérine. A la fin de la MD, la molécule d'eau se stabilise à proximité du cycle benzénique de l'oxyluciférine (voir figure 10.7). Les

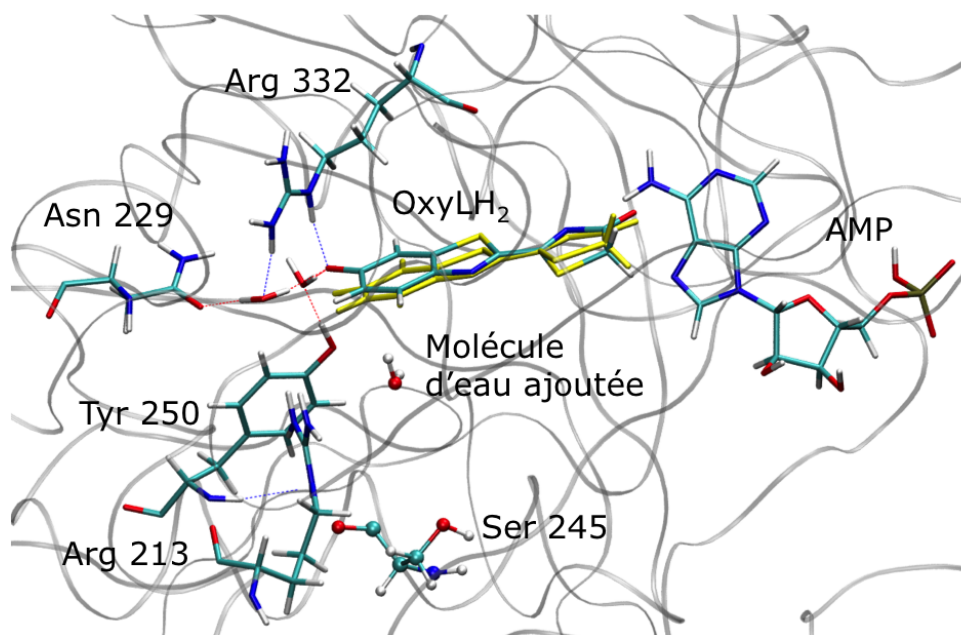


FIGURE 10.7 – Structure du site actif de la  $GB_{Av}$ -B-OxyLH<sub>2</sub>+H<sub>2</sub>O, superposée avec la structure du système  $GB_{Av}$ -B-OxyLH<sub>2</sub>. La comparaison entre les conformations de OxyLH<sub>2</sub> dans l'*image* (2) du tableau 10.2, est représentée en tout atome et OxyLH<sub>2</sub> issu de l'*image* (3) du tableau 10.1 dans le modèle  $GB_{Av}$ -B-OxyLH<sub>2</sub> en jaune. Les structures sont comparées en superposant leur domaines N-terminaux. Le résidu Ser 245 et la molécule d'eau ajoutée artificiellement sont représentés en CPK.

2 *images* extraites de la MD montrent une valeur de  $T_e$  fortement impactée avec un large déplacement de l'émission vers le bleu. La valeur de l'émission dans la deuxième *image* passe même sous la barre des 500 nm (voir tableau 10.2). Ainsi, même si les résultats montrent un déplacement vers le bleu de l'émission, ceux-ci ne sont pas dus à la présence d'une liaison hydrogène entre l'oxyluciférine et le résidu Ser 245. Concernant la géométrie des substrats, il n'y pas de différence avec les géométries sans la molécule d'eau, les valeurs des  $T_e^{vac}$  calculées étant les mêmes.

En conclusion, la présence du résidu sérine 245 chez  $GB_{Av}$  ne permet pas d'expliquer pourquoi cette protéine induit un déplacement vers le bleu de l'émission.

En parallèle, la mutation *in silico* du résidu F250S a été réalisée dans la protéine 4G37. La MD réalisée montre l'apparition de molécule d'eau dans la cavité de la protéine. En outre, après 2 ns, une molécule d'eau se place entre l'atome N1 de l'oxyluciférine et le résidu S250 permettant la création d'un réseau de liaisons hydrogène. Les  $T_e$  calculées sur 2 *images* différentes dévoilent étonnamment un déplacement de l'émission vers le rouge au lieu du déplacement vers le bleu attendu (voir tableau 10.2).

Tableau 10.2 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B-OxyLH<sub>2</sub>+H<sub>2</sub>O et 4G37-OxyLH<sub>2</sub>-F250S. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé dans 4G37 est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que les valeurs expérimentales issues des références<sup>172</sup>.

Modèle	B3LYP eV (nm)	SP in vacuo $T_e^{vac}$ eV (nm)	Exp. eV (nm)
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> +H <sub>2</sub> O (1)	2.36 (526)	2.25 (550)	2.3 (538)
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> +H <sub>2</sub> O (2)	2.49 (498)	2.27 (547)	
4G37-OxyLH <sub>2</sub>	2.23 (555)	2.24 (553)	2.22 (558)
4G37-OxyLH <sub>2</sub> -F250S (1)	2.13 (581)	2.2 (564)	
4G37-OxyLH <sub>2</sub> -F250S (2)	2.19 (566)	2.25 (552)	

### 10.2.3.3 Conclusion

En résumé, la présence de Ser 245 dans GB<sub>Av</sub>-B ne provoque pas de façon claire la formation de liaisons hydrogène avec l'oxyluciférine. De plus, lorsqu'on mute artificiellement le résidu phénylalanine 250 en sérine dans la structure 4G37, le déplacement observé est vers le rouge et une molécule d'eau forme un réseau de liaisons hydrogène entre le résidu 250 et l'oxyluciférine. La présence de ce résidu Ser 245 dans la GB<sub>Av</sub> ne permet pas d'expliquer le déplacement vers le bleu de l'émission observée pour *Amydetes viviani*. Néanmoins, on peut noter que la protéine force l'oxyluciférine à adopter une position à l'intérieur de la protéine différente de celle observée chez les lucioles. De plus, au vu des valeurs du  $T_e$  calculées, on peut voir que cette nouvelle position affecte la longueur d'onde de la bioluminescence, provoquant un déplacement vers le bleu de l'émission.

### 10.2.4 Le repliement de la protéine

La conformation adoptée par GB<sub>Av</sub> dans les structures cristallographiques ne représente peut-être qu'une structure de transition dans la réaction de bioluminescence. La conformation fermée, c'est à dire avec le domaine C-terminal accolé au N-terminal de la luciférase est considéré comme la conformation permettant la réaction. On cherche à replier le domaine C-terminal en vue de reproduire cette conformation fermée. Le repliement du C-terminal se fait au moyen d'une rotation autour d'une charnière de la luciférase, située dans le domaine N-term. Cependant, il a été montré expérimentalement que l'insertion d'une boucle du C-terminal (résidu 519 à 524) dans le N-terminal lors du repliement est primordial dans la réaction de bioluminescence. En effet, cette boucle pénètre dans le site actif de la luciférase et crée plusieurs liaisons faibles (liaisons hydrogène) entre la Lysine 524, l'AMPH et la Glycine 311. L'une des contraintes de la modélisation du repliement de la protéine est de réussir à reproduire l'insertion de cette boucle.

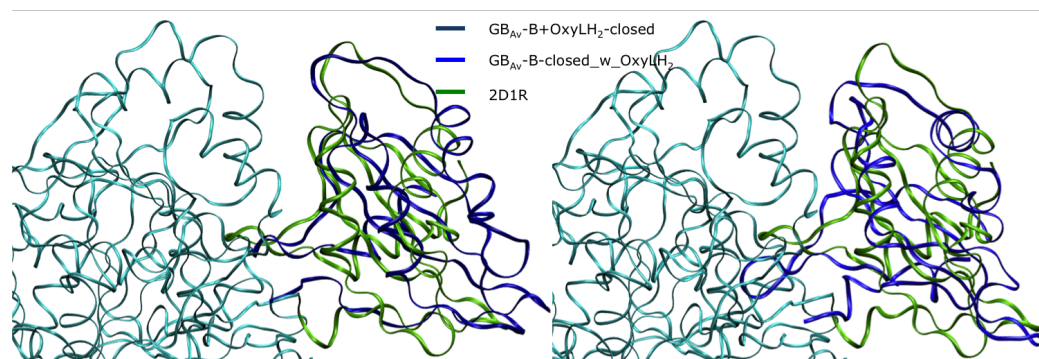


FIGURE 10.8 – Comparaison entre les domaines C-terminaux de la structure  $GB_{Av}-B+OxyLH_2$ -closed et 2D1R à gauche et de la structure  $GB_{Av}-B$ -closed\_w\_OxyLH<sub>2</sub> et 2D1R à droite. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines. La protéine  $GB_{Av}-B+OxyLH_2$ -closed est dessinée en ruban bleu foncé, la 2D1R en vert et la  $GB_{Av}-B$ -closed\_w\_OxyLH<sub>2</sub> en bleu roi.

En vue de réaliser ce repliement, deux structures de départ peuvent être considérées. Dans la première, on part d'une *image* obtenue lors d'une MD réalisée sur  $GB_{Av}-B$  contenant l'oxyluciférine et l'AMPH. Ainsi, l'oxyluciférine et l'AMPH sont déjà présentes à l'intérieur de la cavité, mais leur positionnement est différent de ceux adoptés dans les structures naturelles des lucioles. La deuxième structure de départ correspond à la structure cristallographique  $GB_{Av}-B$  sans substrats à l'intérieur. On réalise le repliement, puis on docke l'oxyluciférine et l'AMPH dans la cavité à la position qu'elles adoptent dans les luciférase des lucioles. Dans ce cas, il est plus difficile de comprendre comment les substrats ont pu intégrer le site actif, cependant la position des deux substrats est la même que dans les structures des luciférase naturelles des lucioles.

Le repliement du C-terminal est réalisé avec une simulation de type umbrella sampling. Dans cette dernière, la coordonnée de réaction qui est contrainte pendant la simulation correspond à la distance entre les deux carbones  $\alpha$  des résidus Gly 311 et Lys 524. On réalise une simulation umbrella sampling pour une distance comprise entre 21 Å et 7 Å. Le pas utilisé pour les différentes fenêtres est de 1 Å. Des informations complémentaires peuvent être trouvées dans la section 10.5.

Le calcul de l'énergie libre lors des simulations umbrella sampling montre qu'il n'y a pas de barrière énergétique permettant de passer de la conformation ouverte à la conformation fermée. Les deux conformations finales obtenues se rapprochent suffisamment de la conformation fermée de la structure 4G36 de la luciole pour être pertinentes. Les deux nouveaux modèles obtenus sont alors nommés  $GB_{Av}-B+OxyLH_2$ -closed et  $GB_{Av}-B$ -closed\_w\_OxyLH<sub>2</sub>, le premier représentant la conformation repliée avec l'oxyluciférine et l'AMPH présents dans la cavité avant le repliement. Le deuxième modèle correspond à la conformation repliée puis le docking des substrats OxyLH<sub>2</sub> et AMPH (voir figure 10.8). Le RMSD entre les 2 nouvelles structures et la 4G36 est de 6 Å si l'on ne considère que le C-terminal et de 7 Å si on considère toute

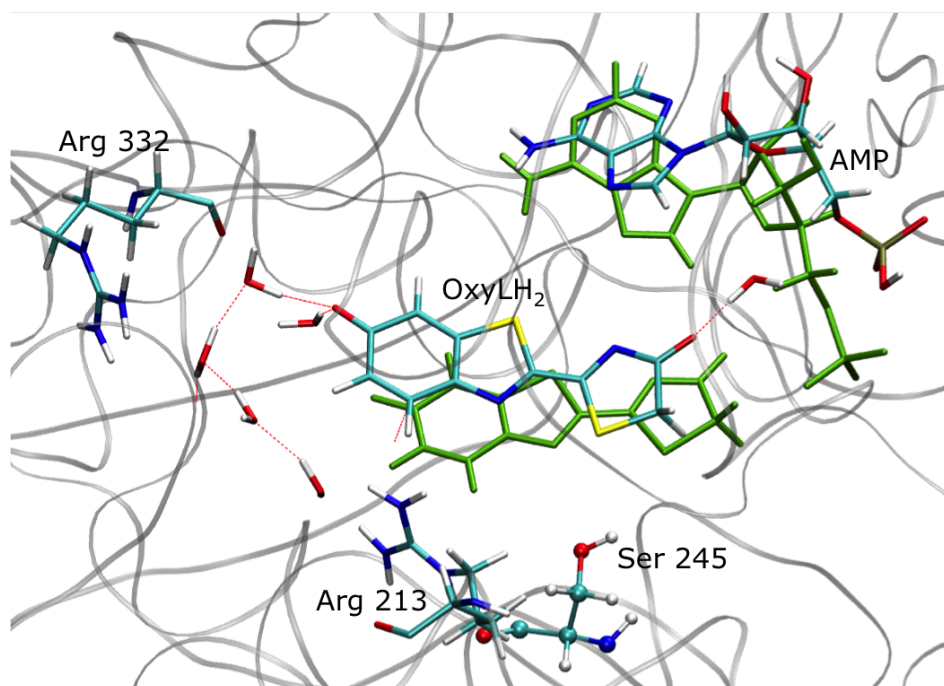


FIGURE 10.9 – Structure du site actif de la  $GB_{Av}$ -B-closed\_w\_OxyLH<sub>2</sub> de l'*image* (2) du tableau 10.3, superposée avec la structure du système de la luciole 2D1R. Les structures sont comparées en superposant leur domaines N-terminaux. L'OxyLH<sub>2</sub> dans l'*image* (2) du tableau 10.3 est représentée en tout atome. De la structure modèle 2D1R, seule OxyLH<sub>2</sub> est représentée en vert pour plus de clarté. Le résidu Ser 245, représenté en CPK correspond à une substitution entre 2D1R (Phe) et  $GB_{Av}$  (Ser).

la protéine sans les substrats.

## 10.2.5 Calcul de l'émission dans les structures repliées

L'obtention de  $GB_{Av}$ -B+OxyLH<sub>2</sub>-closed et  $GB_{Av}$ -B-closed\_w\_OxyLH<sub>2</sub> permet de discuter l'effet du repliement sur la longueur d'onde d'émission et voir si on peut mieux reproduire le comportement observé chez  $GB_{Av}$ . Des MD de 10 ns ont été réalisées sur ces deux modèles.

### 10.2.5.1 Le modèle $GB_{Av}$ -B-closed\_w\_OxyLH<sub>2</sub>

Commençons à détailler le modèle  $GB_{Av}$ -B-closed\_w\_OxyLH<sub>2</sub>. On rappelle que dans ce modèle, l'oxyluciférine et l'AMPH ont été ajoutées après le repliement de la protéine. Après une MD avec les deux substrats dockés dans la protéine, la position adoptée par l'oxyluciférine et l'AMPH est proche de celle dans les luciférase naturelles des lucioles, 4G36 et 2D1R. On observe peu ou pas de mouvement de ces deux substrats à l'intérieur de la cavité lors d'une MD (voir figure 10.9) ce qui se ressent dans le calcul des  $T_e$ . Comme indiqué dans le tableau 10.3, les valeurs de  $T_e$  obtenues sont proches des valeurs calculées dans le modèle 4G37-OxyLH<sub>2</sub>, et proches des

Tableau 10.3 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B-closed\_w\_OxyLH<sub>2</sub>. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé dans 4G37 est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que les valeurs expérimentales issues des références<sup>172</sup>.

Modèle	B3LYP eV (nm)	SP in vacuo $T_e^{vac}$ eV (nm)	Exp. eV (nm)
GB <sub>Av</sub> -B-closed_w_OxyLH <sub>2</sub> (1)	2.21 (560)	2.22 (558)	2.3 (538)
GB <sub>Av</sub> -B-closed_w_OxyLH <sub>2</sub> (2)	2.20 (564)	2.13 (581)	
4G37-OxyLH <sub>2</sub>	2.23 (555)	2.24 (553)	2.22 (558)

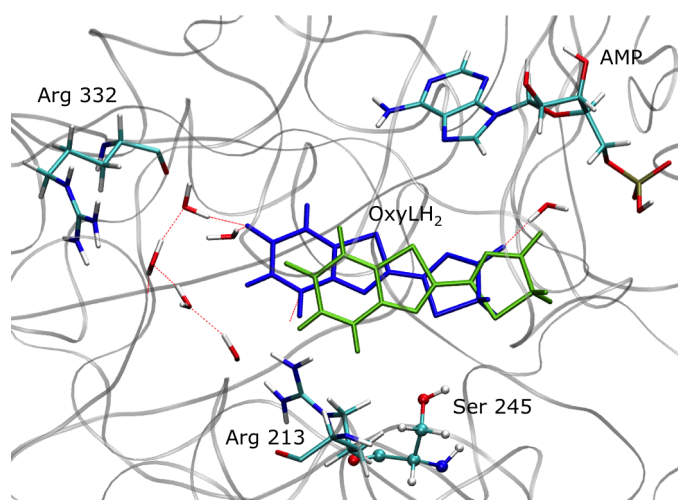


FIGURE 10.10 – Structure du site actif de la GB<sub>Av</sub>-B-closed\_w\_OxyLH<sub>2</sub> dans les différentes *images* (1 à 2). Les domaines N-terminaux ont été superposés. Le substrat OxyLH<sub>2</sub> est représenté dans une couleur différente dépendant de l'*image* référencée dans le tableau 10.3. En bleu, le substrat de l'*image* (1), en vert le (2).

valeurs expérimentales pour les luciférases 4G37 et 2D1R (voir figure 10.10). Le nombre de liaisons hydrogène formées entre l'oxyluciférine et des molécules d'eau ou résidus est compris entre 2 et 3. Ce nombre est similaire à celui trouvé dans le modèle 4G37-OxyLH<sub>2</sub>. Il existe cependant une légère différence entre ce modèle et le modèle 4G37-OxyLH<sub>2</sub> et concerne la position de l'AMPH. La molécule se déplace légèrement en direction de l'interface entre la protéine et le solvant. La distance entre l'oxyluciférine et l'AMPH est alors plus grande que dans les autres modèles. Cela n'a cependant pas d'effet sur la valeur du  $T_e$ , l'AMPH restant dans la deuxième sphère de coordination de l'oxyluciférine.

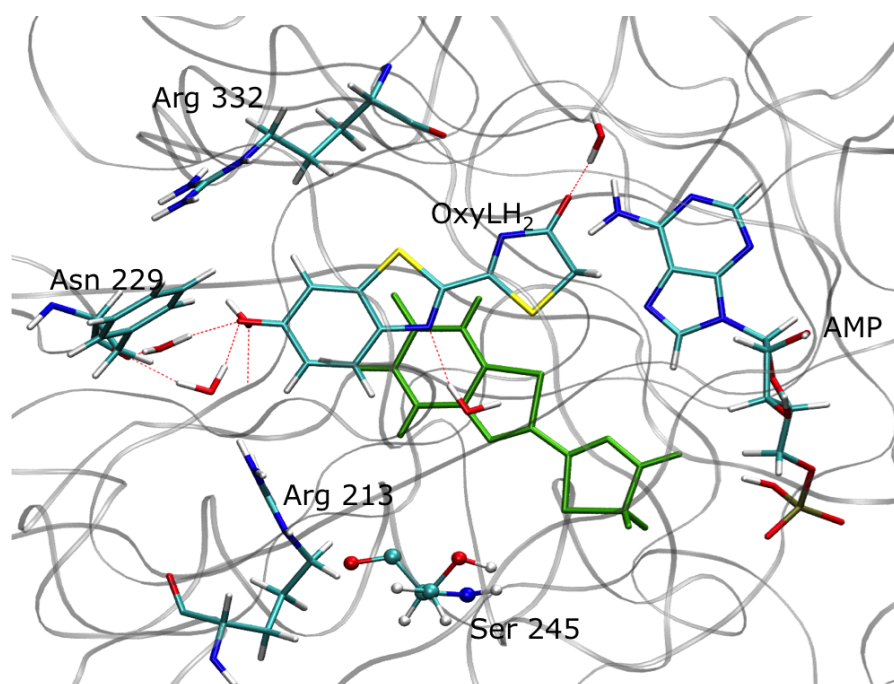


FIGURE 10.11 – Structure du site actif de la GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed de l'*image (2)* du tableau 10.4, superposée avec la structure du système de la luciole 2D1R. Les structures sont comparées en superposant leur domaines N-terminaux. L'OxyLH<sub>2</sub> dans l'*image (2)* du tableau 10.4 est représentée en tout atomes. De la structure modèle 2D1R, seule OxyLH<sub>2</sub> est représentée en vert pour plus de clarté. Le résidu Ser 245, représenté en CPK correspond à une substitution entre 2D1R (Phe) et GB<sub>Av</sub> (Ser).

### 10.2.5.2 Le modèle GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed

Parlons maintenant du modèle GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed. Dans ce modèle l'oxyluciférine et l'AMPH sont déjà à l'intérieur du site actif avant le repliement du domaine C-term de la protéine. Une MD a également été réalisée suite à ce repliement afin de pouvoir équilibrer le système. Lors de cette MD, le substrat reste dans la position initialement obtenue, avant le repliement (voir figure 10.11). Pour rappel l'oxyluciférine adopte un positionnement différent à l'intérieur de GB<sub>Av</sub>-B par rapport à 4G37-OxyLH<sub>2</sub>. Ce résultat montre que la position adoptée par l'oxyluciférine dans cette protéine reste stationnaire malgré un changement important de la conformation de la protéine. Les calculs du T<sub>e</sub> sur plusieurs *images* montrent une faible dispersion des résultats, la différence maximale en terme d'émission étant de 0.1 eV sur 5 *images* extraites de la MD (voir tableau 10.4). Le repliement de la protéine permet donc de stabiliser les valeurs de l'émission pour GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed. Quand aux valeurs du T<sub>e</sub> calculées, elles varient entre 2.42 eV (512 nm) et 2.32 eV (535 nm). Les valeurs calculées possèdent un écart avec la valeur expérimentale de 538 nm compris entre +0.02 eV et +0.12 eV, ce qui est inclus dans la marge d'incertitude de la méthode TD-DFT/MM. En terme de géométries et de valeurs de l'angle de torsion, elles sont similaires à celle obtenue dans la paragraphe 10.2.3, la torsion étant comprise entre 1 et 5° (voir figure 10.12).

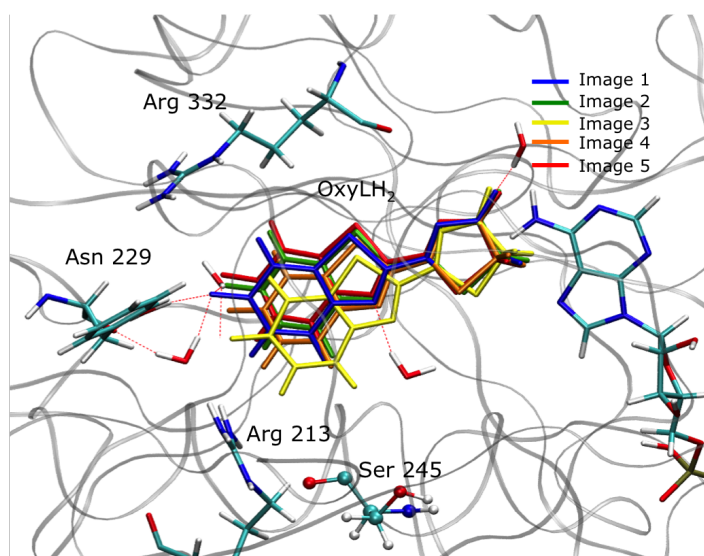


FIGURE 10.12 – Structure du site actif de la GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed dans les différents *images* (1 à 5). Les domaines N-terminaux ont été superposés. Le substrat OxyLH<sub>2</sub> est représenté dans une couleur différente dépendant de l'*image* référencée dans le tableau 10.4. En bleu, celui de l'*image* (1), en vert le (2), en jaune le (3), en orange le (4) et en rouge le (5).

Tableau 10.4 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. TD-DFT/MM avec une base 6-311G(2d,p) sur une structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence<sup>17</sup>.

Modèle	B3LYP eV (nm)	SP in vacuo $T_e^{vac}$ eV (nm)	Exp. eV (nm)
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed (1)	2.32 (535)	2.25 (551)	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed (2)	2.33 (532)	2.24 (553)	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed (3)	2.36 (525)	2.23 (555)	2.3 (538)
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed (4)	2.35 (527)	2.29 (541)	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed (5)	2.42 (512)	2.25 (551)	



### 10.2.5.3 Conclusion

Les valeurs des  $T_e$  calculées avec les modèles possédant le domaine C-terminal replié montrent deux comportements différents. Dans le cas du modèle GB<sub>Av</sub>-B-closed\_w\_OxyLH<sub>2</sub>, le repliement du C-term a été réalisé sans les substrats dans la cavité. Ces derniers ont par la suite été dockés dans la cavité. Après MD, l'oxyluciférine adopte la même position dans le site actif que celle observée dans le modèle 4G37-OxyLH<sub>2</sub>, et les valeurs du  $T_e$  calculées sont très proches de la valeur expérimentale de l'émission dans les lucioles. Ainsi, le repliement de la protéine sans substrat à l'intérieur conduit la protéine GB<sub>Av</sub> à émettre une lumière de la même longueur d'onde que pour les lucioles. Ce résultat ne correspondant pas aux mesures expérimentales observées où l'émission obtenue est déplacée vers le bleu.

Dans le cas du modèle GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed, le repliement du C-term avec les substrats déjà présents dans la cavité n'affecte pas la position de ces derniers dans le site actif. L'oxyluciférine garde la même position décrite que dans la section 10.2.3, et celle-ci est différente de celle du modèle 4G37-OxyLH<sub>2</sub>, issu des lucioles. Ainsi, les valeurs du  $T_e$  calculées sont proches de la valeur expérimentale pour la bioluminescence de *Amydetes viviani*. On montre ainsi que c'est bien la position adoptée par l'oxyluciférine dans la cavité qui permet d'expliquer ce déplacement vers le bleu de l'émission. Dans le cas de *Amydetes viviani*, l'oxyluciférine est exposée à d'autres résidus en comparaison avec sa position dans le site actif chez les lucioles. Les effets électrostatiques exercés sur OxyLH<sub>2</sub> sont alors différents entraînant une modification de l'émission. En revanche, on ne sait pas encore pourquoi l'oxyluciférine change de position dans le cas du lampyre *Amydetes viviani*.

Le point important et nouveau que cette étude apporte est que la position de l'oxyluciférine à l'intérieur du site actif n'est pas fixe, selon l'espèce étudiée, et cette position a une importance notable vis à vis de la longueur d'onde d'émission de la bioluminescence.

### 10.2.6 Etude de mutations chez GB<sub>Av</sub>

L'étude de mutations ponctuelles dans la protéine est l'un des outils les plus efficaces afin de comprendre le fonctionnement d'un système. En outre, l'étude de ces mutations de façon théorique (in silico) ou expérimental est relativement simple et permet de confirmer et d'améliorer la modélisation du système. Dans ce contexte, nous allons ici nous intéresser aux différences de séquence entre les deux structures cristallographiques obtenues par les expérimentateurs d'Abu Dhabi, à savoir GB<sub>Av</sub> qui provient de la luciole *Amydetes viviani* avec une émission dans le bleu-vert et RE<sub>Ph</sub> qui provient de la larve du coléoptère *Phrixothrix hirtus* et émet dans le rouge. Plus précisément on va modéliser les différences en terme de résidus entre ces deux protéines. On considère trois mutations ponctuelles correspondant au passage entre GB<sub>Av</sub> et RE<sub>Ph</sub> : deux mutations proches du site actif et une insertion d'un résidu dans une boucle. En détails, les mutations sont l'arginine 332 en leucine, la leucine 346 en isoleucine et l'insertion d'une arginine en position 351 (voir figure 10.13). Toutes les

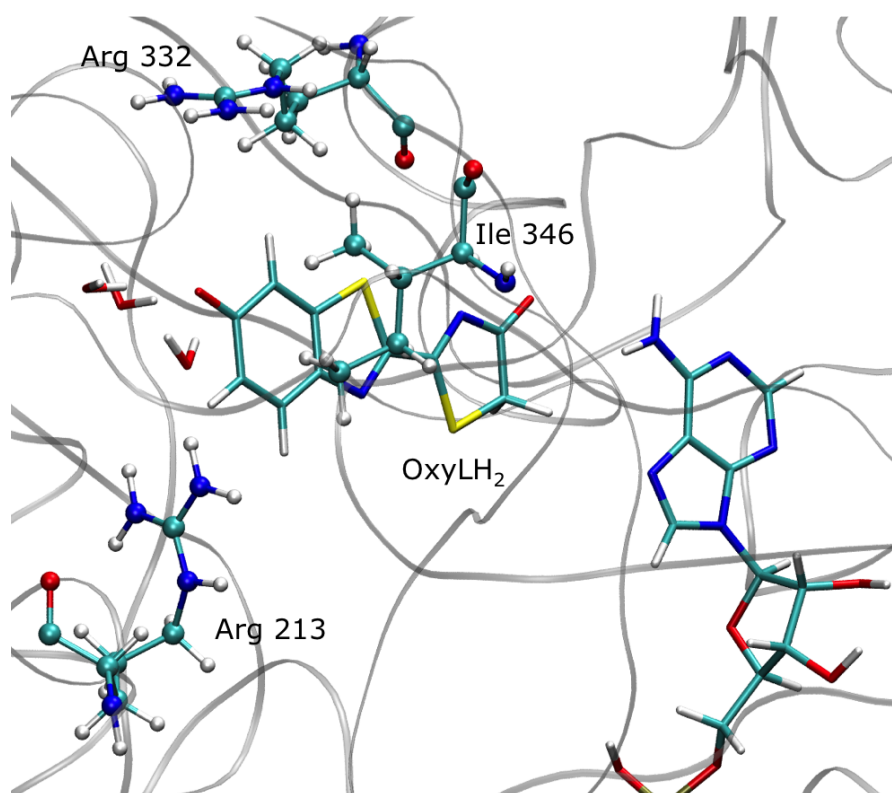


FIGURE 10.13 – Structure du site actif dans la  $GB_{Av}$ -B+OxyLH<sub>2</sub>-closed avec les principaux résidus importants dans l'étude des mutations.

numérotations correspondent à la séquence de  $GB_{Av}$ . Le but de ces mutations est de reproduire le déplacement vers le rouge de l'émission observée chez  $RE_{Ph}$ . Afin de réaliser ces mutations nous allons nous servir du modèle  $GB_{Av}$ -B+OxyLH<sub>2</sub>-closed précédent, celui-ci reproduisant au mieux les valeurs expérimentales d'émission de *Amydetes viviani*.

Pour chacune des deux mutations R332L et I346L dans  $GB_{Av}$ , une MD de 10 ns a été réalisée afin d'équilibrer le système. En outre, une MD de 30 ns a été faite en vue de bien reproduire le mouvement de la boucle lors de l'insertion de l'arginine 351 dans  $GB_{Av}$ . Chaque mutation est étudiée séparément en détail dans les paragraphes suivants.

#### 10.2.6.1 R332L

La mutation R332L, soit l'arginine 332 en leucine change un résidu polaire chargé en résidu apolaire et modifie l'interaction électrostatique exercée par ce résidu. Ce résidu est situé du côté du cycle benzothiazole (voir figure 10.13) et participe souvent au réseau de liaisons hydrogène avec l'oxyluciférine par l'intermédiaire d'une molécule d'eau. Durant la MD réalisée, on voit un léger mouvement de l'oxyluciférine en direction du C-terminal. Lors de cette MD, on observe également le mouvement de l'arginine 213.

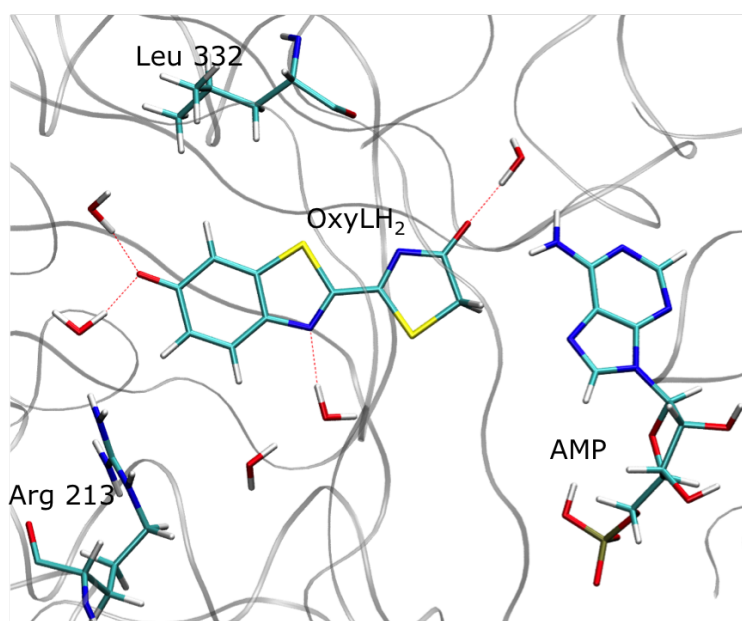


FIGURE 10.14 – Structure du site actif dans la GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-R332L. L'image représentée est la (2) et est issue du tableau 10.5.

Tableau 10.5 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-R332L. Les numéros entre parenthèses correspondent au numéro de l'image extraite de la MD. La valeur obtenue selon le même procédé pour le modèle GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur une structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence<sup>17</sup>.

Modèle	B3LYP eV (nm)	SP in vacuo $T_e^{vac}$ eV (nm)	Exp. eV (nm)
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L (1)	2.18 (568)	2.14 (577)	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L (2)	2.15 (577)	2.22 (559)	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L (3)	2.22 (557)	Erreur	2.3 (538)
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L (4)	2.21 (561)	Erreur	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed (1)	2.32 (535)	2.25 (551)	

Tableau 10.6 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed-I346L. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé pour le modèle GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence<sup>17</sup>.

Modèle	B3LYP eV (nm)	SP in vacuo $T_e^{vac}$ eV (nm)	Exp. eV (nm)
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-I346L (1)	2.35 (528)	2.25 (550)	
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-I346L (2)	2.40 (516)	Erreur	
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-I346L (3)	2.40 (517)	2.22 (557)	2.3 (538)
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-I346L (4)	2.33 (531)	Erreur	
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed (1)	2.32 (535)	2.25 (551)	

Avant la mutation R332L, l'arginine 213 interagit aussi avec l'oxyluciférine mais est souvent située dans la deuxième sphère de solvatation de l'oxyluciférine. Suite à la mutation R332L, l'arginine 213 passe dans la 1<sup>ère</sup> sphère de solvatation (voir figure 10.14). Comme attendu, la mutation de l'arginine 332 affecte de façon importante la valeur de la longueur d'onde d'émission (voir tableau 10.5). Ainsi on observe un déplacement vers le rouge de l'émission avec une longueur d'onde d'émission comprise entre 557 nm et 577 nm. Le nombre de liaisons hydrogène formées entre l'oxyluciférine et l'environnement protéique est similaire à celui dans GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed, soit entre 4 et 6. On remarque cependant que certains calculs sont plus difficiles à converger notamment lorsqu'on cherche à obtenir la valeur du  $T_e^{vac}$ , montrant que la géométrie substrat est instable sans la présence de la protéine. Une MD de 10 ns n'est sans doute pas assez longue pour équilibrer totalement le système.

### 10.2.6.2 I346L

La mutation I346 change une isoleucine en leucine. Plus précisément, cela correspond à bouger un groupement méthyl CH<sub>3</sub> de place, du carbone C<sub>γ</sub> au C<sub>β</sub>. Ce résidu est également dans le site actif avec le groupement CH<sub>3</sub> pointé vers le cycle benzothiazole (voir figure 10.13). Cette mutation impacte exclusivement l'environnement géométrique autour du résidu. En effet, les valeurs de l'émission calculées sur plusieurs *images* montrent peu ou pas de différences avec les valeurs obtenues avec GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed (voir tableau 10.6). Notons cependant que pour certains calculs la valeur du  $T_e^{vac}$  n'est pas accessible. En outre, le nombre de liaisons hydrogène formées avec l'oxyluciférine est plus faible que pour GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed, environ 4-6 pour ce dernier contre 2-3 pour la modèle GB<sub>AV</sub>-B-closed-I346L.

### 10.2.6.3 Insertion de R351

On étudie ici l'insertion d'un résidu dans la chaîne polypeptidique. Dans notre cas on ajoute une arginine à l'emplacement 351 qui correspond au milieu d'une

Tableau 10.7 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed-insert-R351. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé pour le modèle GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence<sup>17</sup>.

Modèle	B3LYP eV (nm)	SP in vacuo $T_e^{vac}$ eV (nm)	Exp. eV (nm)
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-insert-R351 (1)	2.25 (550)	2.18 (567)	
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-insert-R351 (2)	2.25 (550)	2.21 (561)	2.3 (538)
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-insert-R351 (3)	2.23 (553)	2.21 (560)	
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed (1)	2.32 (535)	2.25 (551)	

structure secondaire de type boucle. L'ajout de ce résidu chargé, polaire et surtout très encombrant stériquement risque de déformer la conformation de la protéine. Une MD de 30 ns a alors été réalisée. L'intérêt de l'ajout de ce résidu est de voir s'il est possible de reproduire le comportement adopté dans RE<sub>Ph</sub>. La présence de ce résidu dans la boucle de RE<sub>Ph</sub> force la protéine à adopter une conformation très singulière. La boucle possédant l'arginine semble se détacher des autres boucles de la protéine et "pendre", seulement stabilisée par le solvant environnant.

Lors de cette dynamique, le mouvement de cette boucle n'est pas observée, la conformation de cette dernière restant collée à la protéine. Les calculs de la valeur du  $T_e$  montre un déplacement vers le rouge de l'émission (voir tableau 10.7) bien que celles-ci soit moins importantes que dans le cas de R332L. Il est néanmoins nécessaire de tempérer ces valeurs. En effet, la conformation de la structure secondaire possédant cette boucle est différente de celle observées chez RE<sub>Ph</sub>. Les valeurs de  $T_e$  calculées ne nous permettent pas de conclure sur le rôle du résidu R351 chez RE<sub>Ph</sub>.

L'étude de ces trois mutations donnent des résultats cohérents. La mutation d'un résidu polaire en résidu apolaire influence l'environnement électrostatique de la protéine alors que la mutation d'un résidu apolaire en résidu apolaire de même taille n'a pas d'influence sur le substrat. On peut alors expliquer en partie le déplacement vers le rouge de RE<sub>Ph</sub> par la mutation R322L.

Les MD concernant l'insertion d'un résidu dans la protéine sont cependant assez décevantes. Malgré un temps de dynamique assez long, aucune différence n'est observée dans la conformation des boucles possédant les résidus mutés. On voit ici une limite des calculs théoriques dans les protéines, la difficulté de prédire des comportements influençant directement la conformation de la protéine.

Concernant l'insertion du résidu R351, une étude plus approfondie de la protéine RE<sub>Ph</sub> montre qu'une seconde mutation, proche du résidu inséré, la E354N permet la création d'une liaison hydrogène entre le résidu 351 et le résidu 354. Cette liaison faible n'est pas du tout observée lors de la MD. Une perspective est donc de tester l'influence de la double mutation sur la conformation de la boucle et sur la valeur de

la longueur d'onde d'émission de la bioluminescence.

### 10.3 Prédire le comportement de l'environnement

Au vu des chapitres 9 et 10, la modulation de la longueur d'onde d'émission de la bioluminescence est largement influencée par l'environnement. Une question à se poser serait : peut-on, en observant les résidus du site actif d'une nouvelle protéine prédire la longueur d'onde d'émission ? Avec la question corollaire : peut-on proposer des mutations qui permettent de changer la longueur d'onde émise ?

Grâce à l'expérience acquise et l'étude de mutations *in silico*, un schéma global semble se dessiner. Les liaisons hydrogène formées par l'oxyluciférine avec son environnement proche peuvent provenir des différents hétéroatomes, les deux oxygènes et les deux azotes. En terme de probabilité, le nombre de liaisons hydrogène formées par l'oxygène du cycle phénolate est compris entre 2 et 4, l'oxygène du thiazolone entre 0 et 2 et pour les deux azotes entre 0 et 1.

D'autre part, l'étude des orbitales moléculaires HOMO et LUMO, notamment la localisation des orbitales sur les atomes permet de connaître l'effet des liaisons faibles sur la stabilisation des deux orbitales (voir Figure 10.15). Ainsi, pour l'oxyluciférine, une liaison hydrogène crée avec l'oxygène du cycle phénolate stabilise la HOMO et conduit alors à un déplacement vers le bleu de l'émission. En revanche, la présence d'une liaison hydrogène avec les autres hétéroatomes, les deux azotes et l'oxygène du thiazolone, stabilise la LUMO donnant lieu à un déplacement vers le rouge de l'émission.

Afin de confirmer ces résultats, des calculs théoriques ont été réalisés dans différents

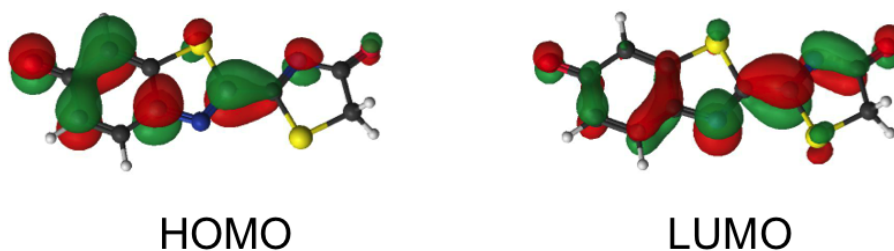


FIGURE 10.15 – Orbitales frontières HOMO et LUMO dans OxyLH<sub>2</sub>.

modèles, i.e. 4G37-OxyLH<sub>2</sub>, GB<sub>AV</sub>-B-OxyLH<sub>2</sub> (2) et GB<sub>AV</sub>-B+OxyLH<sub>2</sub>-closed (1). Pour calculer l'influence d'une liaison hydrogène, d'un résidu ou d'une molécule d'eau, un calcul *single point* est réalisé en fixant la charge des atomes du fragment visé à 0. De cette façon, on peut, par comparaison avec le calcul où les charges sont celles du champ de force, en déduire l'effet électrostatique apporté par ce fragment mais pas son interaction géométrique avec le substrat.

Les calculs préliminaires réalisés dans cette thèse confirme les points précédemment explicités, les liaisons hydrogène formées avec l'oxygène du cycle benzothiazole donnent lieu à un déplacement vers le bleu de l'émission et les liaisons hydrogène avec les azotes et l'autre oxygène de l'oxyluciférine à un déplacement vers le rouge (voir figure 10.16). En détails, les liaisons hydrogène avec des molécules d'eau ou des résidus polaires modifient la valeur du  $T_e$  de l'ordre de 0.02 eV à 0.07 eV. En revanche, la présence d'un résidu chargé comme l'arginine influence plus la valeur du  $T_e$  même si ce résidu ne participe pas au réseau de liaisons hydrogène. L'interaction électrostatique apporté par ce résidu est comprise entre 0.07 eV et 0.17 eV.

Un exemple étudié est la mutation *in silico* du l'arginine 332 en leucine (10.2.6).

Tableau 10.8 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-R332L optimisée dans S1 et comparaison avec les *single points* obtenus en mettant les charges des atomes de l'arginine 332 égales à zéro, appelé "off" dans le tableau. Les numéros entre parenthèses correspondent aux numéros de l'*image* extraite de la MD. Les valeurs sont obtenues par TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée géométriquement avec la même fonctionnelle.

Modèle	B3LYP eV (nm)	SP with R332 "off" eV (nm)
GB <sub>Av</sub> -B-OxyLH <sub>2</sub> (2)	2.30 (538)	2.24 (553)
4G37-OxyLH <sub>2</sub>	2.23 (555)	2.11 (585)
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed (2)	2.33 (532)	2.11 (585)
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L (1)	2.18 (568)	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L (2)	2.15 (577)	
GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L (3)	2.22 (557)	

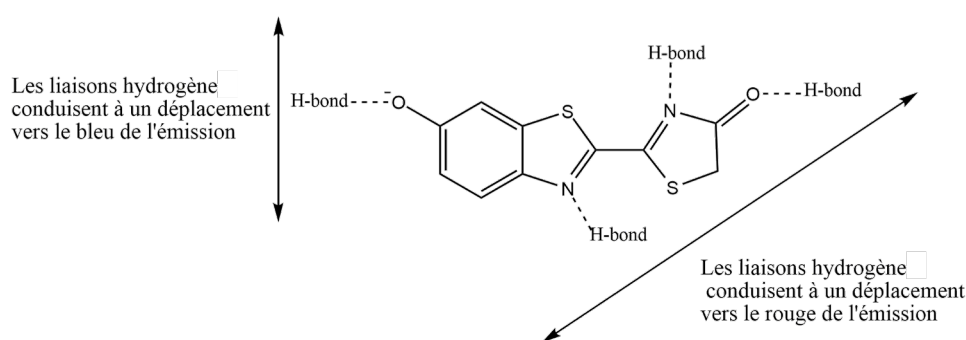


FIGURE 10.16 – Schéma de l'influence des liaisons hydrogène sur la longueur d'onde d'émission de l'oxyluciférine

Les résultats trouvés montrent un déplacement vers le rouge de l'ordre de 0.1 eV. Des calculs *single points* avec les charges du résidu R332 fixées à zéro montrent également un déplacement vers le rouge. L'écart observé est entre 0.06 eV et 0.15 eV selon le modèle (voir tableau 10.8). Ces valeurs sont légèrement surestimées car on

ne prend pas en compte le réarrangement de l'environnement suite à la mutations. En revanche le gain de temps est important, on compare un calcul *single point* avec une MD suivi de calculs TD-DFT.

En revanche certains comportements sont encore difficilement explicables. Par exemple, la liaison hydrogène entre l'AMPH et l'oxygène du thiazolone conduit à un déplacement vers le bleu au lieu du déplacement vers le rouge attendu. De plus tous les résidus et molécules d'eau situés au delà de deux liaisons hydrogène de l'oxyluciférine n'ont aucun effet.

## 10.4 Conclusion

Grâce à des MD suivies de calculs QM/MM on a pu étudier et comprendre l'impact des nouvelles luciférase sur la longueur d'onde de la bioluminescence.

Pour GB<sub>Av</sub>, la conformation obtenue par RX avec le domaine C-terminal ouvert n'est pas la conformation permettant la bioluminescence. Ainsi, comme chez les lucioles, le repliement de ce domaine est un prérequis nécessaire à l'obtention de l'émission. Cependant, la position de l'oxyluciférine dans le site actif de la protéine n'est pas la même dans le cas GB<sub>Av</sub> que dans le cas des lucioles. Seule la position adoptée dans le cas GB<sub>Av</sub> permet de modéliser convenablement la valeur du T<sub>e</sub> de l'émission. Si on force l'oxyluciférine à adopter la position observée dans les luciférase des lucioles, l'émission obtenue est alors proche de celle trouvée chez les lucioles.

Dans un second temps, l'étude de plusieurs mutations *in silico* permettent de mieux comprendre l'influence de l'environnement sur le substrat et l'émission. Ainsi, on peut commencer à discriminer chaque effet interagissant sur l'émission. Le but final serait de pouvoir prédire l'influence exercée par chaque résidu et de pouvoir pointer les mutations pertinentes.

Tous ces résultats restent cependant à être confirmés par des mesures expérimentales. La collaboration expérimentateurs-théoriciens reste toujours le moteur de l'avancée scientifique dans le domaine de la bioluminescence.

## 10.5 Détails de calculs

### 10.5.1 Partie MM

La structure cristallographique GB<sub>Av</sub> issue du système *Amydetes viviani* a été modélisée afin de pouvoir réaliser des MD ainsi que des calculs QM/MM. Dans cette optique, les deux substrats, l'OxyLH<sub>2</sub> et l'AMPH doivent être docké à l'intérieur de la protéine. Ce docking a été fait manuellement en se servant des positions de ces substrats dans le modèle 4G37-LH<sub>2</sub>.

Les résidus ont été protonnés par le programme Leap provenant de la suite logiciel Amber 14<sup>55</sup>. Les cas discutables, notamment pour les résidus histidines ont été élucidés en calculant leur pKa avec le program H++<sup>95</sup>. Selon le modèle étudié, nous



avons choisi de protonés deux fois certaines histidines (HIP). Les numéros des résidus d’histidines doublement protonés se trouvent dans le tableau 10.5.1.

Modèle	Numéro des résidus d’HIP
GB <sub>AV</sub> -B	5, 42, 72, 404 and 426
GB <sub>AV</sub> -A	6, 43, 73, 405 and 427
GB <sub>AV</sub> -B-OxyLH <sub>2</sub>	5, 23, 42, 72, 305, 404 and 426
GB <sub>AV</sub> -B-OxyLH <sub>2</sub> +H <sub>2</sub> O	5, 23, 42, 72, 305, 404 and 426
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed	5, 23, 42, 72, 305, 404 and 426
GB <sub>AV</sub> -B-closed_w_OxyLH <sub>2</sub>	5, 23, 42, 72, 305, 404 and 426
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-R332L	5, 23, 42, 72, 305, 404, 426 and 456
GGB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-I346L	5, 23, 42, 72, 305, 404 and 426
GB <sub>AV</sub> -B+OxyLH <sub>2</sub> -closed-insert-R351	5, 23, 42, 72, 405 and 427
4G37-OxyLH <sub>2</sub>	76, 171, 310, 332, 419, 461 and 489
4G37-F250S-OxyLH <sub>2</sub>	76, 171, 310, 332, 419, 461 and 489
2D1R-OxyLH <sub>2</sub>	457

Tableau 10.9 – Tableau des différents modèles utilisés dans ce chapitre ainsi que les numéros des résidus des histidines doublement protonés dans ces derniers.

## 10.5.2 Partie MD

Les simulations de dynamiques moléculaires classiques ont été réalisées avec la suite logicielle Amber14<sup>55</sup> afin d’obtenir une représentation du système pour une future optimisation QM/MM. Les modèles ont été solvatés par des molécules d’eau de type TIP3P dans une boîte de forme cuboïde. Le système résultant contient environ 90000 atomes au total, dont 28000 molécule d’eau formant une sphère de solvation d’au moins 15 Å autour du substrat. Un champs de force de type AMBER99ff a été utilisé pour modéliser les résidus de la protéine. Des paramètres développés par notre groupe ont été choisis afin de modéliser le substrat OxyLH<sub>2</sub> ainsi que l’AMPH<sup>111</sup>. Ceux-ci n’étant pas optimisés pour des états excités, la structure et les charges des substrats au niveau excité S1 ont été obtenus par calcul QM/MM préliminaire puis réinjectés dans les fichiers paramètres. Le substrat reste gelés dans cette géométrie pendant toute la dynamique. Le déroulement d’une dynamique est le suivant : 3 minimisations sont d’abord réalisées, la 1<sup>ère</sup> minimise seulement la protéine en gelant les molécules d’eaux. La deuxième minimisation ne gèlent que les atomes d’hydrogène alors que rien n’est gelé (à part le substrat) durant la 3<sup>ème</sup>. Le système est ensuite chauffé de 100K à 300K en 20 ps. Puis sous condition NPT avec T=300K, P=1 atm une dynamique de 10 ns ou plus si indiqué est conduite sous condition périodique aux bords avec comme pas de temps 2 fs. La température et la pression sont maintenues en utilisant l’algorithme de Berendsen<sup>61</sup> avec une constante de couplage de 5 ps. Des contraintes de type SHAKE sont appliquées sur toutes les liaisons impliquant des liaisons hydrogène.

Le repliement du domaine C-terminal a été réalisé par une simulation de type umbrella sampling. L’algorithme d’umbrella sampling a été conduit entre les deux

carbones alpha des résidus 311 et 524, le premier situé dans le domaine N-terminal et le deuxième dans le C-terminal. La distance de départ est de 21 Å et celle de fin de 7 Å avec un pas de 1 Å. A chaque pas, 200 ps d'équilibration sont réalisées suivies de 750 ps de production. Puis, la dernière structure de la MD de production est récupérée pour chaque pas, ce qui permet de visualiser le repliement entre les carbones. Le biais introduit par la MD d'umbrella sampling est enlevé par la méthode d'analyse d'histogrammes pondérés WHAM<sup>99,100</sup>, afin de générer un profil d'énergie libre le long du chemin suivi.

### 10.5.3 Partie QM/MM

Les calculs QM/MM sont réalisés en utilisant un schéma de couplage ESPF entre Gaussian<sup>67</sup> et Tinker<sup>56</sup> (Gaussian 09d/Tinker). La méthode ESPF (electrostatic potential fitted)<sup>68</sup> permet de calculer les interactions entre les charges de type Mulliken<sup>114</sup> du QM et le potentiel électrostatique externe du MM situé à moins de 9 Å du QM. La technique dite de microitérations<sup>116</sup> est utilisée afin de faire converger la géométrie de l'espace MM choisi pour chaque pas de minimisation du QM. L'émetteur (ou substrat) est inclus dans la partie QM et le reste du système (AMPH + protéine + molécules d'eau) dans la partie MM. Les calculs TD-DFT en QM/MM correspondent à une optimisation du 1<sup>er</sup> état excité singulet (S1) suivi par le calcul de la transition énergétique vertical ( $T_e$ ) entre ce même état S1 et l'état fondamental S0. Cette transition correspond à une émission de fluorescence. Chez les lucioles l'énergie de la fluorescence est identique à l'énergie de la transition après la réaction de bioluminescence<sup>19</sup>, la valeur de  $T_e$  calculée peut alors être assimilée à l'énergie d'émission de la bioluminescence.

Le niveau de théorie utilisé pour la partie QM dans les calculs QM/MM est le suivant. Des calculs TD-DFT/MM ont été réalisés avec la fonctionnelle B3LYP<sup>44,45,47</sup>, cette dernière donnant les résultats les plus proches des valeurs expérimentales<sup>20</sup>. De même la base choisie, 6-311G(2d,p) est également issue d'un benchmark. Cette base ne possède pas de fonctions diffuses associées car celles-ci peuvent interagir avec la partie MM lors de l'utilisation d'un électrostatique embedding ESPF<sup>68</sup>.

## **Cinquième partie**

### **Etude du mécanisme réactionnel de la bio et chimiluminescence**



Dans cette partie de la thèse, on s'intéresse à l'étude du mécanisme réactionnel de la bio et chimiluminescence. Dans le 1<sup>er</sup> chapitre, une étude conjointe entre expérimentateurs et théoriciens a permis de confirmer la structure chimique de la molécule émettrice chez les 2-coumaranones. Ce travail est réalisé en collaboration avec le groupe de l'institut de chimie organique et macromoléculaire de Jena en Allemagne. Dans le second chapitre on explore une étape de la bioluminescence chez les lucioles encore peu étudiée. On modélise l'approche de la molécule de dioxygène vers le substrat D-luciferyl adénylate. On suit pour cela la molécule de dioxygène de son entrée dans la protéine jusqu'à la formation du cycle dioxétanone. Cette étude permet d'explorer plusieurs zones d'ombres de la réaction de bioluminescence chez les lucioles, comme l'impact de la protéine et des résidus sur la coordination du dioxygène et sa forme chimique.

Enfin, dans le troisième chapitre, on modélise la réaction chimique permettant de relier deux tautomères du substrat émetteur oxyluciférine. Ces deux tautomères, le phénolate-kéto et le phénolate-énol sont en concurrence pour être le véritable émetteur de la bioluminescence chez les lucioles. Par cette étude on cherche à comprendre si il est possible, par une réaction de tautomérisation, de passer d'une molécule à l'autre.



# Chapitre 11

## La chimiluminescence des 2-coumaranones

### 11.1 Contexte

Ce chapitre traite de l'étude théorique de la chimiluminescence des 2-coumaranones, réalisé en collaboration avec un groupe d'expérimentateurs d'Allemagne et a donné lieu à la rédaction d'un article scientifique<sup>85</sup>.

Les 2-coumaranones sont des molécules chimiluminescentes avec un rendement quantique de l'ordre de 5% et possèdent un mécanisme réactionnel proche de la bioluminescence chez les lucioles. Suite à l'action d'une base forte permettant la déprotonation d'une liaison C-H, une molécule de O<sub>2</sub> se coordine à la 2-coumaranone. La structure intermédiaire possède un cycle dioxétanone, qui en se brisant, et après perte du CO<sub>2</sub>, conduit à une émission de lumière bleu, c'est la chimiluminescence. Le but de ce travail est de trouver et étudier la molécule émettrice. En effet, la cassure du cycle dioxétanone peut donner lieu à deux structures excitées précurseur de la chimiluminescence. J'ai également effectué un stage de 1 mois sur la synthèse de nouvelles 2-coumaranones, qui est discuté dans la deuxième de partie de ce chapitre.

### 11.2 Résultats

La compréhension du mécanisme réactionnel est déjà relativement bien étudiée. Les études expérimentales ont permis de trouver la structure du réactif et du produit de réaction après émission de lumière. Contrairement à l'oxyluciférine chez les lucioles, la cassure du cycle dioxétanone induit une cyclisation de la 2-coumaranone en  $\alpha$ -azacoumarine. Cependant, le produit final de la réaction n'est pas l'émetteur de la chimiluminescence, ni de la fluorescence. En effet, les longueurs d'onde d'émission de la fluorescence et de la chimiluminescence obtenues sur le produit de réaction chimiluminescente sont identiques dans un le milieu réactionnel basique, i.e. 434 nm. Or, la fluorescence issue du produit final de réaction noté **6** (voir figure

11.1) émet pour sa part à 366 nm. On remonte alors le mécanisme réactionnel afin de trouver des structures viables.

La figure 11.1 rassemble les différentes molécules appartenant au mécanisme de

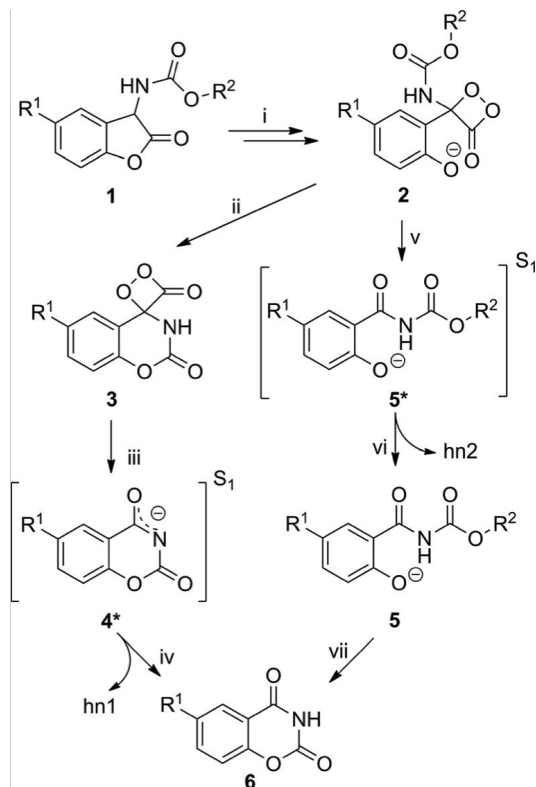


FIGURE 11.1 – Les deux chemins possibles de la décomposition de l'intermédiaire **2** dans la chimiluminescence des 2-coumaranones **1**. R<sub>1</sub> est un atome de fluor et R<sub>2</sub> un groupement éthyl.

la chimiluminescence chez les 2-coumaranones. Les deux structures possiblement émissives sont la **4\*** correspondant à une émission après cyclisation et la **5\***, une émission avant cyclisation.

La structure **4\*** correspond à la structure finale **6** déprotonée. La charge négative est complètement délocalisée sur l'ensemble du cycle donnant lieu à trois formes résonnantes, avec la charge localisée sur l'un ou l'autre des oxygènes ou sur l'azote. Les calculs TD-DFT et MS-CASPT2, conjointement avec des mesures de fluorescence expérimentale montrent cependant que cette structure **4\*** est non-émissive.

L'étude de la structure **5\*** est en revanche plus compliquée. Ce composé est très instable en solution et difficilement étudiable à température ambiante. Les résultats des calculs sur cette structure montrent un très bon accord avec les valeurs expérimentales de la chimiluminescence. Les différences observées, pour les deux méthodes de calculs, aussi bien en absorption qu'en émission sont sous la barre des 0.1 eV en comparaison avec la valeur expérimentale. Ces résultats encourageant ont été confirmés par la synthèse du composé **5** à faible température, reproduisant



Tableau 11.1 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues sur 2 possibles émetteurs de la chimiluminescence chez les 2-coumaranones. TD-DFT obtenu avec une base 6-31+G(d,p) sur un structure optimisée géométriquement avec la même fonctionnelle. MS-CASPT2 obtenu avec un espace actif de 14 électrons dans 12 orbitales en tenant compte de 4 états électroniques et avec une base ANO-RCC-VTZP. Les calculs sont effectués pour les structures **4\*** et **5\***.

	Exp. eV (nm)	B3LYP <b>5*</b> eV (nm)	MS-CAPST2 <b>5*</b> eV (nm)	B3LYP <b>4*</b> eV (nm)	MS-CAPST2 <b>4*</b> eV (nm)
Absorption	3.28 (378)	3.38 (367)	3.32 (374)	4.23 (293)	4.23 (293)
Fluorescence	2.86 (434)	2.89 (429)	2.93 (423)	3.37 (368)	3.45 (359)

presque parfaitement les valeurs d'émissions expérimentales de la chimiluminescence. Le composé **5\*** est alors la molécule émettrice de la chimiluminescence chez les 2-coumaranones.

En parallèle, les expérimentateurs ont remarqué que le produit final se décompose en partie en 4-fluorosalicylamide via une structure proche de celle de **5** confirmée à la fois par des résultats expérimentaux et théoriques (voir Figure 11.2). Cette structure notée **7**, possède également une valeur de l'émission de fluorescence proche des valeurs expérimentales de fluorescence. Cependant, la structure **8** montre des valeurs légèrement déplacées vers le bleu de l'ordre de 10 à 20 nm. Tous ces résultats précédents ont également été vérifiés par calcul (structure **7** par TD-DFT et MS-CASPT2, structure **8** en TD-DFT). On peut alors en déduire que la structure **7** est un possible émetteur de la fluorescence dans les 2-coumaranones. Cet étude de

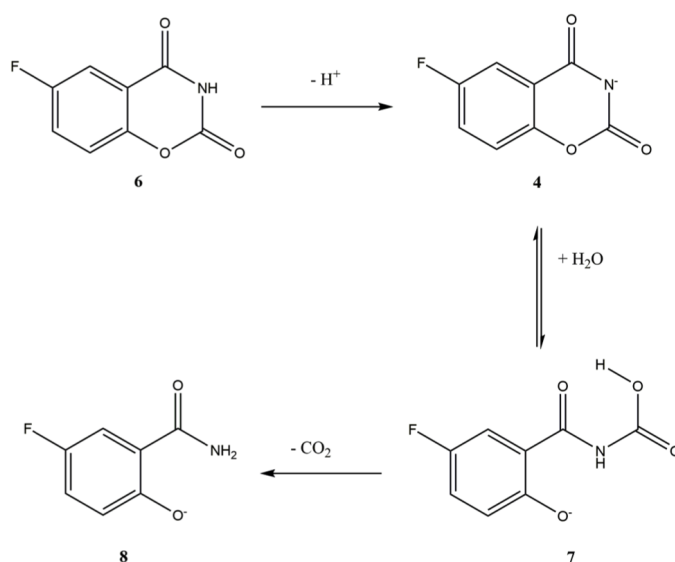


FIGURE 11.2 – Le chemin de décomposition de **6** en **8** avec la structure intermédiaire fluorescente **7**.

la chimiluminescence des 2-coumaranones montre l'intérêt de coupler les calculs théoriques avec des résultats expérimentaux. En outre, la chimiluminescence est un

sujet prometteur et possédant de nombreux parallèles avec la bioluminescence, aussi bien dans la structure des substrats émetteurs que dans le mécanisme de réaction.

### 11.3 La synthèse de nouvelles 2-coumaranones

La collaboration a en outre, donné lieu à la rédaction d'autres articles scientifiques<sup>86</sup> mais également à des échanges de doctorants entre les deux laboratoires. La venue du Dr. Schramm en juin 2015 en France a permis de conclure les travaux débutés sur la chimiluminescence des 2-coumaranones. Pour ma part, j'ai pu aller à Iéna durant le mois de mars 2016.

Lors de cet échange, le but était la synthèse de nouvelles 2-coumaranones incorporant le cycle benzothiazole de la luciférine. De cette façon on cherche à quantifier l'influence de ce groupe sur la chimiluminescence. Une deuxième hypothèse est de pouvoir insérer les 2-coumaranones à l'intérieur d'une protéine afin d'émettre de la lumière via une réaction catalysée par une enzyme comme pour la bioluminescence.

Trois molécules cibles sont choisies en vue de synthèse. La première (1) représente la molécule test, la deuxième (2) incorpore le cycle benzothiazole de la luciférine et la troisième (3) remplace le groupement phénol du cycle benzothiazole par un groupement méthoxy (voir Figure 11.3).

La synthèse de la molécule (1) a été effectuée et celle-ci, à défaut d'être chimiluminescente, présente une émission de type fluorescence dans le bleu. En revanche, malgré plusieurs voies de synthèses différentes il a été impossible de synthétiser les molécules (2) et (3).

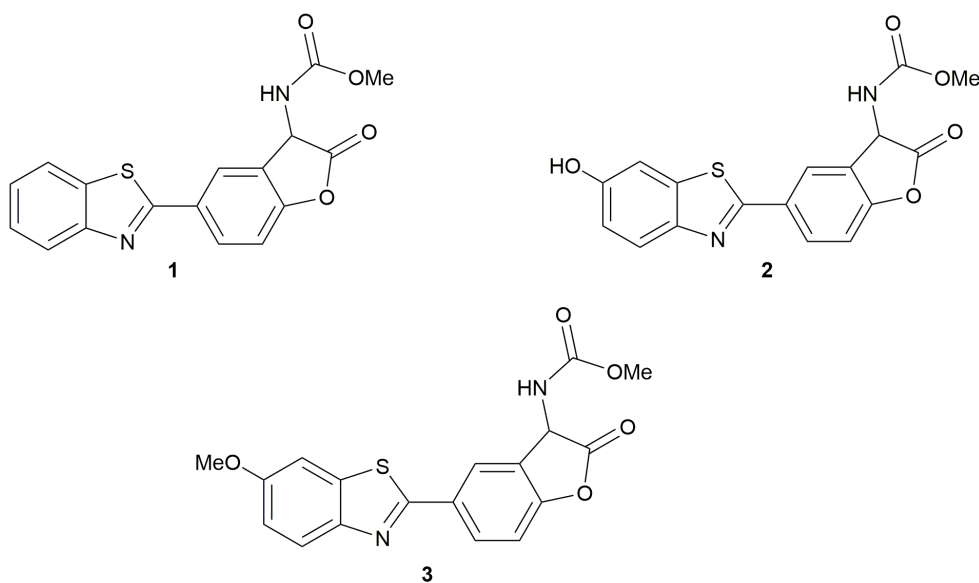


FIGURE 11.3 – Schéma des 3 trois structures cibles.

## 11.4 Détails de calculs

Les calculs QM sont réalisés sur des modèles issus de structures cristallographiques ou construits à la main. Sur ces modèles des calculs DFT et TD-DFT ou MS-CASPT2 sont effectués. Les calculs DFT utilise le logiciel Gaussian 09d et les calculs MS-CASPT2 le logiciel Molcas 8.0.

Les calculs DFT consistent à optimiser dans un premier temps l'état fondamental S0 avant d'optimiser le 1<sup>er</sup> état excité et d'obtenir la valeur du  $T_e$  calculée correspondant à la longueur d'onde d'émission de fluorescence. L'environnement est représenté par un solvant MeCN modélisé de façon implicite et avec un algorithme de type PCM réponse linéaire. Les calculs DFT ont été réalisés avec la fonctionnelle B3LYP avec la base 6-31+G(d,p).

Les calculs MS-CASPT2 utilisent une base ANO-RCC-VTZP et un espace actif de 14 électrons dans 12 orbitales, représentant l'ensemble du système conjugué de la molécule en excluant les fluor (voir Annexes A4 et A5). Des calculs MS-CASPT2 incluant le fluor montrent néanmoins peu de différences en terme d'énergie d'émission. Les calculs MS-CAPST2 sont réalisés dans le vide et tiennent compte de 4 états électroniques.



# Chapitre 12

## Etude de la tautomérisation keto-enol

### 12.1 Contexte

La bioluminescence chez les lucioles donne lieu à une réaction chimique transformant la luciférine en oxyluciférine, le substrat émetteur. L'émetteur peut adopter six formes chimiques différentes présentée dans la figure 12.1. Deux facteurs permettent de distinguer ces conformations, la protonation du groupement phénol et l'équilibre céto-énolique de l'oxyluciférine. Chen *et al.*<sup>117</sup> ont calculé la longueur d'onde d'émission pour chacune des six conformations à l'intérieur de la protéine et dans différents solvants. Les calculs montrent que les formes neutres ne sont pas des formes émissives. Il est maintenant communément admis que c'est le conformère phénolate-kéto qui est obtenue à la fin de la réaction de bioluminescence après coupure du cycle dioxétanone<sup>18 118 119-121</sup>. Le choix du conformère phénolate-kéto comme émetteur de la bioluminescence est un choix raisonnable et celui choisi lors de cette thèse.

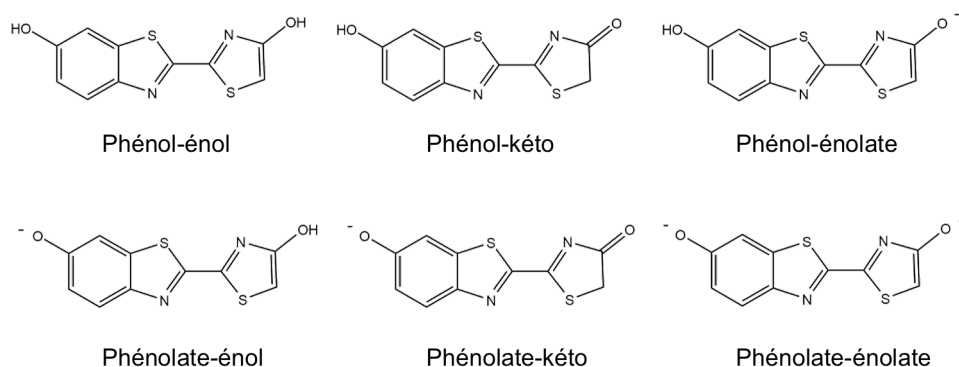


FIGURE 12.1 – Schéma des 6 formes chimiques de l'oxyluciférine.

Cependant, il n'existe aucune preuve expérimentale concernant la forme adoptée par la molécule émettrice. De plus, de nombreuses études réalisées en solution sans protéine donnent des résultats très différents. Dans l'eau, les six formes peuvent exister en solution selon les conditions de pH du milieu et sa polarité<sup>122 123</sup>. D'autres

études expérimentales montrent que la conformation adoptée par la luciférase favorise la forme phénol-énol comme état fondamental le plus stable<sup>124</sup>. Finalement, des mesures expérimentales de la dynamique des états excités font apparaître le rôle du phénolate-énol et du phénolate-énolate comme principaux émetteurs de la fluorescence aussi bien en solution aqueuse que dans la protéine<sup>125</sup>.

Ces différentes études montrent que l'oxyluciférine peut adopter différentes formes chimiques<sup>126</sup>. Dans le cas de la bioluminescence, il n'existe pas de preuves permettant de relier les différentes longueurs d'onde d'émission observées aux six formes chimiques ni de prédire un possible mécanisme réactionnel reliant ces molécules. La possibilité d'une réaction de tautomérisation entre la forme phénolate-kéto et phénolate-énol a déjà été proposée plusieurs fois mais jamais observée. Le but de ce chapitre est double. Dans un premier temps, on cherche à modéliser la tautomérisation kéto-énol à l'intérieur de la protéine puis à suivre ce chemin réactionnel. Le second but de cette étude est de savoir quelle forme chimique est responsable de la bioluminescence chez les lucioles. Une comparaison des barrières énergétiques et de la stabilité des deux conformères est pour cela réalisée aussi bien dans l'état fondamental que dans le premier état excité.

## 12.2 La structure de départ

Afin de pouvoir modéliser l'état de transition représentant la tautomérisation kéto-énol, on choisit de partir d'une structure phénolate-kéto comme structure de départ.

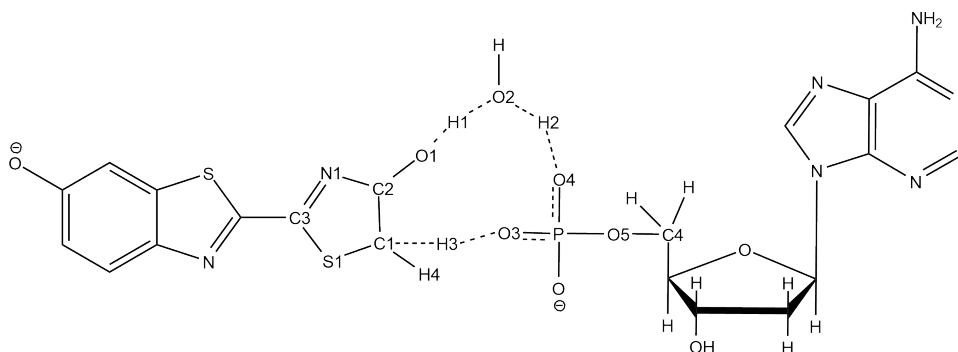


FIGURE 12.2 – Représentation graphique du système étudié dans cette partie. Les traits en pointillés représentent de possibles liaisons hydrogène.

Le modèle choisi est issu du modèle 4G37 où la molécule de DLSA est remplacée par l'oxyluciférine phénolate-kéto et par l'AMPH, la molécule d'AMP avec un oxygène du groupement phosphate protoné. Arbitrairement, l'oxygène pointant vers l'oxyluciférine est celui qui a été protoné. Par la suite 21 MD de 10 ns ont été réalisées afin d'étudier la dynamique du site actif, notamment l'inclusion de molécules d'eau du solvant dans la cavité. Comme évoqué dans les chapitres 9 et 10, la présence de molécules d'eau est cruciale afin de modéliser correctement le système et retrouver

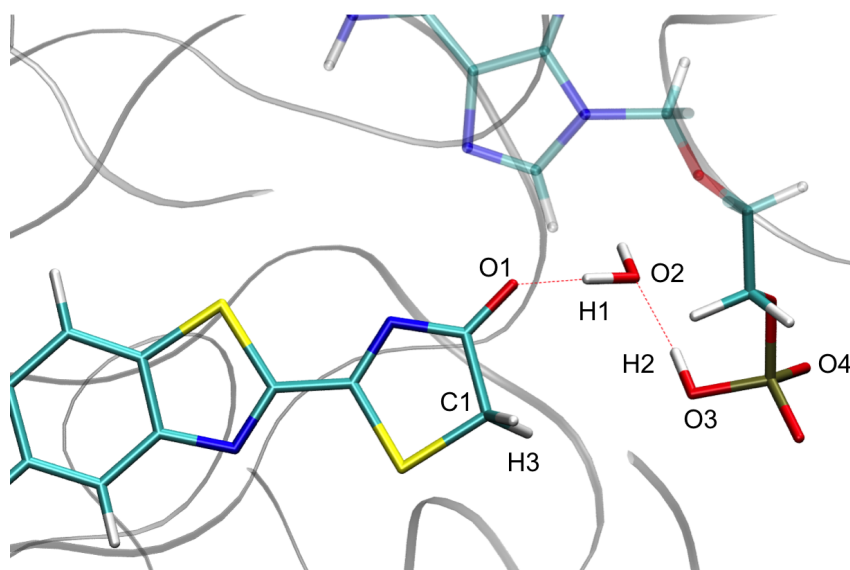


FIGURE 12.3 – Structure kéto choisie comme structure de départ de l'étude de la tautomérisation entre les deux conformères de l'oxyluciférine kéto-énol. Cette structure est issue de la simulation MD numéro 1. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

une valeur de la longueur d'onde d'émission de la bioluminescence proche des valeurs expérimentales. Ces molécules d'eau peuvent également avoir un autre rôle. En effet dans 5 des 21 MD obtenues, une molécule d'eau se place entre l'oxygène kéto de l'oxyluciférine et le groupement O-H du phosphate de l'AMPH, entraînant la formation de liaisons hydrogène avec les deux substrats. Cette molécule d'eau prend cette position plus ou moins rapidement lors des MD. Par exemple, dans la 1<sup>ère</sup> MD celle-ci se place après 1 ns alors que dans la 17<sup>ème</sup> MD, celle-ci arrive vers 6 ns. En revanche, la présence de cette molécule d'eau entre l'oxyluciférine et l'AMPH conduit à une conformation stable. Une fois cette molécule d'eau positionnée, elle reste bloquée entre les deux substrats, et n'est à aucun moment de la MD (sur le reste des 10 ns) remplacée par une autre molécule d'eau. La fréquence d'obtention ainsi que la stabilité de la molécule d'eau montre que le système oxyluciférine, eau et AMPH peut-être considéré comme une structure redondante dans le site actif de la luciférase. Il est envisageable de penser que la molécule d'eau peut servir de pont entre l'oxyluciférine et l'AMPH donnant lieu à un transfert de protons et conduisant à la tautomérisation kéto-énol.

Si on analyse la structure de la figure 12.3, la tautomérisation comporte un transfert de trois protons. L'un des hydrogènes du groupement CH<sub>2</sub> de l'oxyluciférine, nommé H3 peut interagir avec le groupement OH de l'AMPH. L'hydrogène H2 du groupement phosphate interagit avec la molécule d'eau qui en retour donne son hydrogène, nommé H1 à l'oxygène O2 de la fonction cétone de l'oxyluciférine. On a alors un transfert de protons entre quatre centres. Lors de la suite de ce chapitre, le terme phénolate-kéto est également appelé kéto et le terme phénolate-énol, énol.

La structure choisie correspond à celle obtenue dans la 1<sup>ère</sup> MD. Il serait très intéressant d'étudier d'autres structures car on voit apparaître des comportements différents. Ainsi, certains résidus du site actif peuvent également former une liaison hydrogène avec l'oxyluciférine via une molécule d'eau. Les deux principaux acides aminés qui permettent ces liaisons hydrogène relevés dans les MD sont la Lys 443 et la Gly 316. Les calculs présentés à partir de la structure de départ de la 1<sup>ère</sup> MD correspondent déjà à un grand temps de calcul. Nous n'avons pas eu le temps de les reconduire à partir de nouvelles structures issues des autres dynamiques. Nous ne présentons donc dans cette étude que ceux de la structure de départ de la 1<sup>ère</sup> MD.

A partir de la structure de départ, deux études en parallèles sont menées. On modélise le chemin réactionnel entre les deux formes kéto et énol dans l'état fondamental S0 et le 1<sup>er</sup> état excité S1 en prenant en compte soit un environnement implicite, soit un environnement protéique (QM/MM). Le système étudié comporte l'oxyluciférine, l'AMPH et la molécule d'eau située entre les deux (voir figure 12.2).

## 12.3 Dans l'eau en solvant implicite (PCM)

### 12.3.1 Géométrie des états de transition

On réalise dans un 1<sup>er</sup> temps une étude de la tautomérisation kéto-énol dans un solvant implicite afin d'avoir une idée de la forme du TS. De plus, il est possible d'utiliser lors de cette partie l'algorithme QST2, simplifiant l'optimisation du TS. Cette approche n'est pas disponible pour les calculs QM/MM. Des détails supplémentaires sont disponibles dans la partie 12.9.3. Dans le cas du calcul du TS à l'état fondamental dans l'eau, on observe bien comme attendu une réaction de type concertée à 4 centres, C1, O1, O2 et O3, voir figure 12.4. Cependant, seul H3 a une position très différente par rapport à la structure de départ. Cet proton perd sa liaison avec le carbone C1 de l'oxyluciférine et se place maintenant à une distance 1.63 Å du carbone C1. En conséquence ce déplacement conduit l'oxygène O3 du groupement phosphate à porter 2 hydrogènes, H3 et H2 (H3-O3 valant 1.14 Å et O2-H2 1.03 Å). Finalement la molécule d'eau se place de façon à avoir H1 orienté vers O1 de l'oxyluciférine tout en restant fortement lié à O2 (0.98 Å). De plus, l'oxygène O2 de l'eau forme également une liaison hydrogène valant 1.48 Å avec H2 lié à O3, l'oxygène du groupement phosphate de l'AMP. Cette structure du TS dans S0 montre aussi une torsion de l'ordre de 20° du cycle thiazolone. Finalement la structure de transition dans S0 dans l'eau, se rapproche plus de la forme phénolate-kéto que de la forme phénolate-énol.

La géométrie du TS du S1 est une véritable surprise. On passe d'une réaction à 4 centres à une réaction à 5 centres en incluant l'oxygène O4 du groupement phosphate, voir figure 12.5. Lors de l'optimisation du TS, H2 alors situé entre O2 et O3 se lie à l'oxygène de la molécule d'eau, O2 (0.98 Å) et perd complètement sa liaison avec l'oxygène O3 de l'AMPH. En réaction, H1 casse sa liaison avec O2 et se rapproche de l'oxygène O1 de l'oxyluciférine. Ce déplacement de protons conduit également à une rotation de 30° du groupement phosphate de l'AMPH et rapproche de l'oxygène



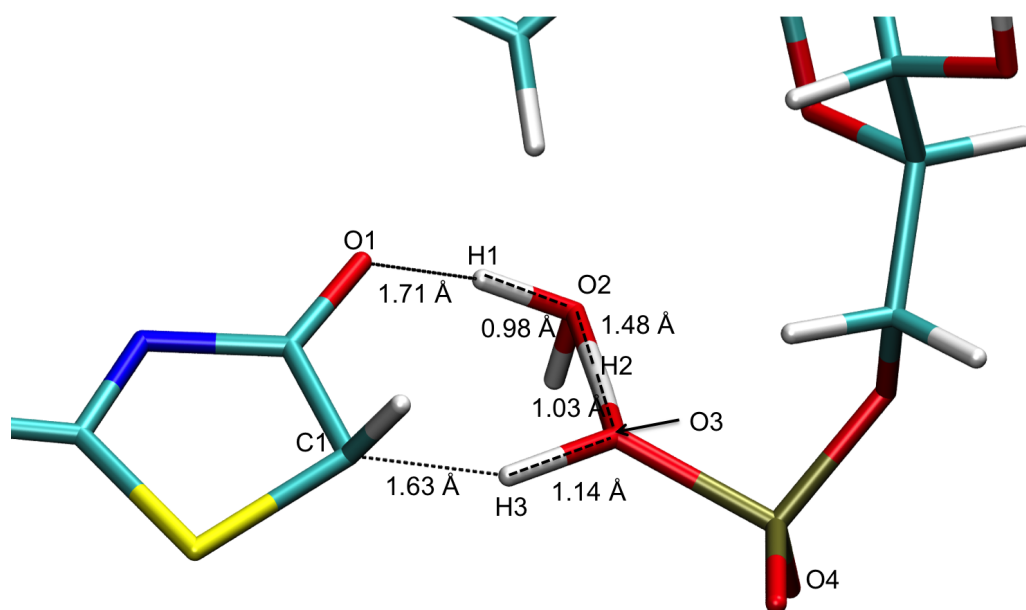


FIGURE 12.4 – Zoom sur la structure de l'état de transition obtenue en PCM dans l'état fondamental S0. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

O4 de H2 (1.76 Å). Dans le TS optimisé au 1<sup>er</sup> état excité, on a H1 lié avec O1 (1.01 Å) et formant une liaison hydrogène avec O2 (1.54 Å). Enfin, H3 conserve sa position entre C1 et O3, sa distance avec O3 augmente légèrement (1.27 Å) et celle avec C1 diminue (1.41 Å) par rapport au TS dans S0. La torsion entre le cycle thiazolone et le cycle benzothiazole augmente et est maintenant de l'ordre de 30°. Cette structure de TS dans l'eau à l'état S1 est plus proche de la forme phénolate-énol que de la forme phénolate-kéto, dû à la présence d'un proton sur l'oxygène O1 de l'oxyluciférine.

### 12.3.2 Les IRC

A partir des structures obtenues pour les deux TS, on a réalisé des profils énergétiques, aussi bien pour le S0 que pour le S1. La molécule obtenue pour une coordonnée intrinsèque de réaction (IRC) positive correspond à la forme kéto alors que pour une coordonnée intrinsèque de réaction négative on obtient une forme énol. Lors du calcul du profil énergétique dans S0 la forme phénol est plus stable que la forme énol de l'ordre de 0.5 eV soit 11.5 kcal/mol (voir Figure 12.6,12.7).

Dans le cas du S0, les deux formes tautomères sont obtenues assez rapidement, pour une coordonnée de réaction de l'ordre de +8 ua. Ainsi à cette coordonnée, H1, H2 et H3 sont liés aux atomes permettant de former les substrats kéto ou énol mais n'ont pas encore adopté leur conformation finale. Par exemple pour H3 dans la forme kéto, à une coordonnée de +8 ua, il forme une liaison covalente avec C1, mais l'angle H4-C1-H3 n'est pas encore de 70°. Les conformations finales des deux

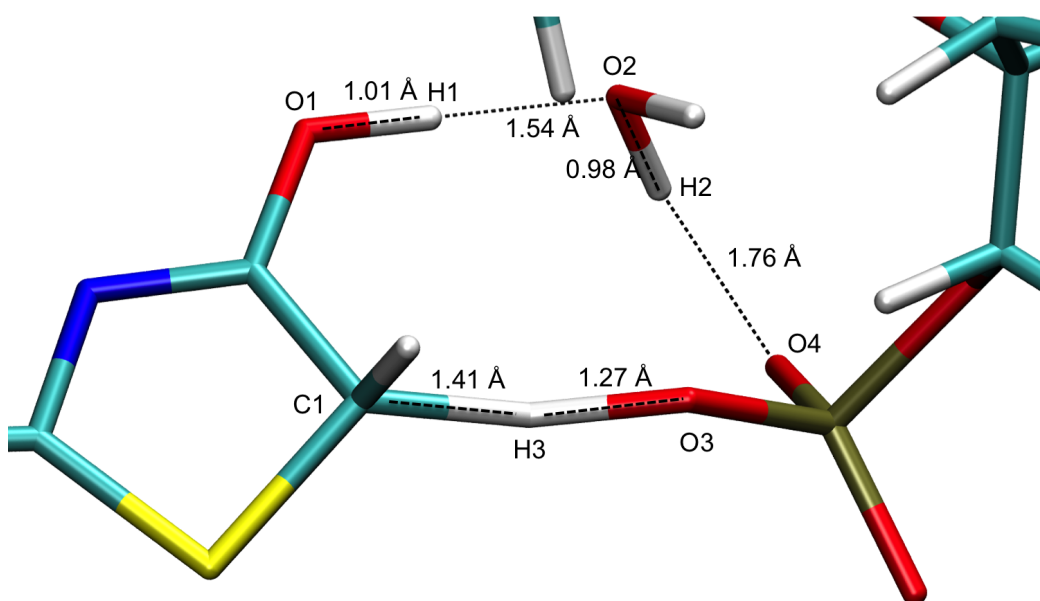


FIGURE 12.5 – Zoom sur la structure de l'état de transition obtenue en PCM dans l'état excité S1. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

tautomères sont observées pour une coordonnée de +12 ua, correspondant à un plateau au niveau de l'énergie.

Le chemin réactionnel vers la forme kéto dans S0 montre une pente importante, similaire à une courbe de type exponentielle décroissante. Lors de ce calcul seul le mouvement de H3 est observé. En revanche entre le TS et la forme énol, le profil énergétique peut se décomposer en deux parties. La première partie correspond au mouvement de H3 qui se déplace vers l'oxygène O3. Lors de cette étape les deux autres hydrogènes H1 et H2 bougent légèrement. Vers une IRC de -3 ua on observe une sorte de palier correspondant à une structure intermédiaire. Puis entre -4 et -8 ua, H1 et H2 bougent de façon concertée afin de former la forme phénolate-énol.

Ce premier chemin réactionnel montre que le mécanisme de tautomérisation est concerté ; on peut passer d'une forme à l'autre par un seul état de transition. Dans le TS, H3 est situé à une position médiane entre C1 et O3, contrairement à H1 et H2. On voit alors que cet proton bouge en premier et de façon indépendante vis à vis des deux autres.

Pour le chemin réactionnel suivi dans le premier état excité, la forme du profil énergétique est inversée par rapport à celui pour S0 (voir Figure 12.7). En effet, comme nous l'avons déjà mentionné, le TS à l'état fondamental à une géométrie proche de la forme phénolate-kéto alors pour le TS au S1, la géométrie se rapproche de la forme phénolate-énol. Ainsi, pour une IRC positif (passage de TS à kéto), on observe le mouvement de H3, suivit d'un palier puis du mouvement de H1 et H2. Le plateau est atteint après environ une CR de +13 ua. En direction de la structure énol, on n'observe que le déplacement de H3, la forme du profil énergétique ressemble

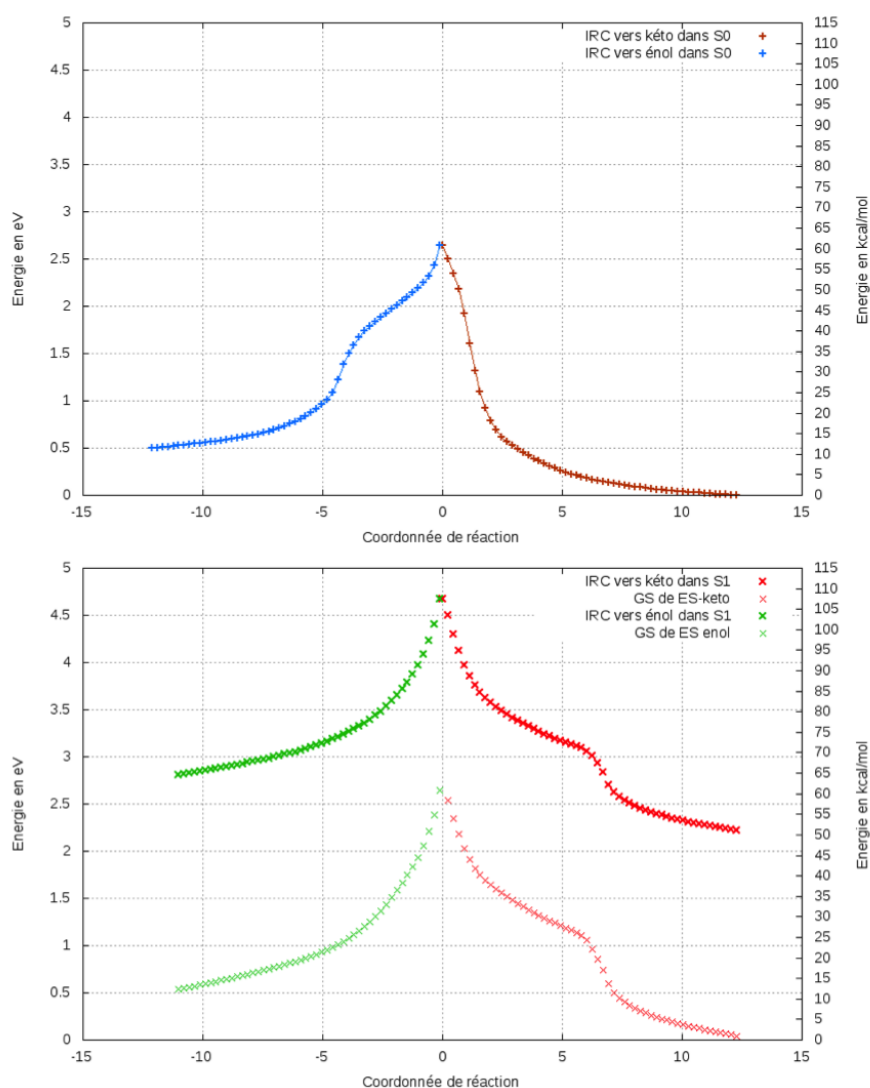


FIGURE 12.6 – En haut, graphique du profil énergétique en fonction de l’IRC en solvant implicite (PCM) dans l’eau pour S0. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0. En marron, la courbe du profil énergétique vers la forme kéto dans S0; en bleu la courbe du profil énergétique vers la forme énol dans S0. En bas, graphique du profil énergétique en fonction de l’IRC en solvant implicite (PCM) dans l’eau pour S1. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS).

alors à une exponentielle décroissante (voir Figure 12.6,12.7).

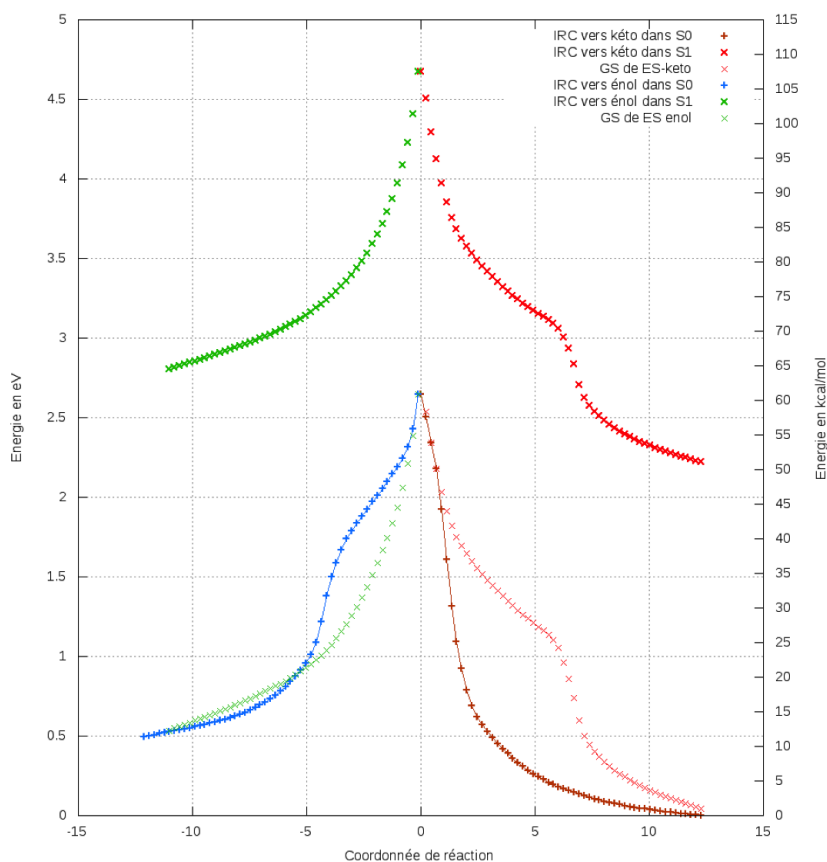


FIGURE 12.7 – Graphique du profil énergétique en fonction de l'IRC en solvant implicite (PCM) dans l'eau. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0. En marron, la courbe du profil énergétique vers la forme kéto dans S0 ; en bleu la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1 ; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1 ; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS).

Le changement de géométrie du TS entre S0 et S1, passant d'une réaction à 4 centres à une réaction à 5 centres, n'affecte pas les étapes de la tautomérisation. La réaction est toujours concertée avec H3 le premier proton affecté. La différence principale est que l'hydrogène porté par l'AMPH n'est plus situé sur le même oxygène selon la forme tautomérique finale obtenue à la fin du profil énergétique le long de la IRC. Pour la forme énol, c'est l'oxygène O3 qui porte l'hydrogène alors que pour la forme kéto c'est l'oxygène O4.

En terme d'énergie, la forme phénolate-kéto dans l'état S1 est comme pour l'état S0

plus stable que la forme phénolate-énol, de l'ordre de 0.56 eV soit 13 kcal/mol, une valeur comparable avec celle obtenue pour l'état fondamental.

Lors de calcul du chemin réactionnel pour le premier état excité, les points correspondants aux valeurs Franck-Condon de l'état fondamental ont également été relevés. Les deux courbes ainsi obtenues ont une représentation très similaire à celles obtenues pour le S1.

Comme les deux TS obtenus pour S0 et S1 ne donnent pas lieu au même mécanisme réactionnel, le TS dans S0 correspondant à une réaction sur 4 centres et le TS dans S1 à une réaction sur 5 centres, on ne compare pas les deux profils énergétiques obtenus.

Finalement, si on calcule les barrières énergétiques on trouve pour le S0 une valeur de 2.23 eV (61 kcal/mol) entre le TS et le dernier point du profil énergétique pour la structure phénolate-kéto ainsi qu'une valeur de 2.14 eV (49 kcal/mol) entre le phénolate-énol et le TS. Pour le S1, la valeur est de 2.45 eV (56 kcal/mol) entre le dernier du point du profil énergétique pour la structure phénolate-kéto et de 1.87 eV (43 kcal/mol) entre le phénolate-énol et le TS. Les barrières obtenues pour le S1 sont légèrement plus basses que celle dans le S0 de l'ordre de 5 kcal/mol. En outre, comme la molécule de phénolate-kéto est plus stable énergétiquement, les barrières calculées pour passer du kéto au TS sont plus élevées que pour celles pour passer au tautomère phénolate-énol.

Les valeurs des barrières calculées dans l'eau sont assez élevées, supérieures à 40 kcal/mol. Elles restent néanmoins accessibles surtout pour la réaction de tautomérisation de l'énol vers le kéto.

### 12.3.3 Conclusion et perspectives

L'étude de la tautomérisation de l'oxyluciférine dans un solvant implicite n'est qu'une étude préliminaire avec de nombreuses approximations et ne représente pas le comportement adopté par l'oxyluciférine dans l'eau. En effet, notre modèle comporte la molécule d'AMPH qui est absente dans les études expérimentales. En outre, le solvant est représenté de façon implicite, on ne tient alors pas compte des interactions qui peuvent s'exercer entre le système et le solvant. Notons par exemple la présence de liaisons hydrogène pouvant stabiliser les deux formes tautomères de l'oxyluciférine.

En revanche, les structures des états de transition obtenues montrent que deux différents mécanismes peuvent exister, un à 4 centres et un à 5 centres. En outre, les barrières calculées montrent que la réaction de tautomérisation entre la forme phénolate-kéto et la forme phénolate-énol de l'oxyluciférine est réalisable. Cette réaction ne contient qu'un seul état de transition et adopte un mécanisme concerté.

En terme de perspectives, la prochaine étape de cette étude est réaliser cette

tautomérisation avec des molécules d'eau explicites sans AMPH. Malheureusement des calculs de MD préliminaires montrent que le réseau de molécules d'eau autour de l'oxyluciférine ne se place pas de telle façon à avoir la tautomérisation kéto-énol possible. Cette étude reste alors à réaliser dans un futur proche. Dans le paragraphe suivant on discute de la réaction de tautomérisation dans la protéine par des calculs QM/MM.

## 12.4 Dans la protéine

Dans cette deuxième partie, on s'intéresse à la tautomérisation kéto-énol à l'intérieur de la protéine. L'apport de la protéine sur le système étudié et sur la réaction de tautomérisation est primordial si on veut étudier ce mécanisme réactionnel lors de la bioluminescence. On a vu dans les chapitres 9 et 10 que la protéine apporte des effets de type électrostatique, via la polarité du site actif et par la création de liaisons hydrogène avec l'oxyluciférine et l'AMP. Le deuxième type d'interaction est un effet de contrainte stérique de la protéine. Les substrats présents à l'intérieur de la cavité ont une marge de mouvement limitée.

### 12.4.1 Géométrie des états de transition

La structure de départ utilisée pour le calcul des TS en QM/MM est dérivée de celle présentée dans le paragraphe 12.2. Elle est extraite d'une *image* issu d'une MD et minimisée de façon classique. A partir de cette structure les atomes de l'oxyluciférine, de l'AMPH et de la molécule d'eau participant à la réaction de tautomérisation sont remplacés par leurs équivalents obtenus pour la géométrie du TS dans l'état fondamental en solvant implicite. La structure de départ lors de l'optimisation du TS dans la protéine adopte donc la conformation du TS calculée lors d'une réaction à 4 centres. On commence par optimiser le TS dans la protéine pour l'état S0. On réalise un calcul d'optimisation DFT/MM avec dans la partie QM, l'oxyluciférine, l'AMPH et la molécule d'eau. Le reste du système appartient à la partie MM. La géométrie du TS obtenue montre une réaction de tautomérisation faisant interagir 5 centres réactionnels comme dans le cas du TS1 trouvé en PCM.

Les calculs d'optimisation DFT/MM présenté précédemment conduisent à un résultat inattendu, deux TS différents ont été obtenus, permettant de réaliser la réaction de tautomérisation de deux façon différentes. Ces deux TS possèdent des géométries proches mais le deuxième TS noté TS2 est plus bas en énergie que le premier TS, TS1. La différence principale entre la géométrie de ces deux TS est la position de H1 et H2. Le TS1 a été obtenu à partir de l'optimisation DFT/MM sur la géométrie du TS dans l'état S0 en PCM. Pour le TS2, c'est lors de l'optimisation DFT/MM du 1<sup>er</sup> point du profil énergétique en direction du tautomère kéto que l'on a obtenue ce TS. L'information importante que montre ces deux TS est que le système peut alors choisir d'emprunter le TS1 ou le TS2 afin de passer de la forme

kéto à la forme énol et inversement.

Intéressons nous au TS1, le premier trouvé lors du calcul d'optimisation (voir Figure 12.8). Dans ce dernier, la géométrie adoptée est proche de celle du TS dans le 1<sup>er</sup> état excité en PCM. En détails, on a une structure de départ proche du tautomère phénolate-énol. La distance O1-H1 est de 1.00 Å, H1-O2 vaut 1.62 Å, O2-H2 vaut 0.99 Å et H2-O4 vaut 1.64 Å. La position de H3 est toujours située entre C1 et O3, la distance C1-H3 valant 1.46 Å et H3-O3 vaut 1.28 Å. La présence de la protéine bloque la torsion entre les deux cycles de l'oxyluciférine, ce dièdre valant 10° soit une valeur proche de celle observée après une simulation MD. L'influence de la protéine se ressent également par la création de liaisons hydrogène entre l'oxyluciférine, l'AMPH et l'environnement protéique proche. La partie phénolate de l'oxyluciférine et l'AMPH conservent les liaisons hydrogène créés lors de la MD.

Contrairement au résultat obtenu en solvant implicite dans le paragraphe 12.2.3, la géométrie du TS dans le S1 est identique à celle du S0, au centième d'angstroms près. Ce 1<sup>er</sup> TS est nommé TS1, avec un proton (H1) déjà présent sur l'oxygène O1 de l'oxyluciférine, comme dans le cas du tautomère phénolate-énol.

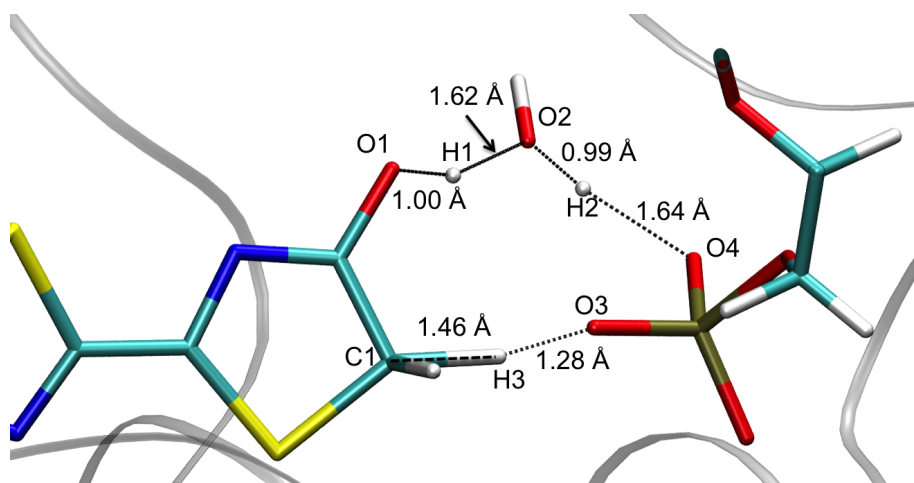


FIGURE 12.8 – Représentation de l'état de transition appelé TS1 et dans l'état excité S1. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

Le 2<sup>ème</sup> TS ressemble pour sa part plus au tautomère phénolate-kéto (voir Figure 12.9). Comme dans le cas du TS1, ce dernier montre une réaction de tautomérisation avec 5 centres. Dans le TS2 on observe une torsion importante du cycle thiazolone par rapport au cycle benzothiazole. Cette torsion est de 20° pour le TS dans l'état fondamental et de 25° dans le S1. Cette torsion affecte notamment les distances entre les centres réactionnels et les hydrogènes associés mais également la position de la molécule d'eau dans la cavité. Cette dernière se déplace en direction du cycle adénosine, la distance absolue relevée entre les oxygènes O2 de l'eau dans le TS1 et le TS2 est de 1.25 Å, lorsqu'on fait une superposition de l'ensemble des deux structures

protéiques. En revanche, la position de H3 entre le carbone C1 et l'oxygène O3 est la même si on la compare avec le TS1. La distance C1-H3 est de 1.62 Å et la distance H3-O3 de 1.13 Å. Pour les autres distance O1-H1 vaut 1.48 Å, H1-O2 est égal à 1.03 Å, O2-H2 vaut 1.43 Å et H2-O2 est égal à 1.04 Å.

Comme énoncé dans un paragraphe précédent les deux TS ne partagent pas la

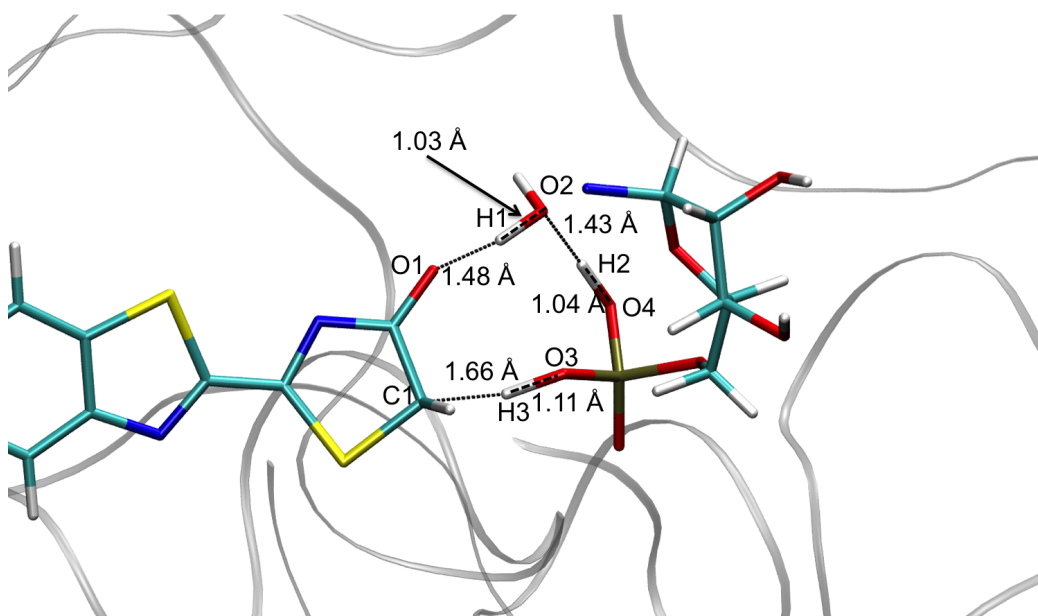


FIGURE 12.9 – Représentation de l'état de transition appelé TS2 et dans l'état excité S1. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

même énergie. Ainsi, le TS2 est plus stable en énergie que le TS1, aussi bien pour S0 que pour S1. La différence d'énergie entre le TS2 et le TS1 pour S0 est de + 1.48 eV soit 34 kcal/mol. Il est pour l'instant très difficile de comprendre pourquoi on a une différence d'énergie si importante entre les deux TS (34 kcal/mol entre les S0 et 26 kcal/mol entre les S1).

## 12.4.2 Les IRC

On va dans un premier temps suivre le chemin réactionnel pour le TS1, puis pour le TS2 et enfin dans une dernière partie on compare les deux chemins obtenus.

### 12.4.2.1 Le TS1

En partant de l'état de transition TS1, les chemins réactionnels, aussi bien en direction du tautomère phénolate-kéto que du phénolate-énol, mettent plus de temps à se stabiliser que pour les profils énergétiques obtenus en PCM. Cependant, si on regarde les structures obtenues au fur et à mesure le long de la IRC, on se



rend compte que les structures désirées sont obtenues bien avant la fin du profil énergétique calculé vers une coordonnée intrinsèque de réaction de l'ordre de 12 à 13 ua. Les points situés après ces valeurs montrent surtout un déplacement erratique de OxyLH<sub>2</sub> et de l'AMPH à l'intérieur de la cavité de la protéine. Comme dans un solvant implicite le tautomère phénolate-kéto est plus stable que le tautomère phénolate-énol. Pour l'état fondamental cette différence est minime en terme d'énergie de l'ordre de 5 kcal/mol alors que pour le S1 elle est plus élevée, de 11 kcal/mol. Le chemin réactionnel est le même qu'observé dans le cas du TS pour l'état S1 dans l'eau (voir Figure 12.10). Si on se dirige vers le tautomère phénolate-kéto, H3 est le premier à se déplacer, suivi de H1 et H2 de façon concertée. Cela est perceptible sur la courbe de l'énergie en fonction du IRC avec un épaulement le long du chemin réactionnel. Avant l'épaulement seul H3 s'est déplacé alors que lorsque le sommet de la bosse est atteint, on a la cassure de liaison entre H1-O1 et H2-O2 et le début de la formation des liaisons H1-O2 et H2-O4 (voir Figure 12.13). Le chemin réactionnel en direction du phénolate-énol montre tout d'abord le déplacement de H3 conduisant à une perte d'énergie jusqu'à la coordonnée de réaction valant -4. Par la suite, les substrats énol et l'AMPH s'écartent progressivement l'un de l'autre, permettant à H3 de rentrer dans le plan formé par le cycle thiazolone (dièdre N1-C2-C1-H4 valant 0°). On remarque également une diminution de l'angle dièdre N1-C2-O1-H1 passant de 150° au TS à 107° à la fin du chemin réactionnel (voir figure 12.11).

Si on compare le profil énergétique pour S0 et pour S1 on remarque que la différence d'énergie entre le TS et la fin du chemin réactionnel est plus importante dans le S0 que dans le S1. En terme de barrières énergétiques, celles obtenues pour le passage du tautomère phénolate-kéto à l'énol sont de 105 kcal/mol pour le S0 et de 88 kcal/mol dans le S1. Pour le tautomère phénolate-énol au kéto on obtient une barrière de 100 kcal/mol pour le S0 et de 71 kcal/mol dans le S1. De plus, au niveau géométrique l'épaulement en direction du phénolate-kéto intervient plus rapidement pour S0 que pour S1. L'épaulement est également plus difficilement perceptible pour S1 alors qu'il est visible dans S0. Dans le cas du phénolate-énol, la pente du profil énergétique est plus importante dans le cas du S0. Si on trace l'énergie des deux tautomères pour l'état fondamental calculé à partir de l'énergie S1 par transition FC (S1GS) on a une bonne corrélation entre ces valeurs et celles calculées dans S0. Dans le cas du phénolate-kéto, l'écart entre S1GS et S0 s'accroît au fur et à mesure du profil énergétique mais l'énergie de S1GS est toujours supérieure à celle de S0. Dans le cas du phénolate-énol, les deux courbes sont superposées jusqu'à la coordonnée de réaction valant -4 ua. L'écart entre S1GS et S0 ne fait que s'agrandir par la suite.

A partir du tracé de S1 et S1GS on peut calculer la valeur du  $T_e$  correspondant à l'émission de type fluorescence. Cette valeur donne également des informations sur l'avancement de la réaction. Pour cela, on compare la valeur du  $T_e$  obtenue pour le dernier point du profil énergétique avec la valeur du  $T_e$  obtenue lors d'un calcul TD-DFT/MM sur une *image* d'un système avec le substrat phénolate-kéto ou le substrat phénolate-énol présent dans la cavité. Par exemple, la valeur de référence du  $T_e$  pour

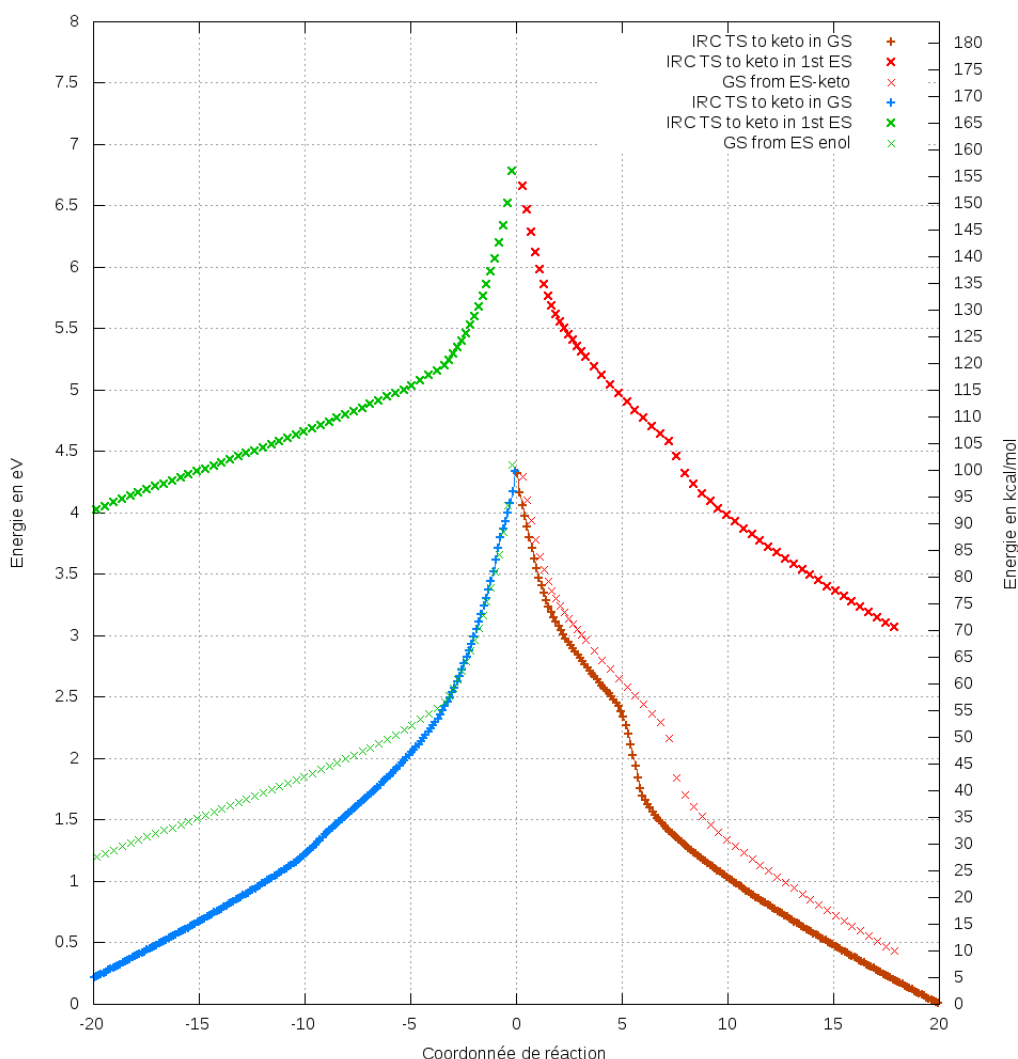


FIGURE 12.10 – Graphique du profil énergétique en fonction de l'IRC dans la protéine pour le TS1. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0 dans la protéine. En marron, la courbe du profil énergétique vers la forme kéto dans S0 ; en bleu la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1 ; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1 ; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS).

le substrat phénolate-kéto est issue d'une calcul TD-DFT/MM sur la structure de départ, discutée dans le paragraphe 12.2 (voir Figure 12.3). Ces valeurs de références sont explicitées dans les détails de calculs 12.9. Cette comparaison permet de voir

Tableau 12.1 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. La valeur de référence provient de calculs TD-DFT/MM effectués sur une *image* extraite d'une MD sur un modèle contenant le substrat phénolate-kéto et sur un modèle contenant le substrat phénolate-énol. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure obtenue dans l'état S1 avec la même fonctionnelle.

TD-DFT/MM		MO6-2X	B3LYP
		$T_e$ eV (nm)	$T_e$ eV (nm)
pour phénolate-kéto	Dernier point de l'IRC	2.64 (472)	2.34 (529)
	Valeur ref	2.55 (486)	2.24 (555)
pour phénolate-énol	Dernier point de l'IRC	2.83 (439)	2.52 (492)
	Valeur ref	2.86 (434)	2.47 (502)

si on retrouve bien la longueur d'onde d'émission des deux substrats kéto et énol à partir des profils énergétiques calculés lors de cette réaction de tautomérisation. Si les valeurs des  $T_e$  sont proches, alors on peut en conclure que l'on a bien atteint les produits finaux de la réaction, c'est à dire le substrat phénolate-kéto et le substrat phénolate-énol.

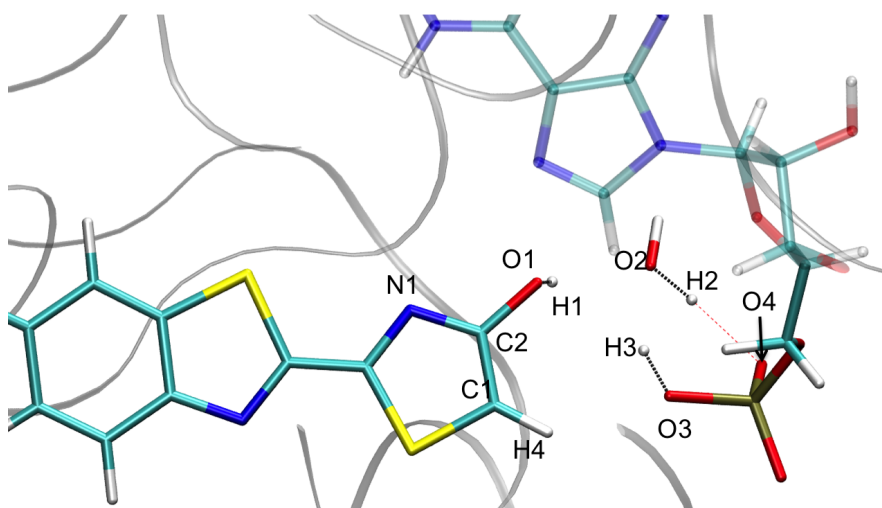


FIGURE 12.11 – Représentation du tautomère phénolate-énol obtenu pour le dernier point du profil énergétique vers la forme énol pour le TS1 et le TS2, dans l'état excité S1 et dans la protéine. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

Les valeurs des  $T_e$  obtenues en fin de réaction sont plutôt proches des valeurs de référence calculées lors de cette thèse. Les résultats obtenus pour le tautomère phénolate-énol ne possèdent que 0.05 eV d'écart avec les valeurs de références. Pour le tautomère phénolate-énol on a une différence de 0.1 eV en moyenne, cette valeur correspondant à la valeur de l'incertitude de la méthode TD-DFT.

L'analyse du TS1 montre que les barrières calculées aussi bien dans le S0 que dans le S1 sont très importantes, toutes supérieures à 70 kcal/mol rendant difficile la réaction de tautomérisation.

### 12.4.2.2 Le TS2

Comme indiqué dans la partie 12.4.1, l'énergie du TS2 est plus faible que pour celle calculée pour le TS1, que ce soit dans S0 ou dans S1. Contrairement au TS1, la géométrie du TS2 se rapproche de la géométrie du tautomère phénolate-kéto. Si on s'intéresse au profil énergétique de la réaction, le tautomère phénolate-kéto est plus stable que le tautomère phénolate-énol de l'ordre de 5 kcal/mol (voir Figure 12.12). Concernant le profil énergétique dans S0, l'épaule est maintenant observé quand on va en direction du tautomère phénolate-énol. En effet, H1 forme une liaison covalente avec la molécule d'eau dans le TS2. Cet épaulement observé le long de la IRC correspond alors à la cassure de cette liaison et au déplacement de H1 vers l'oxygène O1 de l'oxyluciférine. On remarque aussi que la position de l'épaulement intervient à une IRC supérieure à celle de l'épaulement dans le TS1. Le chemin réactionnel en direction du phénolate-kéto montre une baisse très rapide de l'énergie durant les premiers pas du calcul, cette baisse correspondant au mouvement de H3.

D'une façon plus générale, le chemin réactionnel suivi par le TS1 possède le même mécanisme que pour le TS1. L'hydrogène H3 est le premier à bouger suivi de H1 et H2, dans un mouvement concerté.

Pour le profil énergétique calculé dans S1, on ne présente ici que des résultats préliminaires. En direction du tautomère phénolate-kéto, on observe un très léger épaulement au tout début du chemin réactionnel. Celui-ci existe aussi pour S0 mais est moins visible. Par la suite l'énergie diminue drastiquement et se stabilise vers une coordonnée de réaction valant +8 ua. A cette distance, le tautomère phénolate-kéto est déjà visible et on observe seulement un mouvement de l'AMPH dans les pas suivant (voir Figure 12.13). Pour le phénolate-énol, on observe le mouvement de H3 dans un premier temps suivi du déplacement concerté de H1 et H2.

Pour l'état fondamental S0, la barrière phénolate-kéto avec le TS2 est de 65 kcal/mol et la barrière entre le phénolate-énol avec le TS2 est de 58 kcal/mol. Dans le cas du S1, la barrière phénolate-kéto avec le TS2 est de 65 kcal/mol et la barrière entre le phénolate-énol avec le TS2 est de 50 kcal/mol. Si on trace l'énergie des deux tautomères pour l'état fondamental calculé à partir de l'énergie de S1 par transition FC (S1GS) on a une bonne corrélation entre ces valeurs et celles calculées dans S0. Dans le cas du phénolate-kéto, l'écart entre S1GS et S0 s'accroît au niveau de l'épaulement mais reste proche de l'énergie calculée pour le profil énergétique dans l'état S0. Dans le cas du phénolate-énol, les deux courbes restent proches pour l'instant. Pour les deux tautomères l'énergie de S1GS est toujours supérieure à celle de S0 ce qui donne une preuve supplémentaire de la véracité du profil énergétique calculé.

A partir du tracé de S1 et S1GS on peut calculer la valeur de la transition énergétique correspondant à l'émission de type fluorescence comme dans le cas du TS1.

Les valeurs des  $T_e$  obtenues en fin de réaction sont plutôt proches des valeurs de référence calculés lors de la thèse (voir Tableau 12.2). Les résultats obtenus pour le tautomère phénolate-énol ne possèdent que 0.05 eV d'écart avec les valeurs de

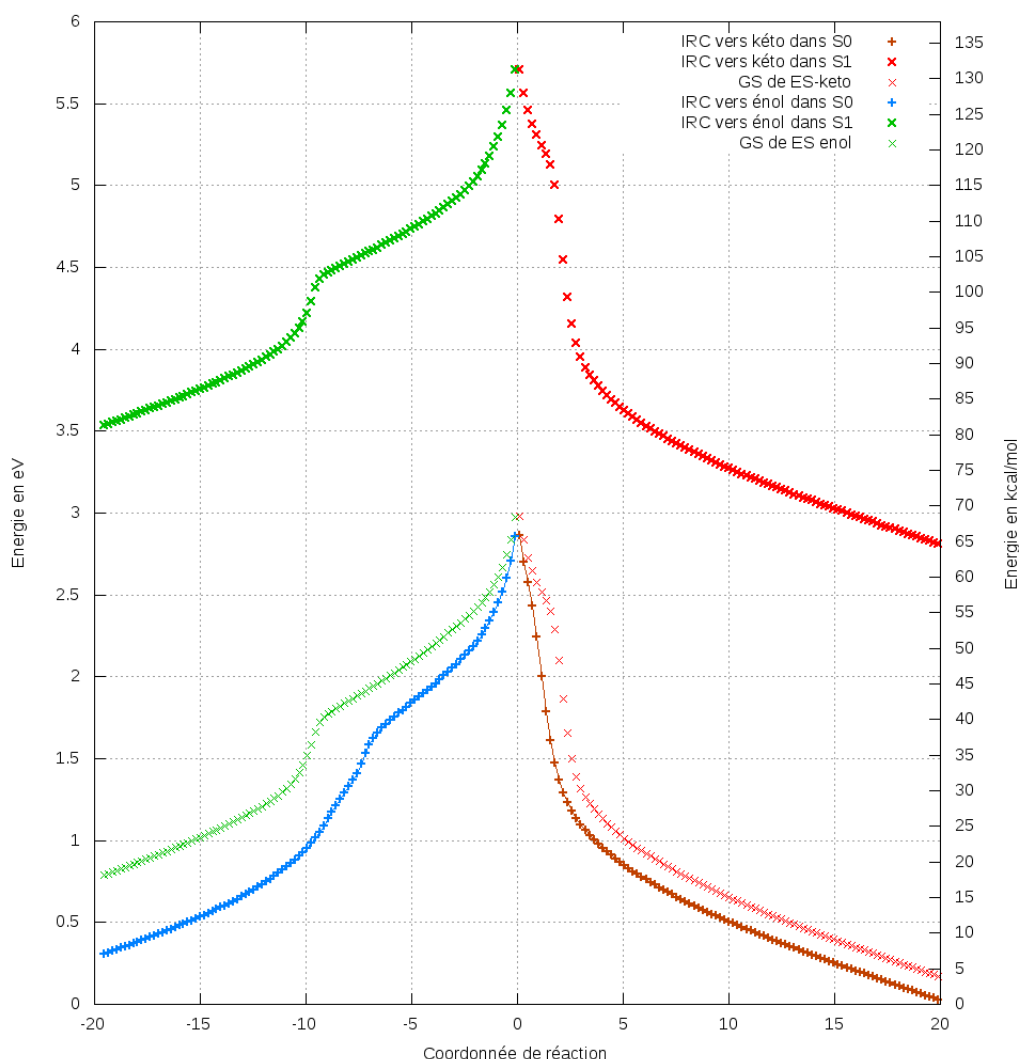


FIGURE 12.12 – Graphique du profil énergétique en fonction de l'IRC dans la protéine pour le TS2. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0 dans la protéine. En marron, la courbe du profil énergétique vers la forme kéto dans S0 ; en bleu la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1 ; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1 ; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS).

références, pour le tautomère phénolate-énol on a en moyenne 0.1 eV, cette valeur correspondant à la valeur de l'incertitude de la méthode TD-DFT.

L'analyse du TS2 montre que l'énergie des barrières obtenues aussi bien dans le

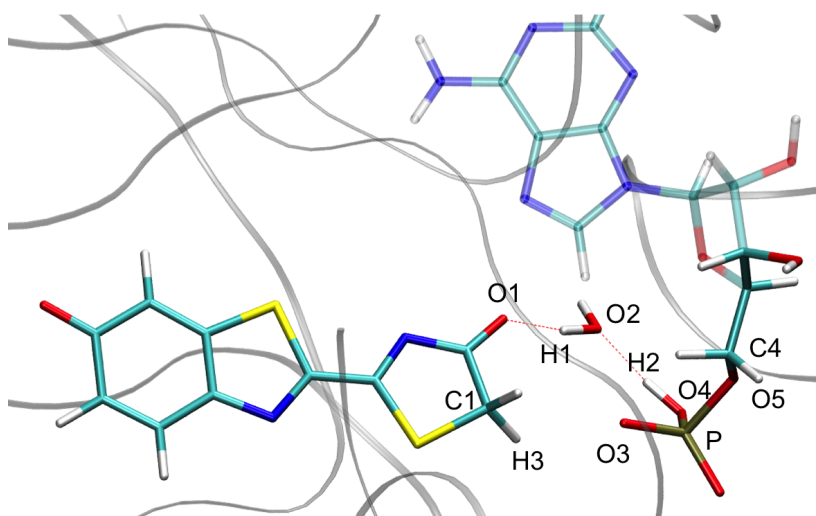


FIGURE 12.13 – Représentation du tautomère phénolate-kéto obtenu pour le dernier point du profil énergétique vers la forme kéto pour le TS1 et le TS2, dans l'état excité S1 et dans la protéine. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

Tableau 12.2 – Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. La valeur de référence provient de calculs TD-DFT/MM effectués sur une *image* extraite d'une MD sur un modèle contenant le substrat phénolate-kéto et sur un modèle contenant le substrat phénolate-énol. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure obtenue dans l'état S1 avec la même fonctionnelle.

TD-DFT/MM		MO6-2X	B3LYP
		$T_e$ eV (nm)	$T_e$ eV (nm)
pour phénolate-kéto	Dernier point de l'IRC	2.64 (470)	2.36 (526)
	Valeur ref	2.55 (486)	2.24 (555)
pour phénolate-énol	Dernier point de l'IRC	2.73 (451)	2.59 (479)
	Valeur ref	2.86 (434)	2.47 (502)

S0 que dans le S1 est un peu supérieure aux valeurs habituellement trouvées dans la littérature<sup>127,128</sup> (entre 25 et 45 kcal/mol). La réaction de tautomérisation semble quand même possible en considérant ces valeurs.

## 12.5 Rotation

On discute dans ce paragraphe de l'effet induit par la rotation de du groupement phosphate de l'AMPH dans la protéine. Lors de la réaction de tautomérisation, lorsqu'on forme le substrat phénolate-kéto à partir de l'état de transition, la géométrie du système obtenue est différente de la géométrie du départ, celle discutée dans la section 12.2. En effet, ce n'est plus le même oxygène de l'AMPH qui porte l'hydrogène. Ce résultat est observé dans les substrats kéto formés à partir du TS1 et du TS2

dans la protéine, aussi bien pour S0 que pour S1. Dans le cas de la structure de départ c'est l'oxygène O3 alors que après la tautomérisation c'est l'oxygène O4. On se propose de calculer le chemin énergétique correspondant à la rotation autour de la liaison P-O5 afin de déplacer l'oxygène O4 vers la position de l'oxygène O3. Pour cela, un calcul de scan le long du dièdre O4-P-O5-C4 est réalisé à partir du dernier point du profil énergétique en direction du tautomère phénolate-kéto obtenu dans l'état fondamental S0 pour le TS1 et le TS2 dans la protéine. Il est en effet plus simple de réaliser ce calcul pour un état fondamental, le temps de calcul requis pour le scan dans S1 est beaucoup trop important.

Dans le graphique 12.14, on voit que cette rotation est facilement réalisable, la valeur

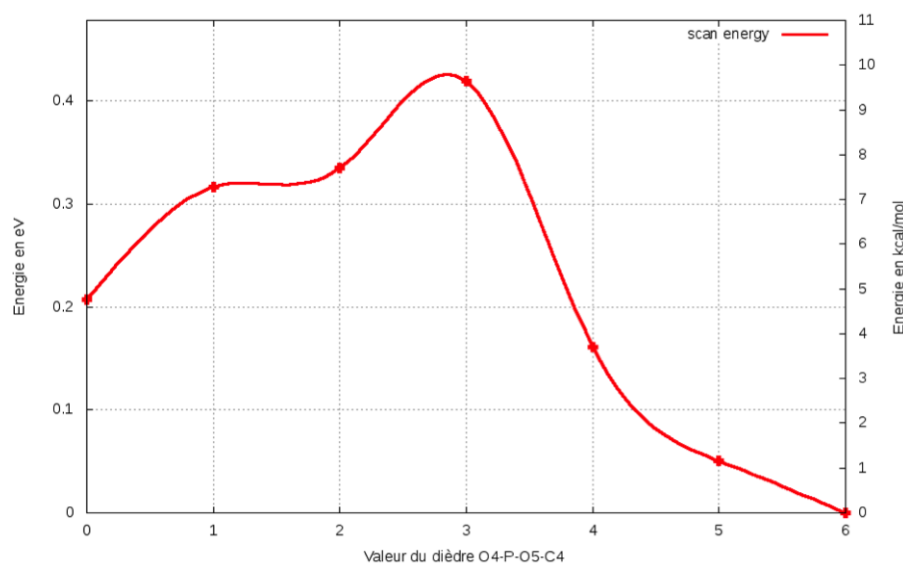


FIGURE 12.14 – Graphique du scan réalisé autour du dièdre O4-P-O5-C4. A gauche la structure issue du dernier point de l'IRC vers le phénolate-kéto, à partir de l'état de transition figure 12.13. A droite, l'énergie de la structure après la rotation de 60° du dièdre. Cette structure correspond à la structure de gauche dans 12.15 et est similaire à la structure de départ.

de la barrière entre les deux conformations est de 5 kcal/mol. De plus, l'énergie de la structure avec l'oxygène O4 pointé vers l'oxyluciférine, comme dans la structure de départ, est plus stable en énergie. La différence d'énergie entre les deux structures du graphique 12.13 et 12.15, est de l'ordre de 4.8 kcal/mol.

On montre qu'il est alors possible de revenir à la structure de départ (voir Figure 12.15) suite au chemin réactionnel de la tautomérisation.

## 12.6 Discussion

On a présenté dans les sections précédentes, la géométrie de TS dans la protéine pour S0 et S1 pour le TS1 et le TS2 ainsi que les IRC correspondants. Ici nous allons comparer ces deux TS et comprendre comment peut se dérouler la tautomérisation

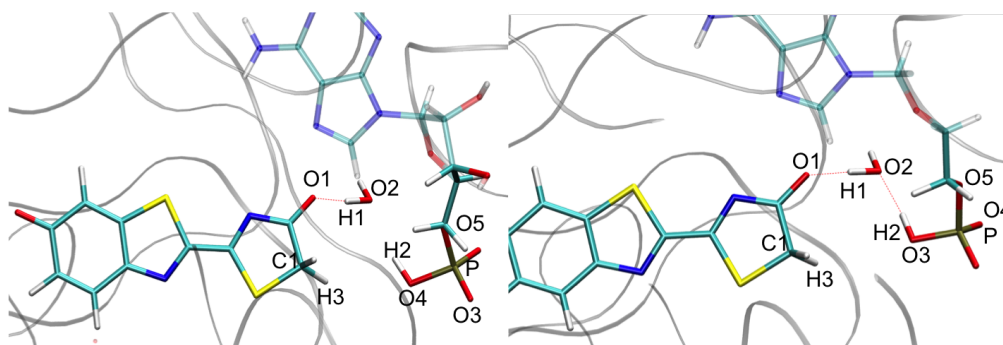


FIGURE 12.15 – A gauche, représentation du tautomère phénolate-kéto après la rotation de  $60^\circ$  du dièdre O4-P-O5-C4 dans la protéine et dans S0. A droite, représentation du tautomère phénolate-kéto dans la structure de départ à l'intérieur de la protéine et dans S0. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons.

kéto-énol.

Premièrement, cette étude montre que la tautomérisation dans la protéine nécessite plus de centres réactionnel qu'attendu. Ce transfert de protons a lieu sur 5 centres réactionnels entre l'oxyluciférine, l'AMPH et une molécule d'eau. Contrairement à une réaction sur 4 centres comme postulé au début de ce chapitre, cela permet d'accroître l'espace alloué au déplacement des hydrogènes. Cependant dans le cas de la formation du tautomère phénolate-kéto cette réaction sur 5 centres ne permet pas de retrouver la structure initiale où l'hydrogène de l'AMPH est situé sur l'oxygène le plus proche de l'oxyluciférine (voir Figure 12.3). On a calculé l'énergie nécessaire à la rotation de l'angle O-P-O. Le chemin énergétique calculé montre qu'il est possible de faire tourner les atomes d'oxygène du groupement phosphate afin de retrouver l'atome d'oxygène possédant l'hydrogène face à l'oxyluciférine. La barrière trouvée est faible dans le cas du TS1 et du TS2, de l'ordre de 5 kcal/mol et la structure calculée, après rotation est plus stable que la structure phénolate-kéto à la fin de l'IRC.

On s'intéresse maintenant au mécanisme de la tautomérisation. Dans un premier temps on compare les structures finales obtenues à partir de ces deux TS. Quand on superpose les deux graphiques pour l'IRC dans le TS1 et dans le TS2 on voit que les énergies des TS ne sont pas les mêmes. Cependant, les structures finales calculées, pour le tautomère kéto et énol possèdent la même énergie, qu'elles soient issues du TS1 ou du TS2.



## CHAPITRE 12 : ÉTUDE DE LA TAUTOMÉRISATION KETO-ENOL

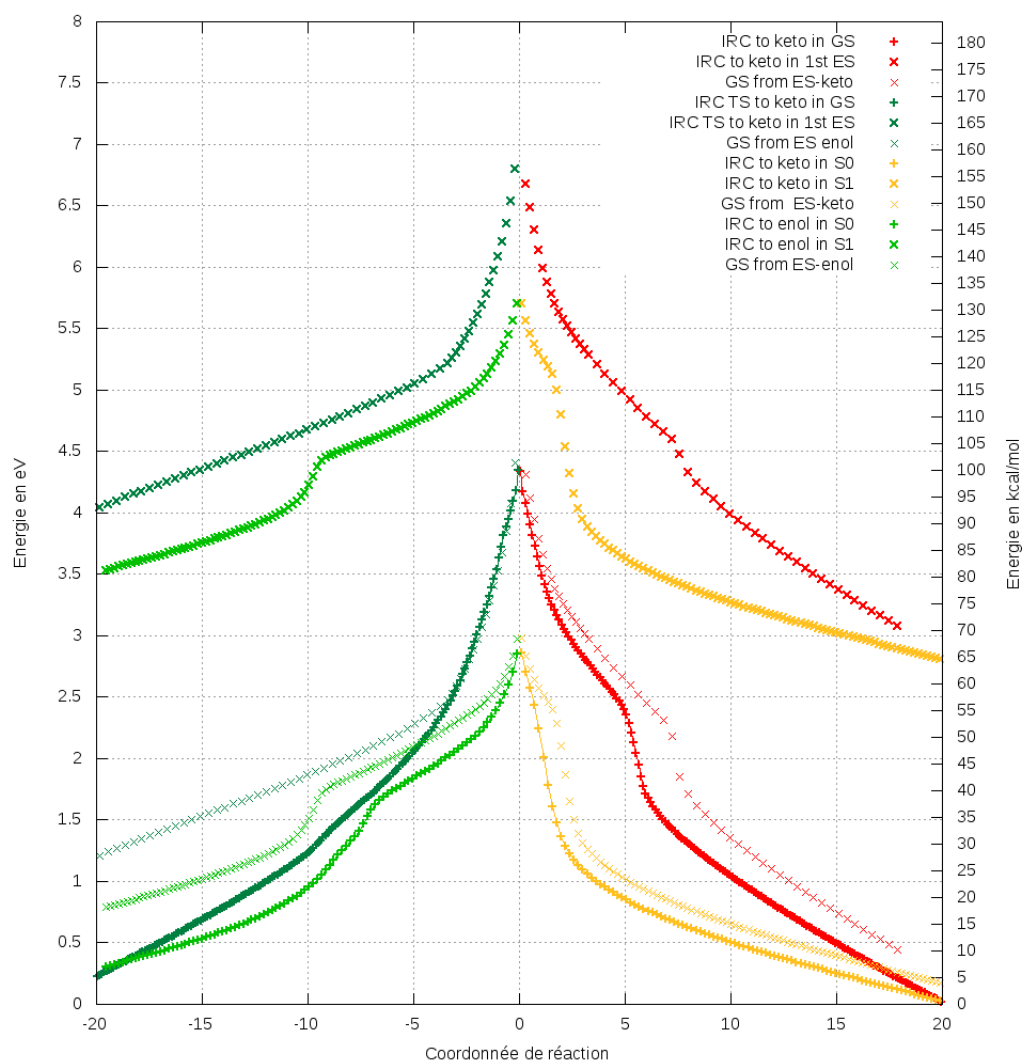


FIGURE 12.16 – Graphique du profil énergétique en fonction de l'IRC dans la protéine pour le TS1 et le TS2. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0 dans la protéine. Pour le TS1, en ligne rouge, la courbe du profil énergétique vers la forme kéto dans S0; en vert la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1. En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1. Pour le TS2, en ligne jaune, la courbe du profil énergétique vers la forme kéto dans S0; en vert clair la courbe du profil énergétique vers la forme énol dans S0. En croix jaune opaque, la courbe du profil énergétique vers la forme kéto dans S1; en croix jaune transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1. En croix verte claire opaque, la courbe du profil énergétique vers la forme énol dans S1; en croix verte claire transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1.

De plus, la géométrie des deux tautomères est également similaire quelque soit le TS de départ. Ces données montrent qu'il existe deux chemins, passant par 2 TS différents permettant de former les deux formes tautomères kéto et énol de l'oxyluciférine (voir Figure 12.6).

Le 2<sup>ème</sup> point important concerne l'énergie des barrières entre la structure finale de l'IRC et l'énergie du TS. Comme l'un des TS est plus bas en énergie, les valeurs des barrières calculées sont plus faibles pour celui-ci que pour l'autre TS. Ainsi, dans le cas du TS2 les barrières montrent une énergie de l'ordre de 55 à 65 kcal/mol alors que dans le TS1, ces barrières sont toutes supérieures à 70 kcal/mol et atteignent même pour l'état S0 des valeurs de 110 kcal/mol.

Enfin, la question restante est : peut-t-on observer la réaction de tautomérisation entre le phénolate-kéto et le phénolate-énol ?

Si on considère l'énergie des barrières pour le TS1, cela est impossible aussi bien dans le S1 que dans le S0. Il faut apporter une énergie de plus de 70 kcal/mol, et pour l'instant il est difficile d'imaginer d'où peut venir cette source d'énergie. Pour le TS2, les énergies des barrières sont plus basses, mais reste supérieur aux énergies de tautomérisation trouvées dans la littérature<sup>127,128</sup>. Néanmoins, une énergie de 56 kcal/mol correspondant à la barrière entre le tautomère phénolate-kéto et le TS semble assez basse pour observer la tautomérisation.

Dans le TS1 et le TS2 on observe que la barrière pour passer du tautomère phénolate-énol au phénolate-kéto est plus basse que pour le chemin inverse. Pourtant expérimentalement le tautomère phénolate-énol est plus souvent observé que

la forme kéto<sup>125</sup>. L'hypothèse explicitée par la suite nous semble raisonnable afin d'expliquer les résultats observés dans la littérature mais il pour l'instant impossible de la vérifier.

On considère les deux chemins réactionnels issus de TS1 et TS2 dans la protéine et seulement pour l'état excité S1. Quand on analyse la géométrie des TS dans la protéine, on remarque que le TS2 est proche de la géométrie du tautomère phénolate-kéto. Ainsi, quand on part du tautomère phénolate-kéto, comme à la fin de la coupure du cycle dioxétanone, il est possible de réaliser la tautomérisation afin d'obtenir le tautomère phénolate-énol via le TS2. Seulement comme la géométrie du TS1 est proche de la géométrie du phénolate-énol on peut penser que c'est ce TS qui est obtenu lorsqu'on part du tautomère phénolate-énol, comme un effet d'hystérésis. Mais cette fois-ci la tautomérisation est impossible à cause de la barrière énergétique élevée (71 kcal/mol). L'oxyluciférine reste alors bloquée dans la conformation phénolate-énol comme observé dans les résultats expérimentaux. Les questions que l'on peut se poser et donnant lieu à de futurs calculs sont : Pourquoi le substrat phénolate-énol ne reprend pas le chemin réactionnel via le TS2 le plus bas en énergie ? Peut-on trouver une preuve afin de vérifier ce mécanisme ? Enfin, on peut également s'intéresser à la formation du substrat phénolate-énolate à partir du phénolate-énol. Cette réaction chimique conduit peut-être à la formation du phénolate-énolate avec une forte stabilisation énergétique et empêchent alors la tautomérisation de l'énol vers le kéto.

## 12.7 Conclusion et perspective

En conclusion, cette étude a montré qu'il est possible de trouver et de suivre un chemin réactionnel permettant de lier deux conformères de l'oxyluciférine, le phénolate-kéto et le phénolate-énol. Cette réaction fait intervenir un transfert de 3 protons entre l'oxyluciférine, l'AMPH et une molécule d'eau.

A l'intérieur de la protéine, deux différents états de transition ont été trouvés, un géométriquement proche du tautomère phénolate-kéto et l'autre proche du phénolate-énol. Le TS proche du phénolate-kéto est plus bas en énergie que pour le deuxième TS. En revanche, l'énergie et la géométrie obtenue pour les deux conformères est la même que l'on emprunte le TS1 ou le TS2. Les valeurs des barrières calculées sont plutôt élevées et comprises entre 55 et 65 kcal/mol pour le TS2 et entre 70 et 110 kcal/mol pour le TS1. Il est néanmoins possible de penser que la tautomérisation peut être observée via le TS2, proche du phénolate-kéto où ces barrières sont de l'ordre de 55 kcal/mol.

Finalement, ces résultats mettent en avant qu'il est possible d'observer plusieurs conformères pouvant jouer un rôle lors de l'émission de la bioluminescence.

Lors de cette étude on n'a étudié que le chemin réactionnel entre 2 conformères de l'oxyluciférine. Or, dans de nombreuses études expérimentales, le rôle du substrat phénolate-énolate est également pointé du doigt dans l'émission de fluorescence.

Par la suite, il faudrait étudier la réaction du substrat phénolaté-énol en phénolate-énolate afin vérifier l'implication de ce substrat dans la bioluminescence.

## 12.8 Détails de calculs

### 12.8.1 Partie MM

Le modèle est construit à partir de la structure cristallographique PDB 4G37 obtenue par le groupe du Prof. Branchini. Cette structure représente le complexe luciférase-luciférine au moment de la coordination du dioxygène. Pour cela, la structure a été modifiée chimiquement en ajoutant un pont disulfure entre les résidus Isoleucine 108 et Tyrosine 447 afin de geler la protéine dans cette conformation. Le fichier PDB 4G37.pdb a été téléchargé à partir du site internet RSCB PDB. Les boucles manquantes de la protéine ont été rajoutées grâce au logiciel Disgro<sup>96</sup>. La molécule DLSA ou 5'-O-[N-(dehydroluciferyl)-sulfamoyl]adénosine a été remplacée par de l'adénosine monophosphate (AMP) et par OxyLH<sub>2</sub>. Les deux molécules SO<sub>4</sub><sup>-</sup> ainsi que le résidu pontant noté XLX ont également été retirés. Le retrait de ce dernier aura peu d'influence sur nos calculs. En effet les dynamiques réalisées n'excèdent pas 10 ns, un temps suffisamment court pour éviter le déplacement de la boucle bloquée par le résidu XLX. Ce modèle est appelé 4G37-OxyLH<sub>2</sub>.

Les résidus ont été protonnés par le programme Leap provenant de la suite logiciel Amber 14<sup>55</sup>. Les cas discutables, notamment pour les résidus histidines ont été élucidés en calculant leur pKa avec le program H++<sup>95</sup>. Les histidines suivantes ont été doublement protonnées : résidus 76, 171, 310, 332, 419, 461 and 489. L'AMPH est chargé -1.

### 12.8.2 Partie MD

Les simulations de dynamiques moléculaires classiques ont été réalisées avec la suite logicielle Amber14 afin d'obtenir une représentation du système pour une future optimisation QM/MM. Les modèles ont été solvatés par des molécules d'eau de type TIP3P dans une boîte de forme cuboïde. Le système résultant contient environ 90000 atomes au total, dont 28000 molécules d'eau formant une sphère de solvation d'au moins 15 Å autour du substrat. Un champ de force de type AMBER99ff a été utilisé pour modéliser les résidus de la protéine. Les paramètres de OxyLH<sub>2</sub> ainsi que ceux de l'AMPH sont développés par notre groupe<sup>111</sup>. Ceux-ci n'étant pas optimisés pour des états excités, la structure et les charges des substrats au niveau excité S1 ont été obtenues par calcul QM/MM préliminaire puis réinjectés dans les fichiers paramètres. Le substrat reste gelé dans cette géométrie pendant toute la dynamique. Le déroulement d'une dynamique est le suivant : 3 minimisations sont d'abord réalisées, la 1<sup>ère</sup> minimise seulement la protéine en gelant les molécules d'eau. La deuxième minimisation ne gèle que les hydrogènes alors que rien n'est gelé (à part le substrat) durant la 3<sup>ème</sup>. Le système est ensuite chauffé de 100K à 300K en 20 ps. Puis sous condition NPT avec T=300K, P=1 atm 21 dynamiques de 10 ns sont conduites sous condition périodique aux bords avec comme pas de temps 2 fs. La

température et la pression sont maintenues en utilisant l'algorithme de Berendsen avec une constante de couplage de 5 ps et par des contraintes de type SHAKE appliquées sur toutes les liaisons impliquant des liaisons hydrogène.

### 12.8.3 Partie QM et QM/MM

Les calculs QM en solvant implicite de type PCM représentant l'eau sont réalisés avec le logiciel Gaussian 09d. La modélisation du solvant implicite est traitée par un algorithme de réponse linéaire pour les calculs dans l'état fondamental et dans le 1<sup>er</sup> état excité.

Les calculs QM/MM sont réalisés en utilisant un schéma de couplage ESPF entre Gaussian<sup>67</sup> et Tinker<sup>56</sup> (Gaussian 09d/Tinker). La méthode ESPF (electrostatical potentiel fitted)<sup>68</sup> permet de calculer les interactions entre les charges de type Mulliken<sup>114</sup> du QM et le potentiel électrostatique externe du MM situé à moins de 9 Å du QM. La technique dite de microitérations est utilisée afin de faire converger la géométrie de l'espace MM choisi pour chaque pas de minimisation du QM. Le substrat, l'AMPH et la molécule d'eau 559 représentent le QM et le reste du système (protéine + molécules d'eau) est inclus dans la partie MM. Les calculs TD-DFT en QM/MM correspondent à une optimisation du 1<sup>er</sup> état excité singulet (S1) suivi par le calcul de la transition énergétique vertical ( $T_e$ ) entre ce même état S1 et l'état fondamental S0. Cette transition correspond à une émission de fluorescence. Chez les lucioles l'énergie de la fluorescence est identique à l'énergie de la transition après la réaction de bioluminescence<sup>19</sup>, la valeur du  $T_e$  calculée peut alors être assimilée à l'énergie d'émission de la bioluminescence.

La fonctionnelle choisie dans cette partie sera la M06-2X. Cette dernière est robuste dans les calculs de réaction chimique ainsi que dans l'optimisation d'état de transition (TS)<sup>129</sup>. En revanche comme vu dans le chapitre 9.3.1 elle ne reproduit pas bien les valeurs du  $T_e$  de l'émission vis à vis des valeurs expérimentales. Des calculs *single points* utilisant la fonctionnelle B3LYP sont réalisés en vue de comparer les valeurs du  $T_e$  le long de l'IRC. La base choisie est la 6-311G(2d,p).

Les deux structures de références utilisées lors de l'étude des valeurs du  $T_e$  le long de l'IRC sont modélisées de cette façon. Pour le substrat phénolate-kéto, une *image* extraite d'une MD sur 4G37-OxyLH<sub>2</sub> a été minimisée de façon MM, puis un calcul TD-DFT/MM a été réalisé sur ce modèle afin d'obtenir la valeur du  $T_e$  en suivant le mode opératoire présenté précédemment. Pour le substrat phénolate-énol, une *image* extraite d'une MD sur 4G37-OxyLH<sub>2</sub>-énol a été minimisée de façon MM, puis un calcul TD-DFT/MM a été réalisé sur ce modèle afin d'obtenir la valeur du  $T_e$  en suivant le mode opératoire présenté précédemment. Ce modèle 4G37-OxyLH<sub>2</sub>-énol correspond à un modèle où le substrat phénolate-kéto a été transformé en phénolate-énol à la main. De nouveaux paramètres ont été développés pour le substrat phénolate-énol.

Le calcul du TS dans l'état fondamental (S0) est dans un premier temps réalisé dans l'eau PCM avec l'algorithme QST2 (The Synchronous Transit-Guided Quasi-

Newton STQN). Ce dernier n'est disponible que pour les calculs dans le vide ou en solvant implicite. L'action de QST2 est de réaliser un calcul d'optimisation de l'état de transition à partir de deux structures de départ, ici la forme phénolate-kéto et la forme phénolate-énol. La structure phénolate-kéto comporte H1 lié à C1, H2 à O3 et H3 à O1. Quand à la structure phénolate-énol, elle comporte H1 lié à O1, H2 lié à O2 et H3 lié à O3. Il est cependant très difficile d'optimiser précisément le TS recherché. Ainsi lorsque l'optimisation du TS atteint un palier en terme d'énergie, la dernière structure est récupérée et une nouvelle optimisation de TS est effectuée à partir de cette dernière structure obtenue en QTS2. La deuxième optimisation du TS est réalisée avec un algorithme classique de type Berny. Afin de simplifier l'optimisation du TS, un certain nombre d'atomes sont gelés pendant l'optimisation. Les cycles adénosine et ribose de l'AMPH ainsi que le cycle benzothiazole de l'oxyluciférine sont gelés lors du calcul. En outre le mot clé "noeigentest" est ajouté. Ce dernier permet de continuer l'optimisation du TS même si plusieurs fréquences négatives sont obtenues.

Le TS optimisé dans le 1<sup>er</sup> état excité (S1), utilise la structure du TS dans le S0. Lors de ce calcul trois états électroniques excités sont inclus dans le calcul.

Les calculs Intrinsic Coordinate Reaction (IRC), sont réalisés en utilisant l'algorithme Local Quadratic Approximation (LQA)<sup>130 131</sup> avec un pas de 0.01 Bohr, noté sous forme de u.a dans le texte. Cet algorithme est le seul fonctionnant à la fois pour le S0 et le S1 et comporte au moins 200 points. Pour l'IRC dans le S1, le mot clé "gradient only" et les IOP 1/11=1, 1/10=5 et 1/71=0 sont ajoutés. Dans la version G09d de Gaussian le calcul des dérivés secondes des états excités n'est pas inclus, le recours à l'algorithme "gradient only" est nécessaire, ce dernier fonctionnant sans ces dérivées secondes. Le premier IOP à le même rôle que noeigentest, le second permet de lire les informations du TS à partir du fichier checkpoint et le dernier permet de calculer la matrice hessienne à partir du fichier checkpoint.

Des calculs de scan de surface d'énergie potentielle relaxée le long d'angle dièdre O4-P-O5-C4 sont réalisés sur le groupement phosphate de l'AMPH. Lors de ce calcul, le ribose, l'OxyLH<sub>2</sub> et l'adénosine sont gelés. Ce dernier comporte 10 pas de -10°.

#### 12.8.4 Discussion sur les méthodes

Les limitations du logiciel Gaussian impacte un peu les résultats. Les approximations utilisées pour le calcul des TS et des IRC donnent lieu à quelques résultats étranges mais ne remettant pas à cause les résultats. Pour les deux TS dans la protéine, aussi bien dans l'état S0 que S1, on obtient 7 fréquences négatives dont une avec une valeur très négative, environ 100 fois supérieure à celle des autres fréquences. C'est cette dernière qui correspond à la tautomérisation kéto-énol. Dans cette fréquence, le mouvement de H3 est visible en premier suivi de H1 et H2. Les autres fréquences correspondent à la respiration des substrats, c'est à dire au mouvement concerté des cycles ribose, adénosine et benzothiazole. Ces fréquences sont

présentes car les atomes de ces cycles sont gelés lors de l'optimisation du TS. La deuxième approximation consiste à utiliser l'algorithme LQA couplée avec le mot clef IOP 1/71=0. Dans ce dernier on utilise la matrice hessienne calculée lors de l'optimisation du TS. Ainsi comme on n'obtient pas l'énergie exacte du TS, le calcul du profil énergétique ne donne pas exactement l'énergie maximale de chemin réactionnel, correspond au TS. Pour les calculs réalisés avec un solvant implicite (PCM), on a testé plusieurs algorithmes de calculs d'IRC différents, à savoir, LQA (Local Quadratic Approximation)<sup>130 131</sup>, HPC (Hessian-based Predictor-Corrector integrator)<sup>132</sup> et DVV (Damped Velocity Verlet Integrator)<sup>133</sup>. Les IRC calculés avec ces trois algorithmes pour l'état fondamental S0 sont totalement identiques. L'algorithme LQA étant le seul disponible pour les calculs dans la protéine, est alors choisi.





# Chapitre 13

## Etude de l'étape de coordination du dioxygène lors de la bioluminescence

### 13.1 Contexte

Depuis une dizaine d'années, de nombreuses études ont permis de mieux comprendre le mécanisme réactionnel de la bioluminescence chez les lucioles. Ces dernières se sont focalisées sur la compréhension du cycle dioxétanone et sur la conformation du substrat lors de l'émission<sup>18 134 117</sup>. Le cycle dioxétanone est un élément primordial dans chaque réaction chimiluminescente. Celui-ci permet d'exciter le substrat émetteur lors de la cassure du cycle conduisant à l'émission de lumière.

En revanche l'étape conduisant à la formation du cycle dioxétanone n'a pas encore été beaucoup étudiée. Le mécanisme communément admis fait intervenir la déprotonation du carbone C1 de l'intermédiaire de réaction noté Int (voir Figure 13.1) par une base, le plus souvent un résidu de la cavité proche du substrat. La suite de la réaction fait ensuite intervenir la molécule de dioxygène et la formation du cycle dioxétanone via un ion superoxyde  $O_2^{\bullet -}$ <sup>135 136 137</sup> (voir Figure 13.2).

Il existe une autre hypothèse pouvant expliquer comment former le cycle dioxétanone. Dans celle-ci c'est la molécule de dioxygène qui va récupérer l'hydrogène attaché sur le carbone C1 via un mécanisme radicalaire. Le formation du dioxétanone passe alors par un intermédiaire hydroperoxyde  $HO_2^{\bullet}$ <sup>2 138</sup> (voir Figure 13.2). Il existe cependant moins de preuves favorisant ce mécanisme.

Le but de ce chapitre consiste à modéliser l'approche et la formation du cycle dioxétanone dans le mécanisme réactionnel de la bioluminescence chez les lucioles. Cette étude est divisé en deux parties, la 1<sup>ère</sup> traite de l'approche de  $O_2$  depuis son entrée dans la protéine jusqu'à arriver à proximité du substrat Int par des simulations MD. Puis, lors de la 2<sup>ème</sup>, on réalise des calculs QM/MM afin de comprendre comment le cycle dioxétanone se forme, selon quel mécanisme et dans quel ordre. Ici, on ne s'intéressera qu'au mécanisme faisant intervenir l'ion superoxyde.

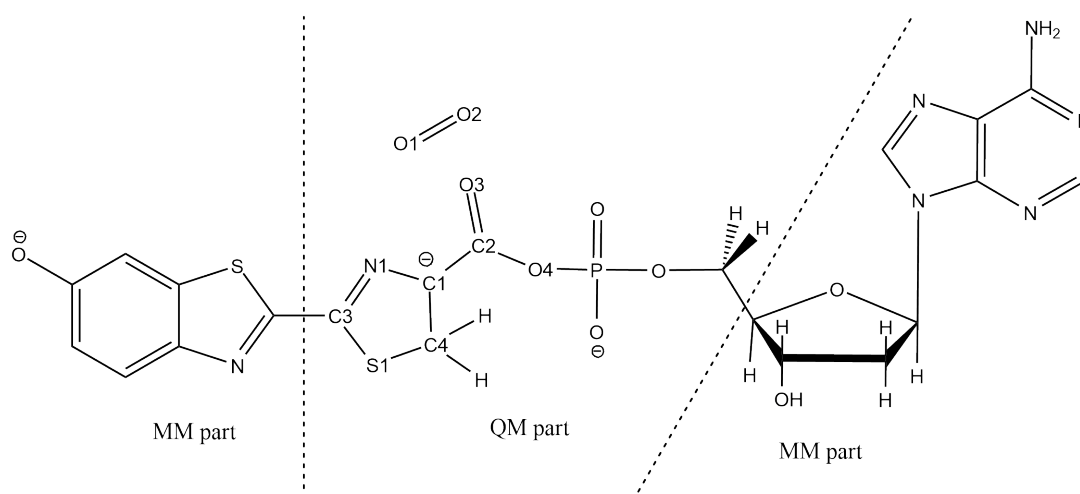


FIGURE 13.1 – Représentation graphique du substrat Int et de la molécule de dioxygène. Les coupures entre la partie QM et les parties MM de ce substrat sont indiqués.

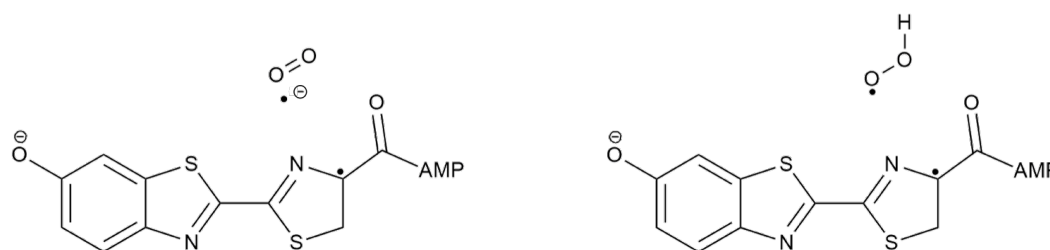


FIGURE 13.2 – Représentation des deux hypothèses proposées pour la coordination de  $O_2$  sur le substrat Int. A gauche via un ion intermédiaire superoxyde (mécanisme étudié dans ce chapitre). A droite via un ion intermédiaire hydroperoxyde.

## 13.2 Déplacement de $O_2$ dans la protéine

La première information à étudier est la possibilité pour le dioxygène  $O_2$  de se déplacer à l'intérieur de la protéine et de se rapprocher de l'intermédiaire de réaction, appelé Int et visible dans la figure 13.1. Selon une publication du Prof. Branchini<sup>2</sup>, la molécule de  $O_2$  passe du solvant à l'intérieur de la protéine par un tunnel que l'on peut voir Figure 13.3.

Un modèle est créé avec la protéine 4G36 de la luciole *Photonus Pyralis* dont on a inséré le substrat Int dans la cavité. On place artificiellement la molécule de  $O_2$  à l'entrée de la cavité, qui correspond au tunnel décrit plus haut. Le modèle est alors noté 4G36- $O_2$ . Le calcul umbrella sampling nous permet d'accéder à la valeur de l'énergie libre du déplacement de  $O_2$  dans la protéine. La molécule de  $O_2$  est placée à 8.5 Å du carbone C1 de Int et alignée de façon à être parallèle au plan du substrat Int. Les minimisations réalisées avant l'umbrella sampling font apparaître un élément

intéressant dans la structure du système. Le résidu Hip 245 situé proche du substrat Int, forme une liaison de type liaison hydrogène avec le carbanion C1, celui qui porte la charge négative.

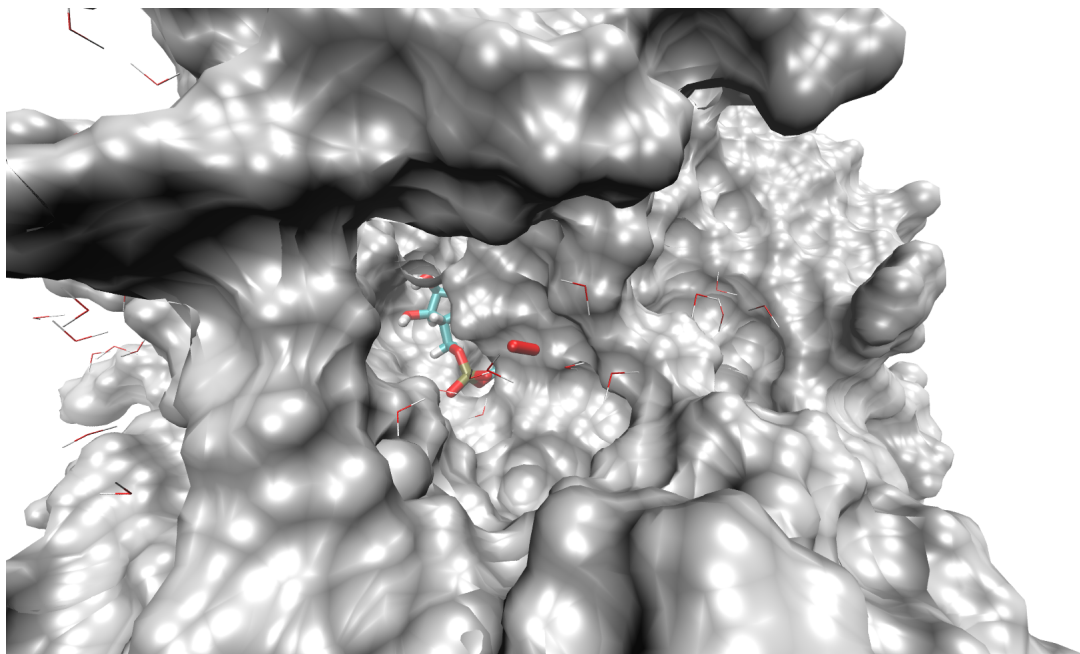


FIGURE 13.3 – Représentation du système avant la simulation umbrella sampling et appelé 4G36-O<sub>2</sub>. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons. O<sub>2</sub> est placé dans un canal joignant le solvant et le site actif de la protéine. La protéine est représentée en dessinant la surface de cette dernière avec un rayon de 1.4 Å et par la couleur grise.

Le calcul d'umbrella sampling est réalisé en prenant comme distance de réaction la distance  $D_{umbrella}$  correspondant à la distance entre le centre de masse de la liaison C1-C2 (molécule Int) et le centre de masse de la liaison O1-O2 (molécule O<sub>2</sub>). Cette distance varie de 8.5 Å à 1.5 Å avec un pas de 0.2 Å. Le graphique obtenu dans la figure 13.4 ne présente aucune barrière énergétique le long de la MD. La molécule de O<sub>2</sub> est libre de se déplacer à l'intérieur de la protéine sans apport d'énergie extérieure. Entre la distance  $D_{umbrella}$  valant 8.5 Å à 3.5 Å la molécule de O<sub>2</sub> se rapproche de plus en plus du substrat Int en restant plus ou moins parallèle au plan du substrat Int incluant la liaison C1-C2. A partir de 3.5 Å les *images* obtenues durant la simulation d'umbrella sampling montrent que la molécule de O<sub>2</sub> ne s'approche désormais plus de la molécule Int et reste dans la même conformation pour les 9 pas restants.

Cette distance de 3.5 Å semble être la distance maximale accessible par une contrainte mécanique avec une approche classique. Au delà il est nécessaire d'utiliser des méthodes quantiques afin de modéliser l'approche de la molécule de O<sub>2</sub>. La structure obtenue à  $D_{umbrella}$  valant 3.5 Å est ensuite extraite et minimisée au ni-

veau MM afin d'obtenir un modèle nommé 4G36-O<sub>2</sub>-min, visible dans la figure 13.5. La structure minimisée montre que la stabilisation de O<sub>2</sub> intervient à une distance  $D_{umbrella}$  égale 3.5 Å. On note la présence de plusieurs liaisons hydrogène reliant le dioxygène aux résidus environnants, notamment l'Hip 245, le résidu choisi pour être protoné en échange de la perte d'un hydrogène sur le substrat Int et une molécule d'eau.

Suite à l'obtention d'une structure stable, la question suivante consiste à savoir si

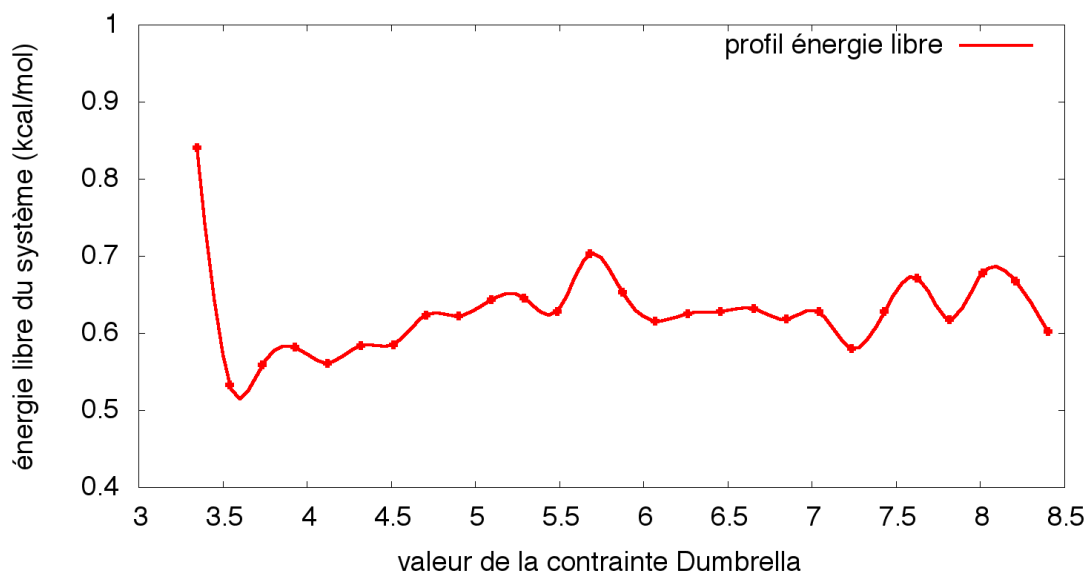


FIGURE 13.4 – Profil de l'énergie libre obtenue lors de la simulation umbrella sampling.

lors d'une MD sans contrainte, O<sub>2</sub> continue à graviter autour de cette position. Ainsi, 5 MD d'une durée comprise entre 30 et 60 ns ont été réalisées. La molécule de O<sub>2</sub> est libre de tout mouvement et continue d'explorer le site actif de l'enzyme. O<sub>2</sub> ne reste jamais proche du site d'ancrage du substrat Int dans toutes les MD effectuées. Cependant, O<sub>2</sub> ne se retrouve jamais de l'autre côté du plan formé par le cycle thiazolone et le cycle adénosine du substrat Int. La position de ces deux plans ainsi que les résidus environnants occupent tout l'espace dans la cavité. Il est alors légitime de penser que l'attaque de O<sub>2</sub> ne peut se faire que d'un seul côté du cycle thiazolone du substrat Int, le côté par lequel la molécule de O<sub>2</sub> est arrivée. L'autre observation de ces MD est que la molécule de O<sub>2</sub> finit toujours par sortir du site actif puis de la luciférase pour se retrouver dans le solvant. Deux chemins sont observés quand à la sortie de O<sub>2</sub>. Cette molécule peut sortir par le même endroit que son entrée dans la protéine, c'est à dire par le canal entre le domaine N-terminal et le domaine C-terminal. La deuxième possibilité est de sortir par l'interface entre le domaine C-terminal, N-terminal et le solvant, la zone étant celle où la luciférine et l'ATP pénètrent dans le site actif quand le domaine C-terminal est en position ouvert.

Finalement, seule une MD avec une contrainte entre O<sub>2</sub> et le substrat Int permet de garder ces molécules proches l'une de l'autre. Suite à l'umbrella sampling, la minimisation de la structure  $D_{umbrella} = 3.5$  Å donne lieu au modèle 4G36-O<sub>2</sub>-min, le point

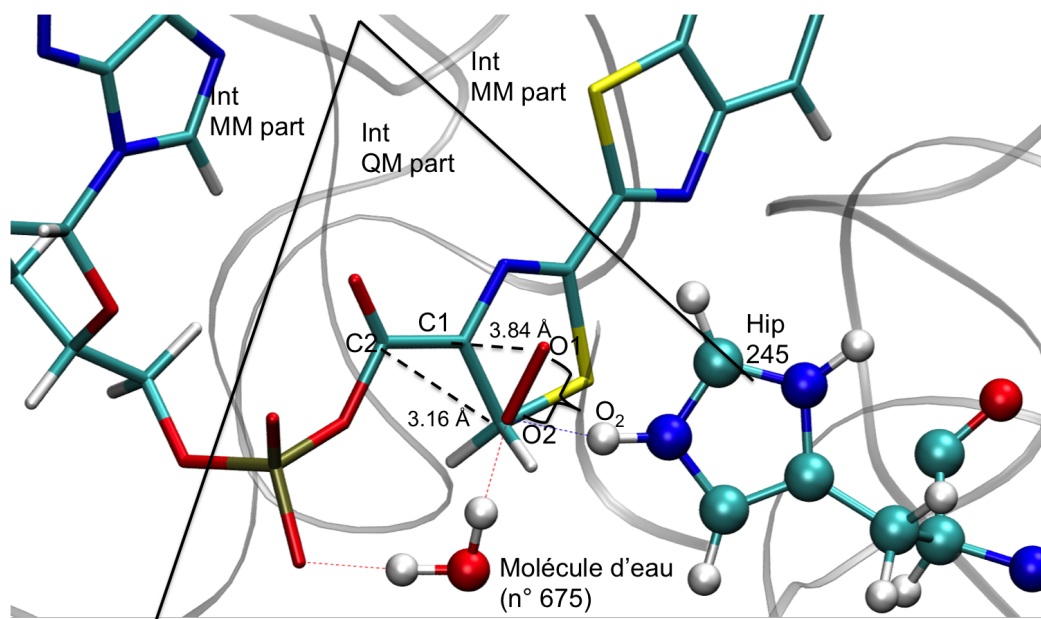


FIGURE 13.5 – Représentation du système après la simulation umbrella sampling et après la minimisation au niveau MM. Le substrat Int et la molécule de  $O_2$  sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par le modèle CPK.

de départ des calculs QM/MM réalisés par la suite. L'orientation de la molécule de  $O_2$  ne facilite pas la coordination au substrat Int. L'oxygène O2 est bien aligné vers le carbone C2 mais l'angle formé entre la liaison C2-C1 et l'autre atome d'oxygène O1 est de  $73^\circ$  au lieu de  $90^\circ$  attendu. En outre, l'atome d'oxygène O2 est plus proche du carbone C2, bien que la formation en premier de la liaison O1 et C1 semble plus logique. La distance C1-O1 est de  $3.77 \text{ \AA}$  et la distance C2-O2 de  $3.09 \text{ \AA}$ . En considérant l'environnement, l'oxygène O2 est stabilisé par le résidu Hip 245 et par une molécule d'eau.

### 13.3 La coordination de $O_2$

Dans cette partie on présente le calcul du chemin réactionnel de la coordination du  $O_2$  au substrat Int et à la formation du cycle dioxétanone. On utilise des calculs QM/MM avec une optimisation sous contrainte pour un ensemble de distances entre les deux molécules cibles.

#### 13.3.1 Calcul préliminaire : minimisation par DFT/MM

Dans un 1<sup>er</sup> temps, on réalise un calcul préliminaire DFT/MM en vue d'avoir un aperçu de la surface à obtenir. A partir de la structure 4G36- $O_2$ -min, une minimisation simple de l'énergie est calculée. L'utilisation de la méthode monodéterminantale DFT limite le nombre d'états électroniques considérés à un, l'état fondamental.

Cette première approche est effectuée pour une multiplicité de spin singulet d'une part et pour une multiplicité de spin triplet d'autre part. On utilise la fonctionnelle CAM-B3LYP ainsi que la base 6-31G(d,p) lors de ce calcul. La partie QM inclut le substrat Int ainsi que la molécule de O<sub>2</sub> (voir les détails de calculs 13.7). Lors d'une minimisation, on ne calcule pas le chemin réactionnel minimal permettant la formation du cycle dioxétanone. En revanche, cela permet de voir si l'état final est obtenu et de donner une idée sur la borne supérieure de la barrière énergétique de la réaction.

La minimisation de l'état S<sub>0</sub> singulet se déroule normalement et la formation du substrat Int-dioxétanone est atteinte en fin de calcul (voir Figure 13.6). Le chemin réactionnel suivi voit la formation de la liaison C1-O1 en premier lieu suivi de la liaison C2-O2. Finalement, le cycle dioxétanone est formé sans départ de l'AMP. On observe la présence de deux barrières énergétiques, une à 0.87 eV (20 kcal/mol) et une à 1.29 eV (30 kcal/mol). Cette énergie est du même ordre de grandeur que celle calculée en MS-CASPT2 lors de la décomposition du cycle dioxétanone, qui est de 25 kcal/mol<sup>18</sup>. En revanche, la minimisation de l'état fondamental triplet ne converge pas.

Cette approche préliminaire est déjà riche d'enseignement. La réaction de forma-

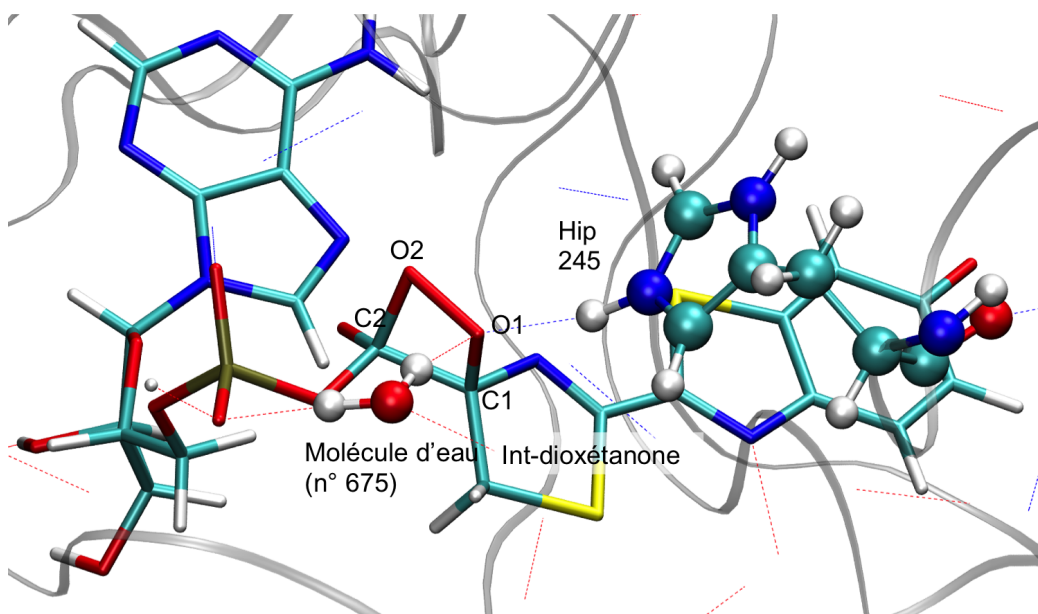


FIGURE 13.6 – Représentation du système après la simulation umbrella sampling et la minimisation en DFT/MM. Le substrat Int-dioxétanone et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK.

tion du dioxétanone est possible et accessible pour un coût énergétique peu élevé. Lors de cette réaction, la formation de la liaison C1-O1 est favorisée en premier lieu. Enfin, seule la réaction avec un état fondamental singulet est observée.

Dans la structure finale obtenue, la distance C1-C2 est de 1.56 Å, C1-O1 est de 1.44 Å, C2-O2 est de 1.56 Å et O1-O2 est de 1.47 Å. En outre, le dièdre impropre C4-N1-C2-C1 est de 23°, le dièdre impropre C1-O3-O4-C2 est de 30° et le dièdre

C1-O1-O2-C2 est de 20°.

Les résultats DFT montrent que le cycle dioxétanone se forme pour une distance moyenne de 1.5 Å entre les carbones C1 et C2 et les oxygènes O1 et O2. Afin de vérifier ce résultat et avoir une approche plus détaillée et précise on utilise des calculs MS-CASPT2/MM afin de modéliser l'approche entre O<sub>2</sub> et le substrat Int.

### 13.3.2 L'approche MS-CASPT2/MM

Dans cette partie, l'approche de O<sub>2</sub> est réalisée en suivant le chemin réactionnel de la formation du substrat Int-dioxétanone, c'est à dire le substrat Int possédant un groupement dioxétanone. On utilise pour cela une suite de calcul MS-CASPT2//SA-CASSCF/MM où on fait varier la contrainte appelée D-coord.

$$\frac{distance(C2 - O2) + distance(C1 - O1)}{2} = D - coord \quad (13.1)$$

Dans cette partie, la structure de départ correspond à la structure 4G36-O<sub>2</sub>-min, également nommée Int-O2 avec D-coord = 3.5 Å. L'ensemble des calculs MS-CASPT2//SA-CASSCF/MM effectués dans ce chapitre sont regroupés dans la Figure 13.7. De plus, afin de simplifier le nom des modèles utilisés, on abrège parfois le nom du modèle en le résumant à la valeur de la contrainte D-coord appliquée. Ainsi le modèle à 3.5 Å correspond au modèle Int-O2 avec D-coord = 3.5 Å.

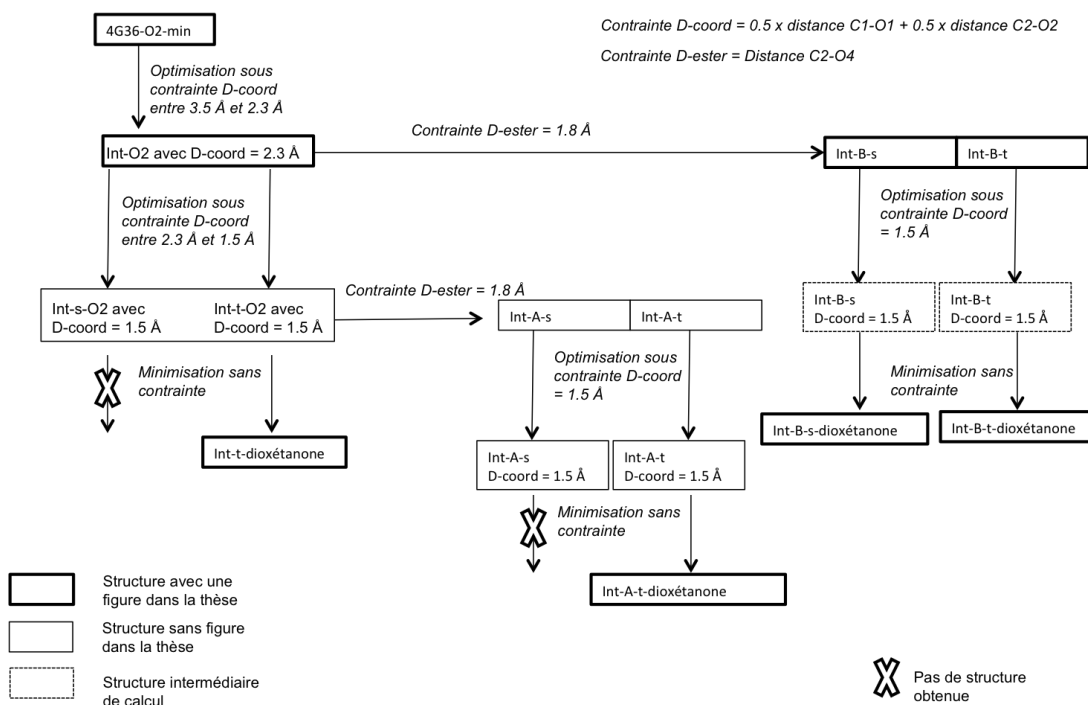


FIGURE 13.7 – Schéma représentant les différentes étapes et approches utilisées lors de l'approche de O<sub>2</sub> vers le substrat Int pour les calculs SA-CASSCF/MM.

### 13.3.2.1 Le choix de l'espace actif

L'approche MS-CASPT2/MM permet de traiter de façon plus précise cette réaction chimique. En outre, grâce au caractère multi-déterminantal de cette méthode, on peut prendre en compte l'impact de plusieurs états électroniques. Enfin, la méthode MS-CASPT2 traite également mieux les électrons non-appariés comme dans les états triplets.

Cependant, la méthode MS-CASPT2 est très lourde en terme de temps de calcul et il est impossible de mettre tous les atomes du substrat Int dans la partie QM. On décide d'inclure une partie du substrat Int dans le sous-système MM. L'algorithme de l'atome de liaison, cf chapitre 3.2.4, est utilisé sur deux liaisons C-C. Au niveau de la luciférine, on place une coupure entre les deux cycles, la partie benzothiazole est alors incluse dans la partie MM. Pour l'AMP, on coupe au niveau du ribose, le groupement méthyl  $\text{CH}_3$  aliphatique est inclus avec le groupe phosphate dans la partie QM alors que le cycle du ribose et l'adénosine sont dans la partie MM. Enfin la molécule de  $\text{O}_2$  appartient à la partie QM (voir Figure 13.1). La structure de départ dans cette étude est celle minimisée à la fin de l'umbrella sampling, 4G36- $\text{O}_2$ -min (voir Figure 13.5).

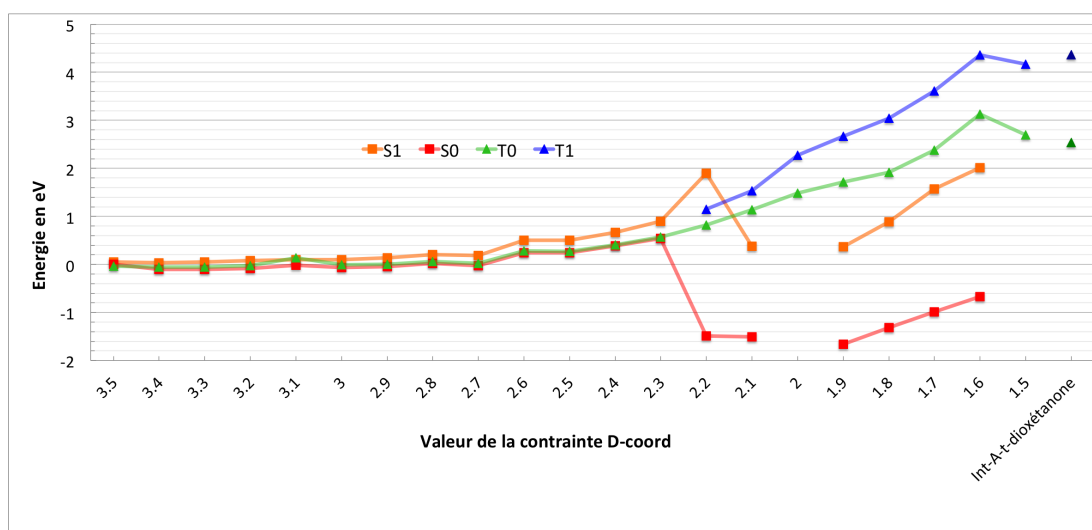


FIGURE 13.8 – Graphique représentant l'énergie du système en eV en fonction de la valeur de la distance contrainte. Le point à zéro eV correspond à l'énergie de la structure pour D-coord valant 3.5 Å et nommée 4G36- $\text{O}_2$ -min. Le point nommé Int-A-t-dioxétanone correspond au point obtenu lors de la formation du cycle dioxétanone après l'application de la contrainte D-ester sur la structure Int-t- $\text{O}_2$  avec D-coord = 1.5 Å. En rouge correspond au chemin réactionnel de l'état S0, en orange S1, en vert l'état T0 et en bleu l'état T1.

Il est important de bien choisir l'espace actif. Ce dernier ne doit pas être trop grand car le nombre de calculs à réaliser lors de cette étude est important. Il ne doit également pas être trop petit afin de bien prendre en compte toutes les orbitales



participant à la réaction. On cherche dans un premier temps à inclure toutes les orbitales  $\pi$  du substrat Int soit 7 électrons dans 6 orbitales. On choisit ensuite un espace actif minimal sur  $O_2$  avec 3 électrons et 2 orbitales. L'espace actif résultant est de 10 électrons dans 8 orbitales (10-in-8). Cet espace semble assez petit mais possède toutes les orbitales importantes. Sur  $O_2$ , les deux orbitales choisies sont celles pointées vers le substrat Int. Le choix de l'espace actif montre déjà que la répartition des électrons dans la partie QM est différente de celle attendue. Le dioxygène possède 3 électrons et le substrat Int 7 électrons dans l'espace actif soit la présence de radicaux libres. En détails, le carbone C1, qui a été déprotoné est censé porter dans un schéma de Lewis une charge -1, soit 2 électrons. Or, on note que cet atome ne porte plus qu'1 électron dans l'orbitale de l'espace actif et les deux oxygènes de la molécule de  $O_2$  portent chacun en moyenne 1,5 électrons (une orbitale à 1,67 et l'autre à 1,33 électrons) (voir Annexes A6). Cet espace actif ne reproduit pas une attaque de  $O_2$  sur un carbanion mais une réaction radicalaire (voir Annexe A6).

### 13.3.2.2 Le chemin réactionnel entre 3.5 Å et 2.3 Å.

On modélise dans un premier temps l'approche de  $O_2$  pour D-coord entre 3.5 Å et 2.7 Å. On observe peu de changement concernant la position des états électroniques lors de ces pas (voir Figure 13.8). De plus, les énergies pour les états S0, S1, T0 et T1 sont comparables et restent quasiment dégénérées pour ces valeurs de la contrainte D-coord. Concernant la géométrie du système, la distance entre les carbones et les oxygènes diminue progressivement et de façon concertée. A 2.7 Å la distance C1-O1 est de 2.85 Å et la distance C2-O2 de 2.50 Å. La molécule de  $O_2$  est toujours parallèle au plan formé par le cycle thiazolone du substrat Int, et encore stabilisée par le résidu Hip 245 et la même molécule d'eau n°675 qui intervient depuis le début de la simulation.

A partir de D-coord = 2.6 Å jusqu'à 2.3 Å l'énergie calculée augmente légèrement par rapport à l'énergie calculée pour la structure D-coord = 3.5 Å. A 2.3 Å l'énergie calculée pour S0 et T0 est de +0.54 eV soit 12,4 kcal/mol. En outre, la séparation des états S0 et S1 est désormais perceptible et vaut environ 0.36 eV à 2.3 Å. Bien que cette distance corresponde à une augmentation de l'énergie cela ne se ressent pas sur la géométrie de la structure. La distance C1-O1 est de 2.40 Å et la distance C2-O2 de 2.20 Å,  $O_2$  est dans la même orientation et toujours stabilisé par Hip 245 et la molécule d'eau n°675 (voir Figure 13.9).

Si on regarde les orbitales de l'espace actif, on voit le couplage entre certaines orbitales du substrat Int et du dioxygène. Ainsi les orbitales centrées sur C2-O3 interagissent avec celles sur O2, l'orbitale centrée sur N1 et celle centrée sur C1 interagissent avec celle de O1. En retour, l'orbitale  $\pi^*$  du dioxygène interagit avec celles centrées sur C3 et N1. En visualisant ces orbitales, on voit qu'il est possible d'avoir la coordination du dioxygène aussi bien avec C1 que C2 (voir Annexes A7). Cette distance marque la séparation des chemins réactionnel entre les états singulets et les

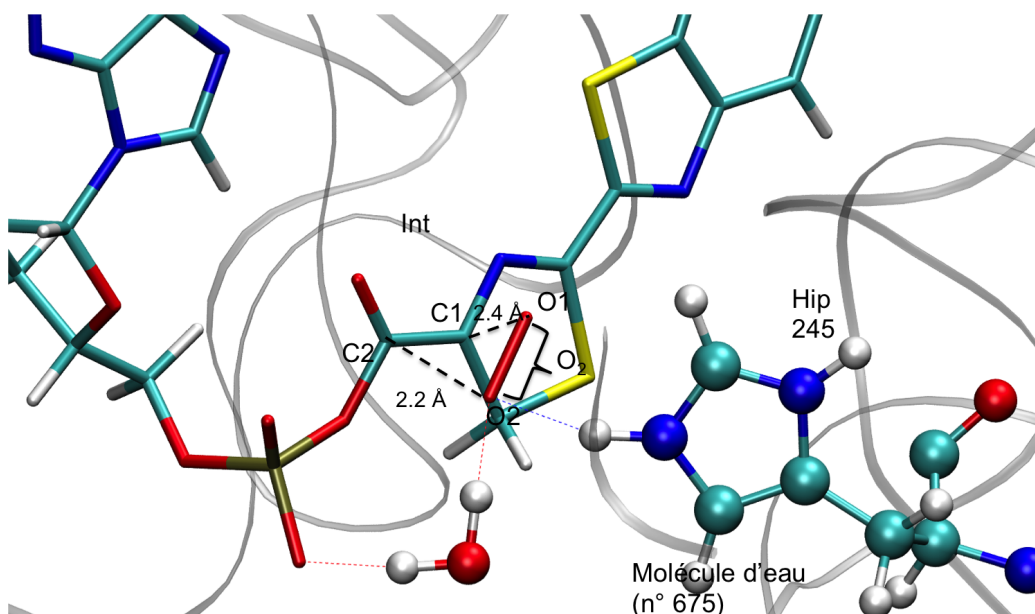


FIGURE 13.9 – Représentation du système après la simulation umbrella sampling et pour le modèle Int-O2 avec D-coord valant 2.3 Å. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord.

états triplets et le début de la coordination de la molécule de O<sub>2</sub>. La distance C1-O1 est de 2.40 Å et la distance C2-O2 de 2.20 Å.

### 13.3.2.3 Le chemin réactionnel entre 2.3 Å et 1.5 Å.

#### Les états singulets

On s'intéresse dans un premier temps aux états singulets. Pour ces derniers la structure calculée pour 2.3 Å correspond à une barrière énergétique. En effet, à D-coord = 2.2 Å l'énergie décroît de façon importante et vaut alors -1.48 eV par rapport à l'énergie de départ. Il est cependant difficile de parler de TS pour la structure à 2.3 Å. Des calculs d'optimisation de TS ont été réalisés mais il est extrêmement difficile d'obtenir la valeur précise de l'énergie de l'état de transition avec le logiciel Molcas. La présence de la protéine autour du substrat complique encore plus les calculs et la minimisation. Cependant la valeur de la barrière calculée correspondant à l'énergie de la structure à 2.3 Å par rapport à la structure à 3.5 Å est de +0.54 eV et du même ordre de grandeur que pour la minimisation en DFT/MM.

La structure à 2.2 Å est très différente du point précédent (voir Figure 13.10). La distance C1-O1 est de 1.50 Å et la distance C2-O2 de 2.90 Å. La diminution brutale de la distance C1-O1 s'explique par le déplacement de l'atome O1 mais aussi par le mouvement de l'atome C1. Cela brise la planéité du substrat Int, notamment du cycle thiazolone. L'angle dièdre impropre entre C4-N1-C2-C1 passe de 2° dans 2.3 Å à 30° dans 2.2 Å. Seul l'atome C1 est affecté par ce mouvement vers le dioxygène.

En outre, l'oxygène O2 est repoussé et perd également son alignement avec C2. L'oxygène O2 conserve néanmoins ses liaisons hydrogène avec Hip 245 et la molécule d'eau n° 675. Au niveau des orbitales de l'espace actif, elles possèdent toutes

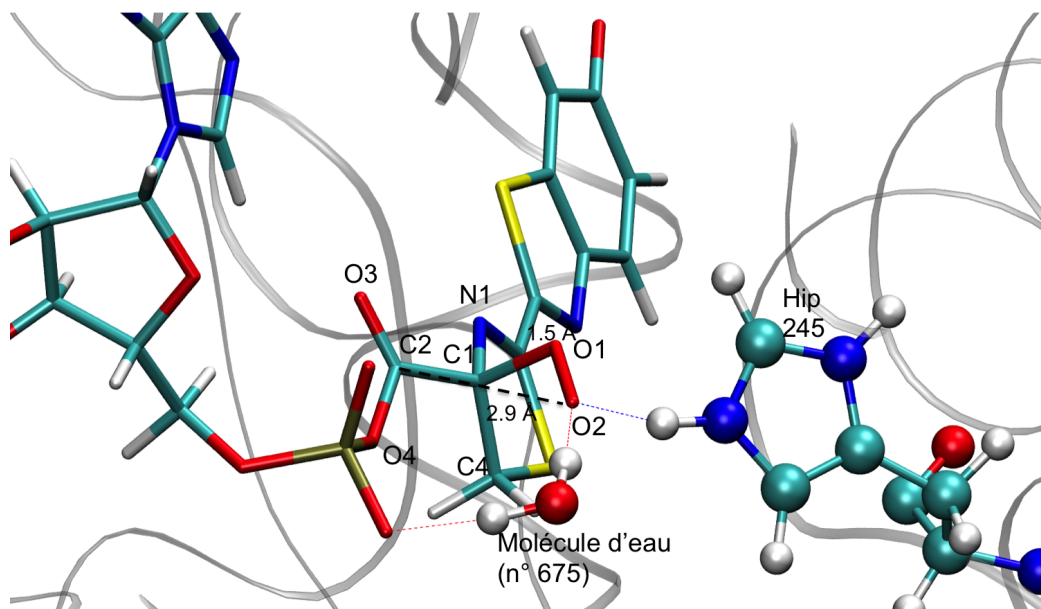


FIGURE 13.10 – Représentation du système après la simulation umbrella sampling et pour Int-s-O2 avec D-coord valant 2.2 Å pour les états singulets. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord.

de la densité électronique répartie entre le substrat Int et O<sub>2</sub>. L'orbitale centrée sur le soufre disparaît mais celle-ci n'était pas importante et n'interagissait jamais avec les autres orbitales. A la place on a l'apparition de plusieurs orbitales représentant les liaisons  $\sigma$  et  $\sigma^*$  entre C1 et O1. Les orbitales montrant ces liaisons possèdent également un caractère d'orbitale  $p_z$  délocalisée. De plus, les deux orbitales  $\pi$  issues du dioxygène interagissent avec le cycle thiazolone mais aussi avec la paire libre de l'azote N1 (voir Annexes A8).

Lors de la suite du chemin réactionnel les énergies du S0 et du S1 recommencent à augmenter entre 1.9 Å et 1.6 Å mais l'énergie de l'état fondamental est toujours plus basse que dans la structure de départ. L'écart entre S0 et S1 diminue et évolue entre 2 eV et 2.6 eV entre 1.9 Å et 1.6 Å. Structuellement, on observe un déplacement concerté de O2 et de C2 l'un vers l'autre. Pour la structure à 1.6 Å la distance C1-O1 est de 1.34 Å et la distance C2-O2 de 1.86 Å et le dièdre C4-N1-C2-C1 de 27° (voir Figure 13.11).

Finalement la distance de 1.5 Å est censée correspondre à la distance où l'on observe la formation du cycle dioxétanone. Malheureusement, le calcul de minimisation n'aboutit pas pour le système dans l'état singulet. La structure observée montre la formation d'une structure stable avec une liaison covalente entre O2 et l'hydrogène

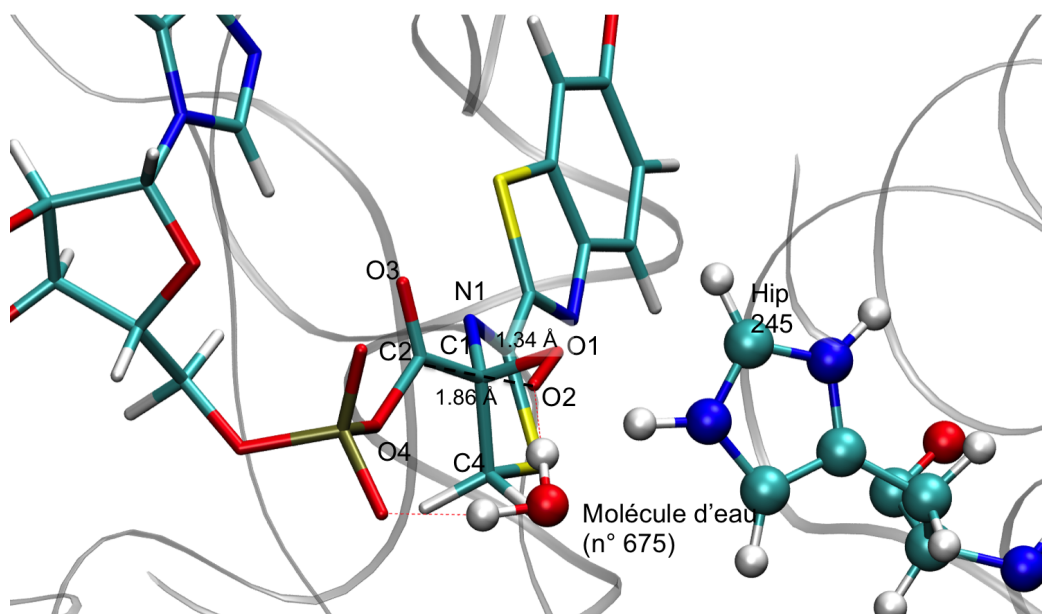


FIGURE 13.11 – Représentation du système après la simulation umbrella sampling et pour Int-s-O2 avec D-coord valant 1.6 Å pour les états singulets. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord.

de la molécule d'eau 675. Il est compliqué d'expliquer pourquoi cette structure se forme contrairement au cycle dioxétanone qui semble pourtant être une structure finale cohérente au vue de l'approche réalisée.

### Les états triplets

L'étude des états triplets montre qu'entre 2.2 Å et 1.6 Å l'énergie de T0 et T1 augmente atteignant un maximum pour une contrainte D-coord valant 1.6 Å de +3.12 eV soit 71 kcal/mol au dessus de l'énergie de la structure 4G36-O<sub>2</sub>-min (voir Figure 13.8). Entre 2.2 Å et 1.6 Å l'écart entre T0 et T1 est compris entre 0.2 eV et 1.2 eV. A une distance D-coord de 1.8 Å, on observe l'apparition d'une liaison entre C2 et O2 (voir figure 13.12). A cette étape, la distance C1-O1 est de 2.17 Å et la distance C2-O2 de 1.43 Å. Comme pour les états singulets, la formation de cette liaison montre la fin de la planéité de l'intermédiaire de réaction. Par contre, alors que pour l'état singulet c'est C1 qui prenait un caractère sp<sup>3</sup>, ici c'est C2 qui perd la planéité vis à vis du cycle thiazolone du substrat, le dièdre impropre C1-O3-O4-C2 est de 29° contre 8° pour la structure à D-coord = 2.2 Å. Si on regarde les orbitales de cette structure, on voit que c'est l'orbitale centrée sur l'oxygène O2 qui permet d'avoir le caractère liaison  $\sigma$  entre C2 et O2. On a peu de changements pour les 7 autres orbitales, on note également que beaucoup d'entre elles sont délocalisées entre le cycle thiazolone et O<sub>2</sub> (voir Annexes A9).

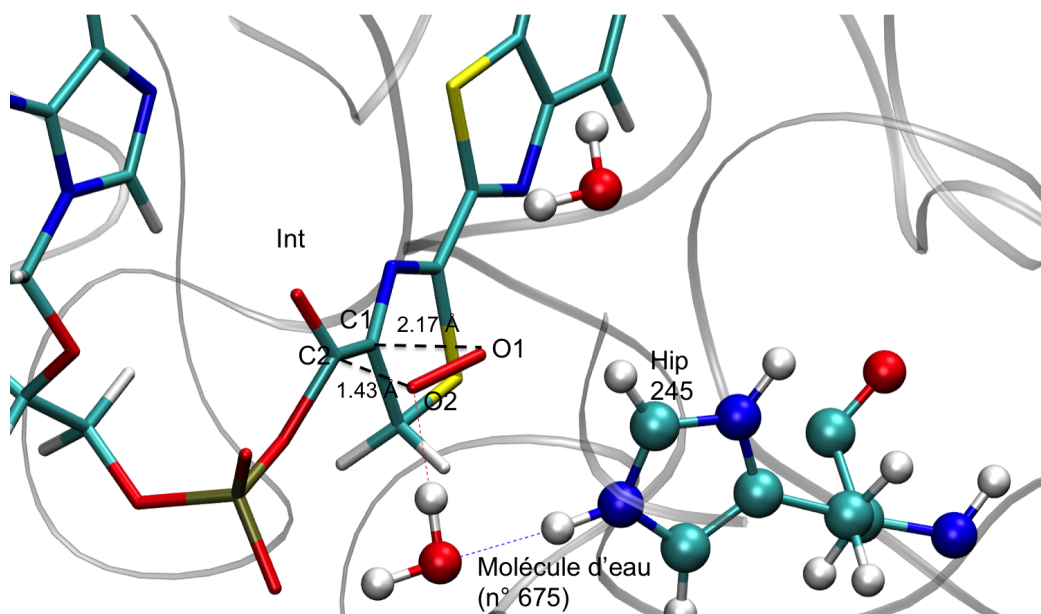


FIGURE 13.12 – Représentation du système après la MD umbrella sampling et pour Int-t-O2 avec D-coord valant 1.8 Å pour les états triplets. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord.

Le point à 1.5 Å montre la formation du cycle dioxétanone avec C1-O1 égal à 1.40 Å, C2-O2 égal à 1.60 Å, O1-O2 égal à 1.41 Å et C1-C2 à 1.55 Å. Le dièdre impropre C4-N1-C2-C1 est de 30°, le dièdre impropre C1-O3-O4-C2 est de 21° et le dièdre C1-O1-O2-C2 est de 25°, voir figur 13.13. Cette structure est nommée Int-t-dioxétanone-1.5. Ces résultats sont proches des valeurs DFT/MM nous confortant dans la conformation de la structure finale. Au niveau de l'environnement l'atome O2 possède toujours une liaison hydrogène avec la molécule d'eau n°675. Cette dernière molécule forme aussi une liaison hydrogène avec le résidu Hip 245 et une avec un oxygène du groupement phosphate du substrat Int.

L'énergie de la structure Int-t-O2 avec D-coord = 1.5 Å affiche une baisse de l'énergie vis à vis de la structure à D-coord = 1.6 Å, qui correspond au maximum énergétique, mais cette baisse est de seulement 0.42 eV. L'énergie de Int-t-O2 avec D-coord = 1.5 Å plafonne à +2.7 eV par rapport à la structure de départ soit 62 kcal/mol. Si on considère l'énergie de la barrière permettant de rejoindre la structure 4G36-O<sub>2</sub>-min à cette structure il faut apporter au système une énergie de 71 kcal/mol. Cette énergie est très élevée et pour l'instant il est difficile de penser quel est le processus permettant de générer cette énergie.

Une autre solution étudiée afin de diminuer l'énergie des structures possédant un cycle dioxétanone consiste en un calcul de minimisation SA-CASSCF/MM en enlevant la contrainte D-coord. Ce calcul est réalisé sur le point Int-t-O2 avec D-coord = 1.5 Å. On observe la cassure du cycle dioxétanone entre les atomes C2 et

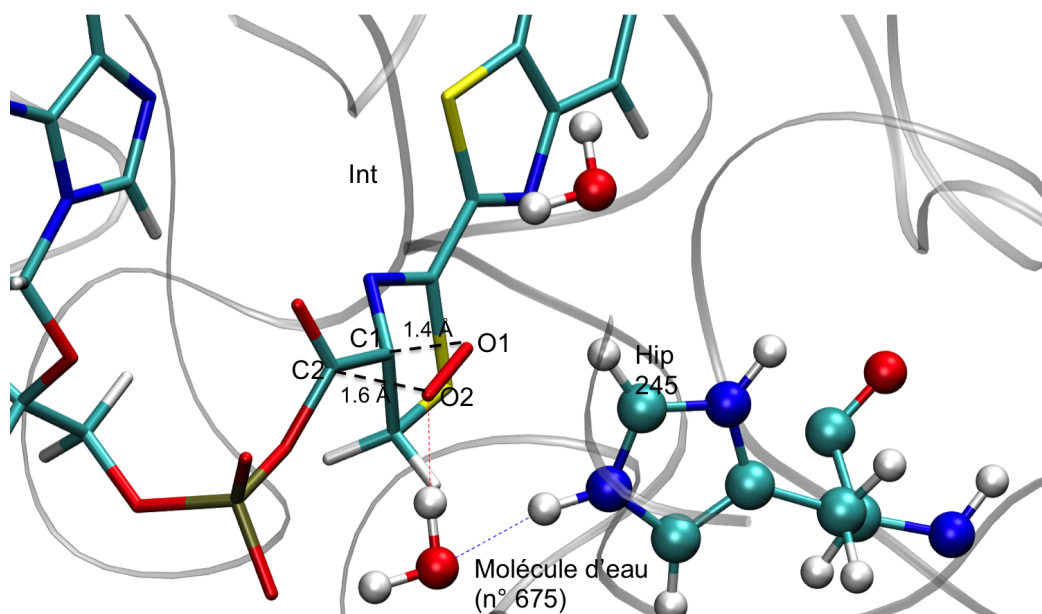


FIGURE 13.13 – Représentation du système après la simulation umbrella sampling et pour Int-t-O2 avec D-coord valant 1.5 Å pour les états triplets. Le substrat Int-dioxétanone et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS.

O<sub>2</sub> alors que la liaison C1-O1 est maintenue.

Au niveau des orbitales de l'espace actif, on perd l'orbitale centrée sur le soufre. Fort heureusement, on observe la présence des orbitales  $\sigma$  et  $\sigma^*$  pour les deux liaisons formées. Les deux orbitales issues du dioxygène interagissent avec la paire libre de l'azote N1 (voir Annexes A10).

#### 13.3.2.4 Retour sur les états singulets

La minimisation dans l'état singulet de la structure Int-s-O<sub>2</sub> avec D-coord = 1,5 Å n'a pas aboutie à la formation de la fonction dioxétanone (structure nommée Int-s-dioxétanone). On propose d'utiliser les résultats obtenus à l'état triplet afin d'estimer l'énergie d'une structure dioxétanone singulet dans l'état fondamental. Pour cela, un calcul *single point* MS-CASPT2/MM sur le modèle Int-t-O<sub>2</sub> avec D-coord = 1.5 est effectué avec un état fondamental singulet. On obtient une énergie de +0.46 eV (11 kcal/mol) par rapport à la structure 4G36-O<sub>2</sub>-min. L'énergie calculée pour un état singulet est plus basse que pour un état triplet mais reste néanmoins supérieur à l'énergie du modèle de départ.

Les premiers calculs représentant l'approche de O<sub>2</sub> vers le substrat Int montrent qu'il est difficile de former le substrat Int avec le cycle dioxétanone, comme dans le cas des états singulets. Pour les états triplet, le groupement dioxétanone est formé, mais l'énergie nécessaire pour le former est d'environ 70 kcal/mol. Une autre ap-

proche est alors étudiée toujours dans le but d'obtenir le cycle dioxétanone, le pré-curseur de la bioluminescence.

### 13.3.2.5 Chemin réactionnel après le départ de l'AMP

La partie précédente a montré qu'il est compliqué de former le substrat Int-dioxétanone. L'approche suivie auparavant bloque surtout quand les deux oxygènes se lient au carbones C1 et C2. En particulier, le carbone C2 doit, lors de la réaction de coordination du O<sub>2</sub>, passer d'une hybridation sp<sup>2</sup> à une hybridation sp<sup>3</sup>. Or, dans les orbitales de l'espace actif obtenue à la distance D-coord valant 1.5 Å on continue à voir le caractère sp<sup>2</sup> de ce carbone, via la présence d'orbitales π entre C2 et O3. On peut réfléchir à une autre approche possible de la formation du substrat Int-dioxétanone tout en conservant l'hybridation sp<sup>2</sup> du carbone C2. Une solution possible consiste à casser la liaison C2-O4, correspondant à la liaison ester entre la luciférine et l'AMP.

L'espace actif choisi lors de cette étude ne comporte pas la liaison sigma entre C2 et O4. La cassure de cette liaison n'était donc pas autorisée lors de l'approche de O<sub>2</sub> réalisée dans les paragraphes précédents. Afin de suivre ce nouveau chemin réactionnel, on suit le protocole suivant : On part d'une structure optimisée avec la contrainte D-coord valant 1.5 Å ou 2.3 Å. Sur cette structure, une nouvelle contrainte est définie, appelée D-ester et correspondant à une distance de 1.8 Å entre le carbone de la luciférine C2 et l'oxygène de l'AMP O4. Un calcul d'optimisation SA-CASSCF/MM est ensuite réalisé avec cette nouvelle contrainte (il n'y a alors plus de contrainte sur D-coord). Lors de cette optimisation, la coupure de la liaison C2-O4 se fait de façon hétérolitique, toute la densité électronique se retrouvant sur l'oxygène. Ainsi, dans les orbitales après cassure, on voit que chaque oxygène du groupement phosphate possède une charge négative.

Puis, sur la structure obtenue, on réutilise une optimisation avec comme seule contrainte la contrainte D-coord valant cette fois-ci 1.5 Å afin d'approcher O<sub>2</sub>. Enfin, suite à cette optimisation, on enlève toute contrainte et on réoptimise afin de relaxer le système. On vérifie à chaque pas que la liaison entre la partie luciférine et AMP, C2-O4 ne se reforme pas. Un récapitulatif des différentes étapes est présentée dans la figure 13.7. Ce protocole est réalisé à partir de 4 structures : Int-s-O2 avec D-coord = 1.5 Å, Int-t-O2 avec D-coord = 1.5 Å, ainsi que Int-O2 avec D-coord = 2.3 Å, obtenue à l'état singulet et à l'état triplet. Nous discutons de ces deux approches des les paragraphes suivants.

#### Quand le départ de l'AMP est réalisé à une contrainte D-coord valant 1.5 Å.

Dans un premier temps on réalise le départ de l'AMP quand la D-coord vaut 1.5 Å. Cela équivaut à casser la liaison entre la luciférine et l'AMP quand la coordination du dioxygène au substrat Int a déjà eu lieu (au moins en partie).

La structure de départ est différente dans le cas du singulet ou du triplet, on traite donc le chemin réactionnel séparément selon le spin total du système. Suite au calcul d'optimisation avec la contrainte D-ester valant 1.8 Å entre C2 et O4, les structures obtenues sont notées Int-A-s (dans le cas du singulet) et Int-A-t (dans le cas du triplet); elles montrent bien la cassure de liaison entre la luciférine et l'AMP. En outre, un calcul *single point* MS-CASPT2/MM montre que l'énergie de ces deux nouvelles structures est identique à l'énergie des structures de départ respectives.

Dans le cas du singulet, on réintroduit ensuite la contrainte D-coord valant 1.5 Å. Une nouvelle fois le cycle dioxétanone n'est pas obtenu. A la place, on observe la cassure de la liaison C1-C2 alors que C1 est toujours lié à la molécule de O<sub>2</sub>. Le calcul de minimisation sans contrainte réalisée sur la structure précédente ne converge pas, nous n'avons donc pas accès à son énergie finale.

Pour le cas du triplet, l'optimisation avec la contrainte D-coord valant 1.5 Å, puis le

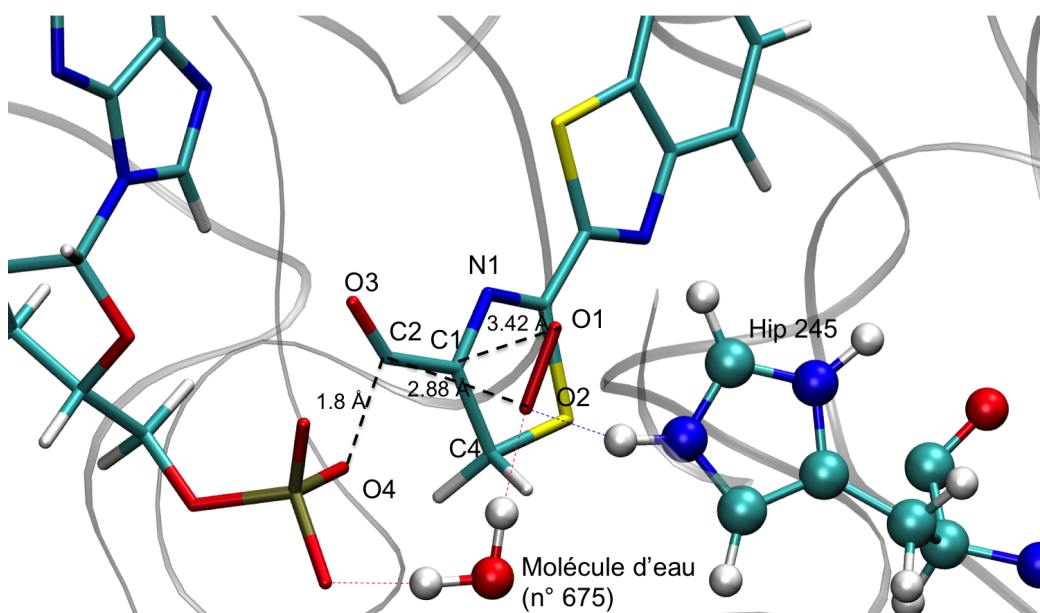


FIGURE 13.14 – Représentation du système après la simulation umbrella sampling, pour la structure Int-B-s, la structure obtenue après une contrainte D-ester entre C2 et O4 de 1.8 Å sur la structure Int-O2 avec D-coord = 2.3 Å et pour les états singulets. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4.

calcul de minimisation sans contrainte montre la formation d'un cycle dioxétanone, et donne le modèle nommée Int-A-t-dioxétanone. Dans ce modèle, C1-O1 est égal à 1.46 Å, C2-O2 vaut 1.38 Å, O1-O2 est égal à 1.54 Å, C1-C2 à 1.50 Å et C2-O4 est de 2.72 Å. Le dièdre impropre C4-N1-C2-C1 est de 25°, et le dièdre C1-O1-O2-C2 est de 5°. L'énergie MS-CASPT2/MM de ce point est de +2.53 eV par rapport à la structure de départ avec la contrainte D-coord valant 3.5 Å. Cette énergie est cependant inférieure à celle obtenue pour la structure Int-t-O2 avec D-coord = 1.5 Å, discutée dans le point



précédent.

Le départ de l'AMP réalisée à une contrainte D-coord valant 1.5 Å donne lieu à des résultats en demi-teinte. Dans le cas des états singulets on n'observe pas la formation du cycle dioxétanone alors que dans le cas des états triplets le cycle dioxétanone est obtenu avec une barrière énergétique élevée. En outre, cette approche en faisant partir l'AMP ne permet pas de stabiliser la structure avec le cycle dioxétanone.

### Quand le départ de l'AMP est réalisée à une contrainte D-coord valant 2.3 Å.

Dans la deuxième approche on réalise le départ de l'AMP quand la D-coord vaut 2.3 Å. Cela équivaut à casser la liaison entre la luciférine et l'AMP avant la coordination du dioxygène au substrat Int.

La structure obtenue pour une contrainte D-coord = 2.3 Å est identique dans le cas de l'état singulet et de l'état triplet. On part donc de cette même structure mais on traite toujours le chemin réactionnel séparément selon le spin total du système. Suite au calcul d'optimisation avec la contrainte D-ester valant 1.8 Å entre C2 et O4, les structures obtenues sont notées Int-B-s (dans le cas du singulet) et Int-B-t (dans le cas du triplet). Ces structures montrent bien la cassure de liaison entre la luciférine et l'AMP.

Pour un état fondamental singulet la structure Int-B-s montre une augmentation de la distance entre O<sub>2</sub> et le substrat Int. On passe d'une distance D-coord de 2.3 Å à une distance de 3.15 Å avec la distance C1-O1 égale à 3.42 Å et C2-O2 égale à 2.88 Å (voir Figure 13.14). Finalement, l'énergie calculée pour cette structure est légèrement supérieure à la structure de départ contenant l'AMP, avec +0.71 eV contre +0.54 eV pour la structure avec D-coord valant 2.3 Å (voir Figure 13.15).

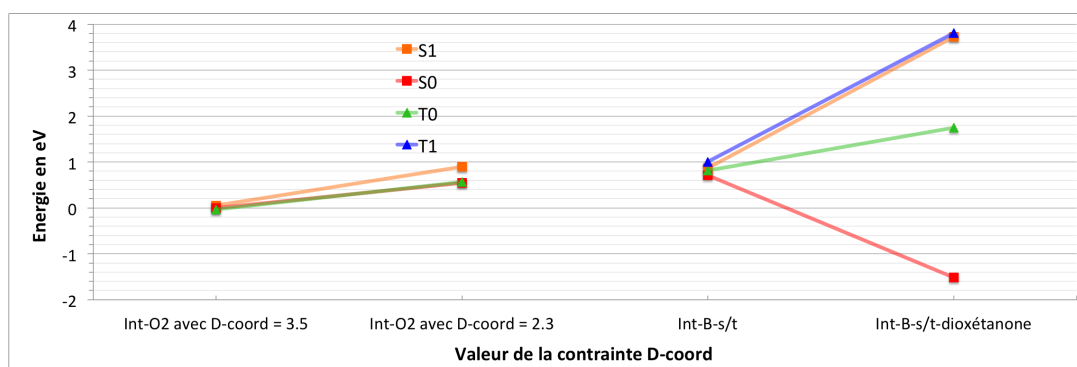


FIGURE 13.15 – Graphique représentant l'énergie du système en eV pour 4 points lors de l'approche de O<sub>2</sub>. Le point à zéro eV correspond à l'énergie de la structure pour D-coord valant 3.5 Å. Le point nommé Int-B-s/t correspond au point obtenu après l'application de la contrainte D-ester sur la structure Int-O2 avec D-coord = 2.3 Å. Le point nommé Int-B-s/t-dioxétanone correspond au point obtenu après l'application de la contrainte D-coord = 1.5 Å suivi d'une minimisation sur la structure Int-B-s/t. En rouge correspond l'état S0, en orange S1, en vert l'état T0 et en bleu l'état T1.

Par la suite la contrainte D-coord, valant 1.5 Å, est appliquée sur la structure Int-B-s et une fois le dioxygène proche du substrat Int, on récupère la structure obtenue pour le dernier pas et on la minimise. On obtient alors la structure Int-B-s-dioxétanone (voir figure 13.16). Dans cette dernière la distance C1-O1 est égale à 1.49 Å, C2-O2 égal à 1.36 Å, O1-O2 égal à 1.40 Å, C1-C2 à 1.50 Å et C2-O4 est de 3.06 Å. Le dièdre impropre C4-N1-C2-C1 est de 22° et le dièdre C1-O1-O2-C2 est de -3°. L'atome O2 ne possède plus de liaisons hydrogène. La molécule d'eau n°675 forme un réseau de liaison hydrogène avec le résidu Hip 245 et un oxygène du groupement phosphate de l'AMP.

Au niveau énergétique, l'énergie diminue fortement et passe à -1,51 eV par rapport

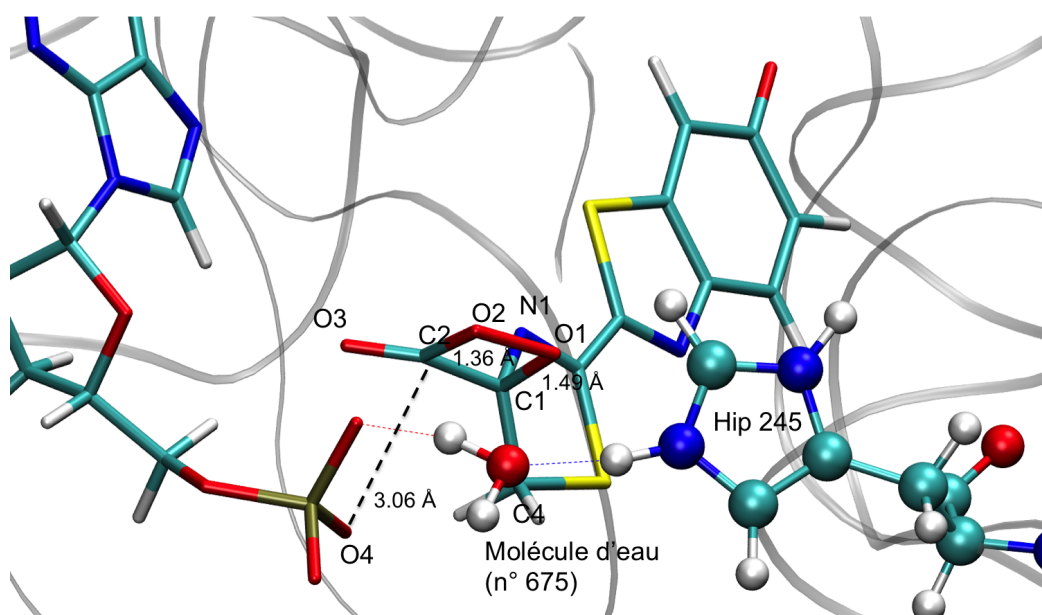


FIGURE 13.16 – Représentation du système après la simulation umbrella sampling, pour la structure Int-B-s-dioxétanone, après une contrainte D-coord valant 1.5 Å suivi d'une minimisation sur la structure Int-B-s pour les états singulets. Le substrat Int-dioxétanone et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4.

à la structure 4G36-O<sub>2</sub>-min soit -34 kcal/mol. L'écart S1-S0 s'accroît beaucoup et passe à 5.31 eV (voir Figure 13.15).

Au niveau orbitalaire, on note la perte de l'orbitale centrée sur le soufre dans l'espace actif. A la place on a l'apparition de plusieurs orbitales représentant les liaisons  $\sigma$  et  $\sigma^*$  entre C1 et O1. Les orbitales montrant ces liaisons possèdent également un caractère d'orbitale  $\pi$  délocalisée. En outre, les deux orbitales  $\pi$  issues du dioxygène interagissent avec le cycle thiazolone et avec la paire libre de l'azote N1. On n'observe pas par contre les orbitales  $\sigma$  et  $\sigma^*$  de la liaison C2-O2. Il existe cependant une orbitale  $\pi$  entre C2 et O2 présente dans l'espace actif (voir Annexes A11).

On a finalement réussi à obtenir une structure contenant un cycle dioxétanone pour des états singulets. De plus, l'énergie calculée est plus basse que celle de la structure de départ, montrant que cette réaction est thermodynamiquement faisable.

Dans le cas de l'état fondamental triplet, la contrainte de 1.8 Å sur C2-O4 éloigne également O<sub>2</sub> du substrat. Dans la structure Int-B-t (voir Figure 13.17), on retrouve la stabilisation de O<sub>2</sub> par le résidu Hip 245 et une molécule d'eau. La distance C1-O1 est égale à 3.39 Å et C2-O2 vaut 2.85 Å. L'énergie de cette structure est légèrement supérieure à celle avec D-coord égale à 2.3 Å, avec +0.82 eV contre +0.54 eV.

Par la suite la contrainte D-coord égale à 1.5 Å est directement appliquée et une

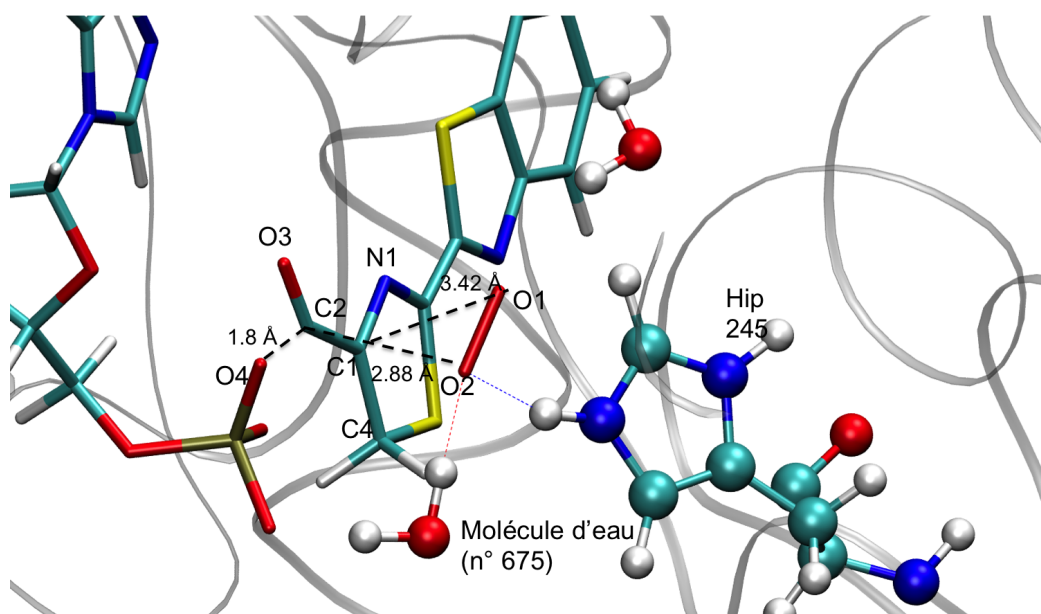


FIGURE 13.17 – Représentation du système après la simulation umbrella sampling, pour la structure Int-B-t, la structure obtenue après une contrainte D-ester entre C2 et O4 de 1.8 Å sur la structure Int-O2 avec D-coord = 2.3 Å et pour les états triplets. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4.

fois le dioxygène proche du substrat Int, la dernière image du calcul est extraite et minimisée sans contrainte. On obtient alors la structure Int-B-t-dioxétanone (voir Figure 13.18 et Annexes A12). Dans cette dernière la distance C1-O1 est égale à 1.50 Å, C2-O2 égal à 1.37 Å, O1-O2 égal à 1.40 Å, C1-C2 à 1.50 Å et C2-O4 est de 5.07 Å. Le dièdre impropre C4-N1-C2-C1 est de 24° et le dièdre C1-O1-O2-C2 est de -1°. La distance entre Int-B-t-dioxétanone et l'AMP est très grande car le groupement phosphate subit une rotation de 60° et O4 ne pointe plus en direction du substrat. En outre, O2 possède toujours une liaison hydrogène mais avec une autre molécule d'eau n°675. Un réseau de liaisons hydrogène est formé entre Hip 245, la molécule d'eau n°675 et l'oxygène de l'AMP qui appartient à la liaison PO dirigée vers C2.

Au niveau énergétique l'énergie augmente quand on passe de Int-B-t à Int-B-t-

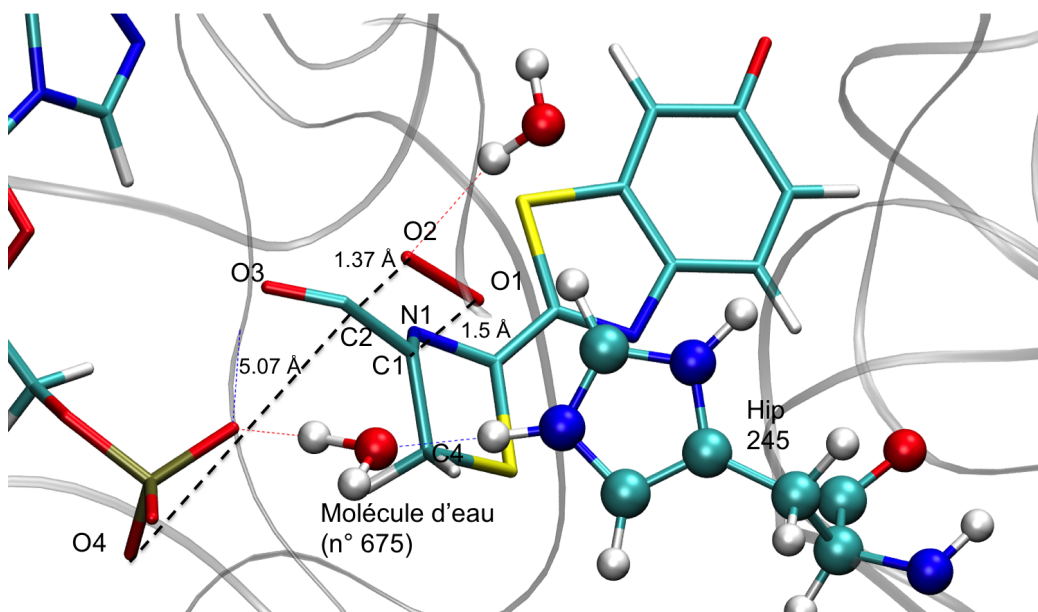


FIGURE 13.18 – Représentation du système après la simulation umbrella sampling, pour la structure Int-B-t-dioxétanone, après une contrainte D-coord valant 1.5 Å suivi d'une minimisation sur la structure Int-B-t pour les états triplets. Le substrat Int-dioxétanone et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et les molécules d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4.

dioxétanone et atteint la valeur +1.75 eV par rapport à l'énergie de 4G36-O<sub>2</sub>-min (voir Figure 13.15).

Pour les 3 différents substrats Int-dioxétanone obtenus dans un état fondamental triplet, l'énergie trouvée est toujours supérieure à celle de départ. Il est alors raisonnable de penser qu'il est difficile de former le dioxétanone pour un mécanisme comportant des états triplets.

## 13.4 Discussion

L'étude de la coordination du dioxygène a montré de nombreux points plus ou moins ambigus que nous allons discuter maintenant.

### 13.4.1 L'espace actif en CASSCF

Le premier point concerne le choix des orbitales dans l'espace actif pour les calculs MS-CASPT2//SA-CASSCF/MM effectués pour le substrat Int et la molécule de dioxygène. Le choix d'un espace actif de 10 électrons dans 8 orbitales semble

raisonnable en temps de calcul. De façon pratique il est pour l'instant impossible d'utiliser un espace actif plus grand pour l'optimisation de notre système. Pour rappel, l'espace actif comprend le système conjugué du cycle thiazolone, c'est à dire les orbitales  $\pi$  des atomes C3, N1, C1, C2, O3 et la paire libre du soufre de S1. Sur le dioxygène les deux  $\pi^*$  sont sélectionnées. Le détail des orbitales se trouve en Annexes, de A6 à A13. Les orbitales sélectionnées à l'état triplet sont les mêmes.

Cependant, il est possible pour les *single points* MS-CASPT2/MM de choisir un espace actif plus grand. Nous avons choisi de garder le même espace actif pour les calculs d'optimisation SA-CASSCF/MM et *single points* MS-CASPT2/MM afin d'avoir une certaine cohérence dans les calculs. L'espace choisi est correct pour les calculs effectués mais pourrait être amélioré. Premièrement, la présence d'une orbitale centrée sur la paire libre du soufre est inutile, l'occupation de cette orbitale reste constante à 2 électrons pendant toute l'approche. De plus, quelques orbitales importantes manquent à l'espace actif, il aurait été intéressant de rajouter plus d'orbitales présentes sur le dioxygène comme des orbitales  $\pi$ . De même l'orbitale représentant la liaison  $\sigma$  entre O1 et O2 aurait été un ajout plutôt pertinent dans le calcul afin d'étudier la possible cassure de cette liaison lors de la coordination. Finalement l'ajout de la liaison  $\sigma$  entre C2 et O4 dans l'espace actif symbolise la coupure de liaison entre la luciférine et l'AMP. Dans cette étude on force la cassure de cette liaison par l'ajout d'une contrainte D-ester. En ajoutant cette orbitale on aurait pu voir si le départ de l'AMP se fait de façon concertée avec l'approche de O<sub>2</sub>. Ces idées sont à tester mais prennent beaucoup de temps.

A mesure de l'approche du dioxygène et la formation du cycle dioxétanone, on observe certaines modifications dans les orbitales de l'espace actif. On ne discute ici que les orbitales obtenues dans les structures possédant un cycle dioxétanone, soit Int-t-dioxétanone, Int-A-t-dioxétanone, Int-B-t-dioxétanone et Int-B-s-dioxétanone. Pour les états singulets, l'espace actif obtenu contient la plupart des interactions attendues et la perte de l'orbitale centrée sur le soufre est apprécié. Ce qui est très intéressant est que la plupart des orbitales sont largement délocalisées sur l'ensemble de la partie QM. Les deux orbitales  $\sigma$  et  $\sigma^*$  représentant la liaison C1-O1 sont présentes et délocalisées sur d'autres orbitales comme les orbitales  $\pi$  des atomes N1 et C3. Les deux orbitales  $\pi^*$  du dioxygène sont très délocalisées sur la paire libre de l'azote N1 et un peu sur O4. Les orbitales  $\pi$  et  $\pi^*$  entre C2 et O3 sont les moins délocalisées mais présentent un peu de densité électronique sur O1 ou O2. Afin d'avoir un espace actif parfait il manque sans doute les orbitales représentant les liaisons  $\sigma$  et  $\sigma^*$  entre C2 et O2.

Pour les états triplets, les orbitales obtenues pour les structures Int-t-dioxétanone sont similaires à celle de Int-A-t-dioxétanone et Int-B-t-dioxétanone et semblent beaucoup moins pertinentes que dans le cas des calculs singulets. On note la présence de deux orbitales plus ou moins centrées sur le soufre et donc inutiles. On perd ainsi les deux orbitales  $\pi^*$  centrées sur le dioxygène. Le reste des orbitales sont similaires à celles obtenues pour l'état singulet.

Finalement on discute de l'occupation de ces orbitales. Le but de cette étude, outre la formation du dioxétanone est de comprendre quel mécanisme réactionnel est à l'œuvre, radicalaire ou chargé. Dans cette optique on a décidé de déprotone le carbone C1, rajoutant une charge -1 à la charge globale de la molécule. On s'attend dans une représentation de Lewis à avoir une densité de charge proche de -1 sur le carbone C1. Seulement dans les orbitales créées durant toute l'approche i.e. de D-coord égal à 3.5 Å à 1.5 Å, l'orbitale centrée sur ce carbone possède une occupation proche de 1 et les deux orbitales  $\pi^*$  centrées sur le dioxygène proche de 3 électrons. Les calculs CASSCF et le choix de cet espace actif force le système à adopter un mécanisme radicalaire avec un carbone  $C1^\bullet$  et un dioxygène  $O_2^\bullet-$ . On se rapproche alors du mécanisme réactionnel détaillé dans la littérature utilisant un ion superoxyde.

### 13.4.2 La formation du dioxétanone

En tenant compte des éléments développés dans le paragraphe précédent, on peut analyser le chemin réactionnel formé par l'approche de  $O_2$  et la création du cycle dioxétanone.

Le premier point à aborder est la valeur des énergies obtenues entre D-coord valant 3.5 Å et 2.3 Å. Pour tous ces points, l'énergie de S0, S1, T0 et T1 est quasiment dégénérée. Deux hypothèses peuvent être proposées. Dans la 1<sup>ère</sup>, la molécule de  $O_2$  et le substrat Int n'interagissent pas avant d'être suffisamment proche. Ainsi le système se trouve dans un état inactif jusqu'à D-coord valant 2.5 Å où le transfert d'un électron est possible entre Int et  $O_2$ . Ce transfert se traduit également par une levée de dégénérescence des états électroniques. La deuxième hypothèse consiste à dire qu'à un moment de l'approche de  $O_2$  les états électroniques doivent être dégénérés afin de permettre à des croisements entre états d'exister et ainsi modifier la nature de l'état électronique fondamental.

Pour les états singulets la stabilisation de l'énergie pour des valeurs de D-coord plus petites que 2.3 Å correspond à la formation de la liaison C1-O1 et au recouvrement des deux orbitales mono-occupées (à caractère radicalaire). On observe une perte d'énergie car la disparition des radicaux du milieu est un fort motif de stabilisation. L'énergie augmente faiblement lors de l'approche du second oxygène. Il est cependant compliqué d'expliquer pourquoi on n'observe pas la formation de la liaison C2-O2 conduisant au cycle dioxétanone.

Il est cependant clair que le départ de la partie AMP est un prérequis nécessaire à la formation du cycle dioxétanone. La présence de l'AMP tord la conformation du cycle dioxétanone empêchant la stabilisation de ce dernier. Dans le papier de Liu *et al.*<sup>18</sup>, les structures de départ avant cassure du cycle dioxétanone affichent une valeur de l'angle dièdre C1-O1-O2-C2 de l'ordre de 0°. Ces structures sont réalisées avec des calculs SA-CASSCF dans le vide sur le système thiazole dioxétanone. On

retrouve bien ce résultat dans les cycles dioxétanones formés après départ de l'AMP. Le cycle dioxétanone appelé Int-t-dioxétanone et formé pour les états triplet à la distance D-coord égale à 1.5 Å possède une valeur de l'angle dièdre de 25°, cette valeur se rapprochant de la structure TS du composé dioxétanone calculée dans le même papier.

On compare maintenant l'approche utilisant les méthodes multi-configurationnelles MS-CASPT2//SA-CASSCF/MM avec la minimisation DFT/MM. Avec les calculs DFT/MM, la structure Int-dioxétanone est obtenue par une simple minimisation sans départ de l'AMP et pour un état fondamental singulet. En revanche le calcul avec un état fondamental triplet ne converge pas. Pour les calculs SA-CASSCF/MM, le départ de l'AMP est un prérequis nécessaire afin de former le cycle dioxétanone. Une explication possible est que les calculs DFT/MM forcent la minimisation du système, qu'importe le moyen. La minimisation du système ne suis également pas le chemin réactionnel cohérent représentant l'approche attendu du O<sub>2</sub>. Il est alors préférable de ne tenir compte que des calculs MS-CASPT2//SA-CASSCF/MM.

Il est impossible pour l'instant d'expliquer pourquoi dans le cas de Int-A-s on ne peut pas former le cycle dioxétanone. Les calculs ne convergent pas et montrent plutôt la cassure de la liaison C1-C2, étape correspondant à la suite du mécanisme réactionnel. Il est alors possible que le structure possédant un cycle dioxétanone ne soit qu'une structure métastable.

Dans le cas de Int-B-s la molécule de dioxygène est repoussée et revient proche de la position de départ qu'elle occupe dans 4G36-O<sub>2</sub>-min. La formation du cycle dioxétanone se réalise relativement facilement en comparaison des autres simulations précédemment présentées. Pour cette approche, consistant à faire partir la molécule d'AMP avant la coordination de O<sub>2</sub>, on n'a malheureusement calculé que deux points le long du chemin réactionnel. Une perspective est alors de réaliser le chemin complet quand le substrat Int se détache de l'AMP.

L'énergie obtenue pour Int-B-s-dioxétanone est bien inférieure à l'énergie de départ et très légèrement supérieure au point le plus bas obtenu dans le chemin réactionnel (0.11 eV de différence), le point Int-s-O<sub>2</sub> avec D-coord valant 1.9 Å.

Pour les états triplets, l'augmentation croissante de l'énergie jusqu'à Int-t-O<sub>2</sub> avec D-coord égale à 1.6 Å est explicable car la liaison formée C2-O<sub>2</sub> ne correspond pas à la liaison formée entre les deux radicaux. Cependant pour les 3 structures différentes possédant un cycle dioxétanone soit Int-t-dioxétanone, Int-A-t-dioxétanone et Int-B-t-dioxétanone, l'énergie calculée est toujours supérieure à l'énergie de la structure 4G36-O<sub>2</sub>-min (entre +1.79 eV et +2.70 eV). Ces énergies sont plutôt importantes. Il est compliqué d'expliquer pourquoi le mécanisme avec les états triplet donne des énergies si importantes. Une conclusion évidente est de rejeter pour l'instant le mécanisme avec un état fondamental triplet.

## 13.5 Conclusion

Dans cette étude, on a réussi à obtenir, difficilement, la formation du cycle dioxétanone suite à l'approche de dioxygène. Des calculs de MD montrent que la molécule de dioxygène se déplace sans contrainte dans la protéine et sans la présence de barrière. Il est néanmoins nécessaire de fixer une contrainte afin de rapprocher le dioxygène du substrat Int. Une nouvelle fois cette approche ne met pas en exergue une barrière énergétique.

La formation du cycle dioxétanone par des calculs MS-CASPT2//SA-CASSCF/MM montre que ce mécanisme est radicalaire avec la présence d'un dioxygène superoxyde. Cependant, seuls les calculs avec un état fondamental singulet permettent d'obtenir une conformation finale avec une énergie inférieure à celle de la structure de départ. En outre, le départ de la molécule d'AMP est nécessaire afin de former le cycle dioxétanone. On retrouve également dans cette étude un mécanisme faisant intervenir un ion superoxyde via un transfert d'électron entre le substrat Int et la molécule de O<sub>2</sub> comme dans un certain nombre de publications<sup>135 136</sup>. Cette étude permet de mettre en valeur certaines informations inédites et surprenantes mais il reste encore beaucoup de travail afin de connaître la véritable histoire du mécanisme réactionnel de la bioluminescence chez les lucioles.

## 13.6 Perspective

Dans cette partie perspective on mettra en valeur certains calculs ou idées pouvant accroître les connaissances pour ce chapitre.

Comme on l'a développé dans les parties précédentes tester d'autres espaces actifs permettrait de mieux comprendre cette réaction. Dans un premier temps, on cherche à augmenter le nombre d'orbitales choisies sur O<sub>2</sub> afin de tenir compte par exemple de la liaison  $\sigma$  O-O. Dans un second temps, il faut vérifier si seul le mécanisme radicalaire est accessible via les calculs CASSCF ou si en changeant l'espace actif on peut arriver à un mécanisme entre deux charges.

On va également chercher à comprendre pourquoi il est compliqué d'obtenir le cycle dioxétanone dans certains cas, et si la structure Int possédant le cycle dioxétanone est une vraie structure stable ou seulement mésostable.

Enfin, la perspective la plus souhaitable et enviable serait de réaliser un calcul de type *ab initio* MD et de réaliser l'approche du dioxétanone seulement par dynamique *ab initio*. Le choix de l'espace actif sera alors déterminant.

Une perspective qui me tient à cœur consiste à observer les structures finales avec le substrat possédant le cycle dioxétanone. Dans ces dernières, un réseau de liaisons hydrogène se forme entre le résidu Hip 245, une molécule d'eau et l'AMP. On sait que le résidu Hip 245 a été protoné à la main et que ce dernier est présent sous forme de HIE dans la structure cristallographique de la luciférase. Un calcul



très intéressant est alors de modéliser le transfert de protons entre ces 3 molécules pour obtenir la molécule d'AMP protonée, l'AMPH, et ainsi avoir une vraie preuve de la molécule d'AMPH dans le milieu, cette molécule jouant un rôle important notamment dans la tautomérisation keto-énol.

Enfin, on pourrait maintenant s'intéresser à l'autre mécanisme réactionnel décrit dans la littérature, via l'ion hydroperoxyde<sup>2138</sup>.

## 13.7 Détails de calculs

### 13.7.1 Partie MM

Dans cette étude on utilise la structure cristallographique 4G36 issu de la luciole nord-américaine *Photinus Pylalis* obtenue par Branchini *et al.*. Cette structure représente la luciférase dans une conformation fermée au début de la réaction de bioluminescence. Le fichier 4G36.pdb a été téléchargé sur le site RSCB PDB, et les boucles manquantes ajoutées avec le logiciel Disgro<sup>96</sup>. Cette structure cristallographique a été résolue avec un substrat présent dans le site actif. Ce substrat se nomme le DLSA (5'-O-[N-(dehydroLuciferyl)-sulfamoyl]adenosine) et correspond à un analogue de l'intermédiaire réactionnel Int lors de la réaction de bioluminescence. Afin d'obtenir le modèle avec l'intermédiaire Int, la structure du DLSA est modifiée pour avoir les atomes correspondant à ceux de Int. Pour cela, l'atome de soufre est changé en phosphore et l'amine adjacente remplacé par un atome d'oxygène. L'atome C1 du cycle thiazolone est déprotoné pour avoir l'intermédiaire Int du mécanisme réactionnel de la bioluminescence présenté figure 13.1. Le résidu Hie 245 est un résidu proche et important dans le mécanisme de la bioluminescence. En outre, ce dernier peut capter un proton et alors être considéré comme Hip. On fait l'hypothèse de protoner ce dernier avec le proton issu du carbone C1. L'hydrogène déprotoné du substrat Int est donc placé sur le résidu Hip 245. Le substrat Int est donc chargé -3, une charge négative venant du cycle benzothiazole, une du cycle thiazole et la dernière du groupement phosphate. Les résidus de la protéine sont protonés avec le programme Leap de la suite logicielle Amber. Afin d'obtenir un système neutre plusieurs histidines ont été doublement protonées selon le pKa calculé avec le programme H++. Les histidines doublement protonées correspondent aux résidus 245, 310, 332, 419, 431 and 489. Finalement un ion Na<sup>+</sup> est ajouté dans le milieu. On obtient alors le modèle nommé 4G36-Int.

L'étude de la coordination du dioxygène nécessite la présence d'une molécule de O<sub>2</sub> dans le modèle. Une molécule de O<sub>2</sub> est ajoutée et positionnée comme indiqué 13.3. Afin de pouvoir effectuer des calculs de MD des paramètres de champ de force ont été développés pour le substrat Int et pour O<sub>2</sub>. Ces paramètres sont issus de paramètres précédents utilisés pour traiter l'oxyluciférine ou l'AMP<sup>111</sup>. Les charges ponctuelles affectées dans le champ de force proviennent de précédents calculs QM/MM. Pour O<sub>2</sub>, les deux oxygènes possèdent une charge nulle. Le modèle est alors appelé 4G36-O<sub>2</sub>.

### 13.7.2 Partie MD

Les simulations de dynamiques moléculaires classiques ont été réalisées avec la suite logicielle Amber14<sup>55</sup> afin d'obtenir une représentation du système pour une future optimisation QM/MM. Les modèles ont été solvatés par des molécules d'eau de type TIP3P dans une boîte de forme cuboïde. Le système résultant contient environ 90000 atomes au total, dont 28000 molécules d'eau formant une sphère de solvata-tion d'au moins 15 Å autour du substrat. Un champ de force de type AMBER99ff a été utilisé pour modéliser les résidus de la protéine. Le substrat reste gelé dans cette géométrie pendant toute la dynamique. Le déroulement d'une dynamique est le suivant : 3 minimisations sont d'abord réalisées, la 1<sup>ère</sup> minimise seulement la protéine en ge-lant les molécules d'eau. La deuxième minimisation ne gèle que les atomes d'hydro-gène alors que rien n'est gelé (à part le substrat) durant la 3<sup>ème</sup>. Le système est ensuite chauffé de 100K à 300K en 20 ps. Puis sous condition NPT avec T=300K, P=1 atm une dynamique de 10 ns ou plus si indiqué est conduite sous condition périodique aux bords avec comme pas de temps 2 fs. La température et la pression sont maintenues en utilisant l'algorithme de Berendsen<sup>61</sup> avec une constante de couplage de 5 ps et par des contraintes de type SHAKE appliquées sur toutes les liaisons impliquant des atomes d'hydrogène.

L'approche de O<sub>2</sub> dans la protéine a été réalisé par une simulation de type um-brella sampling. L'algorithme d'umbrella sampling a été conduit en prenant comme contrainte la distance  $D_{umbrella}$  correspondant à la distance entre les deux centres de masse des deux liaisons C9-C2 et O2-O1. La distance de départ est de 8.5 Å et celle de fin de 1.5 Å avec un pas de 0.2 Å. A chaque pas, 50 ps d'équilibration sont réalisée suivie de 150 ps de production. La dernière structure de la MD de produc-tion est récupérée pour chaque pas, ce qui permet de suivre les effets du repliement dans la simulation. Le biais introduit par la MD d'umbrella sampling est enlevé par la méthode d'analyse d'histogrammes pondérés WHAM<sup>99,100</sup>, afin de générer un profil d'énergie libre le long du chemin suivi.

### 13.7.3 Partie QM/MM

Les calculs QM/MM sont réalisés en utilisant un schéma de couplage ESPF entre Gaussian<sup>67</sup> et Tinker<sup>56</sup> ou Molcas<sup>113</sup> et Tinker (Gaussian 09d/Tinker ou Molcas 8.0/Tinker). La méthode ESPF (electrostatical potentiel fitted)<sup>68</sup> permet de prendre en compte le potentiel électrostatique externe crée par les points de charge du MM dans le calcul de la fonction d'onde de la partie QM. Un cutoff de 9 Å est pris pour éviter une sommation infinie. Au niveau MM, les charges de type Mulliken<sup>114</sup> de la partie QM sont prises en compte dans le calcul électrostatique. La technique dite de microitérations<sup>116</sup> est utilisée afin de faire converger la géométrie de l'espace MM choisi pour chaque pas de minimisation du QM.

Les calculs DFT/MM sont réalisés avec la fonctionnelle CAM-B3LYP et une base 6-31G(d,p).

L'approche du O<sub>2</sub> est également étudiée par des calculs MS-CASPT2//SA-CASSCF/MM. Dans ces derniers, le substrat Int est divisé en une partie QM et deux parties MM avec l'algorithme de l'atome lié. La partie contenant le cycle thiazole, le groupement phosphate et le groupement CH<sub>2</sub> entre le phosphate et le ribose correspond à la partie QM. Le cycle benzothiazole ainsi que la partie ribose+adénosine appartient à la partie MM. La molécule de O<sub>2</sub> est incluse dans la partie QM, qui contient alors 21 atomes et une charge de -2.

L'espace actif a été construit à partir de la structure 4G36-O<sub>2</sub>-min (voir Annexe A6). L'espace actif choisi correspond à 8 électrons dans 10 orbitales. Il comprend le système conjugué du cycle thiazolone, c'est à dire les orbitales  $\pi$  des atomes C3, N1, C1, C2, O3 et la paire libre de S1. Sur le dioxygène les deux  $\pi^*$  sont sélectionnées. Le détail des orbitales se trouve en annexe. Les orbitales sélectionnées à l'état triplet sont les mêmes.

Le chemin réactionnel correspondant à l'approche O<sub>2</sub> vis à vis du substrat est réalisé par une succession de calculs d'optimisation sous contrainte. Pour chaque pas un calcul SA-CASSCF est réalisé avec une base ANO-RCC-VTZP et 3 états électroniques. Les calculs sont réalisés pour un état fondamental singulet ou triplet. Puis, des calculs *single points* MS-CASPT2 sont effectués sur les géométries optimisées en SA-CASSCF. Ces calculs prennent en compte 3 états électroniques et utilisent un level shift de 0.1.

Les calculs SA-CASSCF/MM sont réalisés en commençant par la structure 4G36-O<sub>2</sub>-min, où la distance  $D_{umbrella}$  vaut 3.5 Å.



# **Conclusion**



Le développement des méthodes théoriques couplé avec l'augmentation de la puissance de calcul des machines permet aujourd'hui de répondre à des questions toujours plus pointues sur le fonctionnement de systèmes moléculaires complexes.

Lors des parties 1, 2 et 3 de cette thèse on a développé les principales méthodes théoriques utilisées, notamment l'approche QM/MM, primordiale dans l'étude des systèmes protéiques. Puis on a discuté des différents mécanismes chimiques permettant d'obtenir une émission de lumière, et détaillé les principales applications de la bioluminescence. Enfin, on s'est intéressé à la structure des protéines et à leur utilité.

Dans les parties 4 et 5 on a présenté les résultats obtenus lors de cette thèse concernant la bioluminescence chez les lucioles. Les chapitres 9 et 10 traitent des facteurs chimiques modifiant la longueur d'onde d'émission. Parmi ces facteurs, on a montré l'influence de la taille du système  $\pi$  conjugué du substrat sur la modulation de la couleur de l'émission. En outre, les différentes interactions apportées par la protéine ont été relevées et explicitées. Ainsi, l'effet électrostatique produit par les résidus polaires et par la formation d'un réseau de liaisons hydrogène dans la cavité de la protéine sont les principaux responsables du changement de la longueur d'onde de la bioluminescence. Il est alors possible de jouer avec ces différents facteurs afin de pouvoir nuancer l'émission du système.

Dans le chapitre 12, on a traité d'une des questions les plus discutées dans le domaine de la recherche sur les lucioles, la conformation du substrat. Alors que les chercheurs théoriciens utilisent majoritairement le substrat sous forme phénolate-kéto afin de reproduire la bioluminescence, les expérimentateurs affirment que le substrat phénolate-énol est également le responsable de l'émission. Cette étude a permis de mettre au jour un chemin réactionnel liant ces deux formes chimiques de l'oxyluciférine. Bien que ces travaux n'en soient qu'à leurs débuts, ils posent les bases de l'étude de la réactivité chimique dans les substrats émetteurs.

Dans le chapitre 13, on a délaissé la forme émissive oxyluciférine afin de s'intéresser à la réaction de bioluminescence en elle-même. L'étape chimique développée ici correspond à la coordination d'une molécule de  $O_2$  à un intermédiaire de réaction et à la formation du cycle dioxétanone, le cycle responsable de l'excitation chimique du substrat. Cette étape chimique est jusqu'à maintenant la moins étudiée dans l'ensemble du mécanisme de bioluminescence chez les lucioles. Les deux mécanismes proposés datent des années 1970 et restent pour l'instant à l'état d'hypothèse. Dans notre étude, on se focalise sur un seul mécanisme, qui nous semble plus réaliste au vu de l'environnement protéique et de notre connaissance du système. Ce mécanisme radicalaire fait intervenir un ion  $O_2$  superoxyde issu d'un transfert d'un électron entre le substrat et  $O_2$ . On montre que la réaction est réalisable et favorisée dans le cas d'un état fondamental singulet. En outre, le départ de la molécule d'AMP, liée de façon covalente à la luciférine est obligatoire si on veut observer la formation du cycle dioxétanone. L'étude du second mécanisme réactionnel représente la suite logique afin de comprendre un peu plus le fonctionnement de la bioluminescence

chez les lucioles.

Enfin, dans le chapitre 11, on a étudié la chimiluminescence d'une classe de molécules appelées 2-coumaranones. Ces molécules possèdent un comportement similaire à celui observé dans les lucioles dans la production de lumière. Chez les 2-coumaranones, et suite à la décomposition du cycle dioxétanone, 2 molécules identifiées peuvent correspondre à la molécule émettrice. Une étude conjointe expérimentateurs/théoriciens a permis de trouver la molécule jouant le rôle de précurseur de la chimiluminescence.

Cette thèse a montré l'apport d'une méthodologie couplant des simulations de dynamique moléculaire avec des calculs QM/MM dans la modélisation et la compréhension de systèmes bioluminescents.

L'utilisation de la méthode hybride QM/MM est indispensable afin de modéliser, analyser et comprendre les systèmes complexes issus des espèces bioluminescentes. Malgré quarante ans d'existence et un développement considérable, la méthode QM/MM reste encore difficile à appréhender. La mise en place de calculs utilisant cette approche hybride nécessite aussi bien une connaissance rigoureuse du système chimique étudié qu'une connaissance approfondie des méthodes théoriques. Une des difficultés de cette thèse fût par exemple le temps et la complexité requis pour bâtir le 1<sup>er</sup> modèle du complexe oxyluciférine/luciférase, qui a pris 3 mois.

Les simulations de dynamique moléculaire réalisées dans cette thèse ont apporté une vraie plus-value sur la modélisation du système. Parmi les résultats obtenus, l'attention a été focalisée sur l'influence exercée entre l'environnement protéique et le substrat chimique des lucioles suite à ces dynamiques. Dans les chapitres 9 et 10, on a pu expliciter les effets apportés par la protéine et jouant un rôle dans la modulation de la longueur d'onde d'émission de la bioluminescence. La principale découverte concerne l'effet du réseau de liaisons hydrogène à l'intérieur de la cavité. La présence de molécules d'eau dans la cavité formant des liaisons hydrogène entre le substrat et certains résidus était déjà observée dans de précédentes publications, mais on a pu ici la rationaliser et comprendre son impact. Ainsi, on peut moduler la longueur d'onde d'émission selon la position de ces molécules d'eau autour du substrat. De plus, certains résidus comme l'arginine ont un fort impact sur la valeur de l'émission, même en étant situé dans la seconde sphère de solvation du substrat. Finalement, ces dynamiques permettent d'étudier des mouvements de la structure protéique. Dans le chapitre 10, on a pu replier le domaine C-terminal de la luciférase et montrer l'impact de ce dernier sur l'émission et sur la stabilité des substrats à l'intérieur de la cavité. Ces enseignements sont maintenant à réutiliser pour de nouvelles protéines, comme la RE<sub>Ph</sub> dont les calculs débutent en ce moment.

Le rôle des molécules d'eau dans la cavité est également observé dans l'ensemble des études réalisées dans la partie 5 de cette thèse. Dans le chapitre 12, j'ai montré que la présence d'une molécule d'eau entre OxyLH<sub>2</sub> et l'AMPH permet la tautomérisation du substrat phénolate-kéto en phénolate-énol. C'est la première



fois que le rôle de cette molécule d'eau est mise en avant dans la tautomérisation, et sa présence est indispensable dans la réaction. Enfin dans le chapitre 13, la protéine permet de stabiliser le molécule de  $O_2$  pendant toute l'approche vers le substrat Int via un réseau de liaisons hydrogène avec l'Hip 245, une molécule d'eau et  $O_2$ . A nouveau, la molécule d'eau joue un rôle important dans la formation du cycle dioxétanone lors de la réaction de bioluminescence. De plus, le positionnement de cette molécule d'eau, entre le résidu Hip 245 et l'AMP permet peut-être d'expliquer comment obtenir la molécule d'AMPH et ainsi retrouver les structures de départ utilisées dans le calcul de la bioluminescence des lucioles.

Un autre point important abordé dans cette thèse est la relation entre expérimentateurs et théoriciens. J'ai abordé plusieurs fois dans certaines parties la difficulté de modéliser des systèmes enzymatiques sans l'aide de groupe d'expérimentateurs. Heureusement des collaborations se mettent progressivement en place afin de réfléchir et d'échanger sur des travaux communs. Ainsi, lors de cette thèse, j'ai pu participer à deux collaborations avec deux groupes d'expérimentateurs. La première collaboration avec un groupe allemand a également permis de mettre en commun des compétences en plus des idées. Un doctorant allemand a pu venir en France apprendre à réaliser des calculs de chimie quantique et je suis allé en Allemagne synthétiser des composés chimiques. Cette collaboration a donné lieu à une publication visible dans cette thèse et étudiée dans le chapitre 11.

La deuxième collaboration est toujours en cours et donnera bientôt lieu à une publication. Le groupe de cristallographes d'Abu Dhabi synthétise de nombreux mutants de la luciférase et a obtenu récemment de nouvelles structures cristallographiques. Comme on a pu le voir dans le chapitre 10, ces chercheurs nous ont posés beaucoup (énormément) de questions auxquelles on a essayé de répondre. Notre connaissance mutuelle du système a permis des échanges très constructifs et de nombreuses nouvelles idées d'expériences ou de calculs à faire, notamment lors de l'étude de mutations ponctuelles dans la protéine. La suite de cette collaboration va essayer de coupler encore plus ces deux aspects de la recherche sur les systèmes bioluminescents via l'étude systématique de mutations imaginées par les deux parties.

Finalement, le travail sur la tautomérisation kéto-énol ne représente qu'un prélude à une collaboration entre chercheurs de Strasbourg, Marseille et Paris regroupés sous un contrat ANR qui a débuté en janvier dernier. En plus d'une étude expérimentale et théorique sur la conformation de l'émetteur oxyluciférine, ce projet ANR s'intéresse aussi à la modélisation d'un système protéique pH dépendant, permettant l'étude de l'effet de l'acidité du solvant sur le site actif et sur le substrat.

Finalement le message que j'aimerais faire passer concerne l'importance des simulations de dynamique moléculaire dans l'étude de la bioluminescence. Il est nécessaire de franchir un nouveau pas dans la complexité des calculs théoriques et ne plus se limiter à une approche statique du système modélisé en QM/MM mais évoluer vers une approche dynamique.

## CONCLUSION

# Bibliographie

- [1] Aristote. *De anima*. Number ISBN 9782711620463. Librairie Philosophique Vrin, 1995.
- [2] Jesse A. Sundlov, Danielle M. Fontaine, Tara L. Southworth, Bruce R. Branchini, and Andrew M. Gulick. Crystal structure of firefly luciferase in a second catalytic conformation supports a domain alternation mechanism. *Biochemistry*, 51(33) :6493–6495, 2012. PMID : 22852753.
- [3] Elena Conti, Nick P Franks, and Peter Brick. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure*, 4(3) :287 – 298, 1996.
- [4] Emil H. White, Frank. McCapra, and George F. Field. The structure and synthesis of firefly luciferin. *Journal of the American Chemical Society*, 85(3) :337–343, 1963.
- [5] Zinaida M. Kaskova, Aleksandra S. Tsarkova, and Ilia V. Yampolsky. 1001 lights : luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. *Chem. Soc. Rev.*, 45 :6048–6077, 2016.
- [6] Michio Kakiuchi, Soichiro Ito, Minoru Yamaji, Vadim R. Viviani, Shojiro Maki, and Takashi Hirano. Spectroscopic properties of amine-substituted analogues of firefly luciferin and oxyluciferin. *Photochemistry and Photobiology*, 93(2) :486–494, 2017.
- [7] Chihiro Miura, Masahiro Kiyama, Satoshi Iwano, Kazuto Ito, Rika Obata, Takashi Hirano, Shojiro Maki, and Haruki Niwa. Synthesis and luminescence properties of biphenyl-type firefly luciferin analogs with a new, near-infrared light-emitting bioluminophore. *Tetrahedron*, 69(46) :9726 – 9734, 2013.
- [8] Carolyn C. Woodroffe, Poncho L. Meisenheimer, Dieter H. Klaubert, Yumi Kovic, Justin C. Rosenberg, Curran E. Behney, Tara L. Southworth, and Bruce R. Branchini. Novel heterocyclic analogues of firefly luciferin. *Biochemistry*, 51(49) :9807–9813, 2012. PMID : 23164087.
- [9] Amit P. Jathoul, Helen Grounds, James C. Anderson, and Martin A. Pule. A dual-color far-red to near-infrared firefly luciferin analogue designed for multiparametric bioluminescence imaging. *Angewandte Chemie International Edition*, 53(48) :13059–13063, 2014.

## BIBLIOGRAPHIE

- [10] Rachel C. Steinhardt, Jessica M. O'Neill, Colin M. Rathbun, David C. McCutcheon, Miranda A. Paley, and Jennifer A. Prescher. Design and synthesis of an alkynyl luciferin analogue for bioluminescence imaging. *Chemistry - A European Journal*, 22 :3671–3675, 2016.
- [11] Xue Qin Ran, Xin Zhou, and John D. Goddard. The spectral–structural relationship of a series of oxyluciferin derivatives. *ChemPhysChem*, 16(2) :396–402, 2015.
- [12] Carolyn C. Woodroffe, John W. Shultz, Monika G. Wood, Jean Osterman, James J. Cali, William J. Daily, Poncho L. Meisenheimer, and Dieter H. Klaubert. N-alkylated 6'-aminoluciferins are bioluminescent substrates for ultra-glo and quantillum luciferase : New potential scaffolds for bioluminescent assays. *Biochemistry*, 47(39) :10383–10393, 2008. PMID : 18771284.
- [13] Narges Kh. Tafreshi, Saman Hosseinkhani, Majid Sadeghizadeh, Mehdi Sadeghi, Bijan Ranjbar, and Hossein Naderi-Manesh. The influence of insertion of a critical residue (arg356) in structure and bioluminescence spectra of firefly luciferase. *Journal of Biological Chemistry*, 282(12) :8641–8647, 2007.
- [14] Yulia Modestova and Natalia N. Ugarova. Color-shifting mutations in the c-domain of *l. mingrelica* firefly luciferase provide new information about the domain alternation mechanism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1864(12) :1818 – 1826, 2016.
- [15] Mikhail I. Koksharov and Natalia N. Ugarova. Strategy of mutual compensation of green and red mutants of firefly luciferase identifies a mutation of the highly conservative residue e457 with a strong red shift of bioluminescence. *Photochem. Photobiol. Sci.*, 12 :2016–2027, 2013.
- [16] Bruce R. Branchini, Danielle M. Ablamsky, Martha H. Murtiashaw, Lerna Uzasci, Hugo Fraga, and Tara L. Southworth. Thermostable red and green light-producing firefly luciferase mutants for bioluminescent reporter applications. *Analytical Biochemistry*, 361(2) :253 – 262, 2007.
- [17] Vadim R. Viviani, Danilo Amaral, Rogilene Prado, and Frederico G. C. Arnoldi. A new blue-shifted luciferase from the brazilian amydetes fanestratus (coleoptera : Lampyridae) firefly : molecular evolution and structural/functional properties. *Photochem. Photobiol. Sci.*, 10 :1879–1886, 2011.
- [18] Fengyi Liu, Yajun Liu, Luca De Vico, and Roland Lindh. A casscf/caspt2 approach to the decomposition of thiazole-substituted dioxetanone : Substitution effects and charge-transfer induced electron excitation. *Chemical Physics Letters*, 484(1) :69 – 75, 2009.
- [19] Isabelle Navizet, Daniel Roca-Sanjuán, Ling Yue, Ya-Jun Liu, Nicolas Ferré, and Roland Lindh. Are the bio- and chemiluminescence states of the firefly oxyluciferin the same as the fluorescence state? *Photochemistry and Photobiology*, 89(2) :319–325, 2013.

## BIBLIOGRAPHIE

- [20] Romain Berraud-Pache and Isabelle Navizet. Qm/mm calculations on a newly synthesised oxyluciferin substrate : new insights into the conformational effect. *Phys. Chem. Chem. Phys.*, 18 :27460–27467, 2016.
- [21] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28 :1049–1070, Dec 1926.
- [22] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20) :457–484, 1927.
- [23] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry : Introduction to Advanced Structure Theory*. Number ISBN 0486691861. Dover Publications, 1996.
- [24] Massimo Olivucci. *Computational photochemistry*. Number ISBN 0444521100. Elsevier, 2005.
- [25] J. C. Slater. The theory of complex spectra. *Phys. Rev.*, 34 :1293–1322, Nov 1929.
- [26] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23 :69–89, Apr 1951.
- [27] G. G. Hall. The molecular orbital theory of chemical valency. viii. a method of calculating ionization potentials. *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, 205(1083) :541–552, 1951.
- [28] Per E. M. Siegbahn, Jan Almlöf, Anders Heiberg, and Björn O. Roos. The complete active space scf (casscf) method in a newton–raphson formulation with application to the hno molecule. *The Journal of Chemical Physics*, 74(4) :2384–2396, 1981.
- [29] Bjorn O. Roos, Peter R. Taylor, and Per E.M. Sigbahn. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chemical Physics*, 48(2) :157 – 173, 1980.
- [30] Kerstin Andersson, Bjorn O. Roos, and Per-Ake Malmqvist. Second order perturbation theory with a complete active space self consistent field reference function. *The Journal of Chemical Physics*, 96(2) :1218–1226, 1992.
- [31] Chr. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46 :618–622, Oct 1934.
- [32] Björn O Roos and Kerstin Andersson. Multiconfigurational perturbation theory with level shift — the cr2 potential revisited. *Chemical Physics Letters*, 245(2) :215 – 223, 1995.
- [33] Giovanni Ghigo, Björn O. Roos, and Per Åke Malmqvist. A modified definition of the zeroth-order hamiltonian in multiconfigurational perturbation theory (caspt2). *Chemical Physics Letters*, 396(1–3) :142 – 149, 2004.

## BIBLIOGRAPHIE

- [34] James Finley, Per Åke Malmqvist, Björn O. Roos, and Luis Serrano-Andrés. The multi-state caspt2 method. *Chemical Physics Letters*, 288(2) :299 – 306, 1998.
- [35] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136 :B864–B871, Nov 1964.
- [36] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140 :A1133–A1138, Nov 1965.
- [37] Narbe Mardirossian and Martin Head-Gordon. [small omega]b97x-v : A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.*, 16 :9904–9924, 2014.
- [38] John P. Perdew, Adrienn Ruzsinszky, Jianmin Tao, Viktor N. Staroverov, Gustavo E. Scuseria, and Gábor I. Csonka. Prescription for the design and selection of density functional approximations : More constraint satisfaction with fewer fits. *The Journal of Chemical Physics*, 123(6) :062201, 2005.
- [39] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations : a critical analysis. *Canadian Journal of Physics*, 58(8) :1200–1211, 1980.
- [40] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77 :3865–3868, Oct 1996.
- [41] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Peder-son, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces : Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46 :6671–6687, Sep 1992.
- [42] Yan Zhao and Donald G. Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements : two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(1) :215–241, 2008.
- [43] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters : The pbe0 model. *The Journal of Chemical Physics*, 110(13) :6158–6170, 1999.
- [44] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45) :11623–11627, 1994.
- [45] Axel D. Becke. Density functional thermochemistry. iii. the role of exact exchange. *The Journal of Chemical Physics*, 98(7) :5648–5652, 1993.

## BIBLIOGRAPHIE

- [46] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38 :3098–3100, Sep 1988.
- [47] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the collesalvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37 :785–789, Jan 1988.
- [48] Takeshi Yanai, David P Tew, and Nicholas C Handy. A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chemical Physics Letters*, 393(1–3) :51 – 57, 2004.
- [49] Stefan Grimme. Accurate description of van der waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, 25(12) :1463–1473, 2004.
- [50] Stefan Grimme. Semiempirical hybrid density functional with perturbative second-order correlation. *The Journal of Chemical Physics*, 124(3) :034108, 2006.
- [51] Tobias Schwabe and Stefan Grimme. Towards chemical accuracy for the thermodynamics of large molecules : new hybrid density functionals including non-local correlation effects. *Phys. Chem. Chem. Phys.*, 8 :4398–4401, 2006.
- [52] Björn O. Roos, Valera Veryazov, and Per-Olof Widmark. Relativistic atomic natural orbital type basis sets for the alkaline and alkaline-earth atoms applied to the ground-state potentials for the corresponding dimers. *Theoretical Chemistry Accounts*, 111(2) :345–351, 2004.
- [53] Björn O. Roos, Roland Lindh, Per-Åke Malmqvist, Valera Veryazov, , and Per-Olof Widmark. Main group atoms and dimers studied with a new relativistic ano basis set. *The Journal of Physical Chemistry A*, 108(15) :2851–2858, 2004.
- [54] A. R. Leach. *Molecular Modelling : Principles and Applications*. Number ISBN 0582239338. Longman Pub Group, 1996.
- [55] D.A. Case and al. Amber 2015. 2015.
- [56] J. W. Ponder. Tinker, software tools for molecular design, version 5.1 ; department of biochemistry and molecular biophysics, washington university school of medicine, st. louis, mo. available at : <http://dasher.wustl.edu/tinker>. 2005.
- [57] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. Charmm : The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10) :1545–1614, 2009.

## BIBLIOGRAPHIE

- [58] Jay W. Ponder, Chuanjie Wu, Pengyu Ren, Vijay S. Pande, John D. Chodera, Michael J. Schnieders, Imran Haque, David L. Mobley, Daniel S. Lambrecht, Robert A. DiStasio, Martin Head-Gordon, Gary N. I. Clark, Margaret E. Johnson, and Teresa Head-Gordon. Current status of the amoeba polarizable force field. *The Journal of Physical Chemistry B*, 114(8) :2549–2564, 2010. PMID : 20136072.
- [59] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffrey D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2) :926–935, 1983.
- [60] Michael W. Mahoney and William L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics*, 112(20) :8910–8922, 2000.
- [61] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8) :3684–3690, 1984.
- [62] D. J. Evans and B. L. Holian. The nose–hoover thermostat. *The Journal of Chemical Physics*, 83(8) :4069–4074, 1985.
- [63] Hans C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4) :2384–2393, 1980.
- [64] Paul Sherwood, Alex H. de Vries, Martyn F. Guest, Georg Schreckenbach, C. Richard A. Catlow, Samuel A. French, Alexey A. Sokol, Stefan T. Bromley, Walter Thiel, Alex J. Turner, Salomon Billeter, Frank Terstegen, Stephan Thiel, John Kendrick, Stephen C. Rogers, John Casci, Mike Watson, Frank King, Elly Karlsen, Merethe Sjøvoll, Adil Fahmi, Ansgar Schäfer, and Christian Lennartz. Quasi : A general purpose implementation of the qm/mm approach and its application to problems in catalysis. *Journal of Molecular Structure : THEOCHEM*, 632(1) :1 – 28, 2003.
- [65] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions : Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2) :227 – 249, 1976.
- [66] Stefan Dapprich, István Komáromi, K. Suzie Byun, Keiji Morokuma, and Michael J. Frisch. A new {ONIOM} implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives1. *Journal of Molecular Structure : {THEOCHEM}*, 461–462 :1 – 21, 1999.
- [67] 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,



## BIBLIOGRAPHIE

- 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, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian~09 Revision D.01. Gaussian Inc. Wallingford CT 2009.
- [68] Nicolas Ferré and János G. Ángyán. Approximate electrostatic interaction operator for qm/mm calculations. *Chemical Physics Letters*, 356(3–4) :331 – 339, 2002.
- [69] Xavier Assfeld and Jean-Louis Rivail. Quantum chemical computations on parts of large molecules : the ab initio local self consistent field method. *Chemical Physics Letters*, 263(1) :100 – 106, 1996.
- [70] Dean M. Philipp and Richard A. Friesner. Mixed ab initio qm/mm modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *Journal of Computational Chemistry*, 20(14) :1468–1494, 1999.
- [71] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A generalized hybrid orbital (gho) method for the treatment of boundary atoms in combined qm/mm calculations. *The Journal of Physical Chemistry A*, 102(24) :4714–4721, 1998.
- [72] Kristian Støchkel, Christian Nygaard Hansen, Jørgen Houmøller, Lisbeth Munksgaard Nielsen, Kelvin Anggara, Mathieu Linares, Patrick Norman, Fernando Nogueira, Oleg V. Maltsev, Lukas Hintermann, Steen Brøndsted Nielsen, Panče Naumov, and Bruce F. Milne. On the influence of water on the electronic structure of firefly oxyluciferin anions from absorption spectroscopy of bare and monohydrated ions in vacuo. *Journal of the American Chemical Society*, 135(17) :6485–6493, 2013. PMID : 23557511.
- [73] B. Mennucci, E. Cancès, , and J. Tomasi. Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method : theoretical bases, computational implementation, and numerical applications. *The Journal of Physical Chemistry B*, 101(49) :10506–10517, 1997.
- [74] E. Cancès, B. Mennucci, and J. Tomasi. A new integral equation formalism for the polarizable continuum model : Theoretical background and applications to isotropic and anisotropic dielectrics. *The Journal of Chemical Physics*, 107(8) :3032–3041, 1997.
- [75] A. Bondi. van der waals volumes and radii. *The Journal of Physical Chemistry*, 68(3) :441–451, 1964.

BIBLIOGRAPHIE

- [76] Darrin M. York and Martin Karplus. A smooth solvation potential based on the conductor-like screening model. *The Journal of Physical Chemistry A*, 103(50) :11060–11079, 1999.
- [77] Giovanni Scalmani and Michael J. Frisch. Continuous surface charge polarizable continuum models of solvation. i. general formalism. *The Journal of Chemical Physics*, 132(11) :114110, 2010.
- [78] Marius Wanko, Michael Hoffmann, Thomas Frauenheim, and Marcus Elstner. Effect of polarization on the opsin shift in rhodopsins. 1. a combined qm/qm/mm model for bacteriorhodopsin and pharaonis sensory rhodopsin ii. *The Journal of Physical Chemistry B*, 112(37) :11462–11467, 2008. PMID : 18698712.
- [79] Denis Jacquemin, Eric A. Perpète, Adèle D. Laurent, Xavier Assfeld, and Carlo Adamo. Spectral properties of self-assembled squaraine-tetralactam : a theoretical assessment. *Physical Chemistry Chemical Physics*, 11(8) :1258–1262, 2009.
- [80] Detlef Lohse, Barbara Schmitz, and Michel Versluis. Snapping shrimp make flashing bubbles. *Nature*, 413 :477–478, 2001.
- [81] E. Fred Schubert, Thomas Gessmann, and Jong Kyu Kim. *Light Emitting Diodes*. John Wiley Sons, Inc., 2000.
- [82] Ullrich Mitschke and Peter Bauerle. The electroluminescence of organic materials. *J. Mater. Chem.*, 10 :1471–1507, 2000.
- [83] A.M Jiménez and M.J Navas. Chemiluminescence detection systems for the analysis of explosives. *Journal of Hazardous Materials*, 106(1) :1 – 8, 2004.
- [84] Geoffrey J. Lofthouse, Hans Suschitzky, Basil J. Wakefield, Roger A. Whittaker, and Brian Tuck. Synthesis and chemiluminescent reactions of some 3-alkoxycarbamoylbenzo[b]furan-2(3h)-ones. *J. Chem. Soc., Perkin Trans. 1*, pages 1634–1639, 1979.
- [85] Stefan Schramm, Isabelle Navizet, Panče Naumov, Naba K. Nath, Romain Berraud-Pache, Pascal Oesau, Dieter Weiss, and Rainer Beckert. The light emitter of the 2-coumaranone chemiluminescence : Theoretical and experimental elucidation of a possible model for bioluminescent systems. *European Journal of Organic Chemistry*, 2016(4) :678–681, 2016.
- [86] Stefan Schramm, Luiz Francisco Monteiro Leite Ciscato, Pascal Oesau, Reimar Krieg, Jan Frank Richter, Isabelle Navizet, Daniel Roca-Sanjuan, Dieter Weiss, and Rainer Beckert. Investigations on the synthesis and chemiluminescence of novel 2-coumaranones - ii. *ARKIVOC*, pages 44–59, 2015.
- [87] Steven H.D. Haddock, Mark A. Moline, and James F. Case. Bioluminescence in the sea. *Annual Review of Marine Science*, 2(1) :443–493, 2010. PMID : 21141672.

## BIBLIOGRAPHIE

- [88] Miranda A. Paley and Jennifer A. Prescher. Bioluminescence : a versatile technique for imaging cellular and molecular features. *Med. Chem. Commun.*, 5 :255–267, 2014.
- [89] Dan M. Close, Tingting Xu, Gary S. Sayler, and Steven Ripp. In vivo bioluminescent imaging (bli) : Noninvasive visualization and interrogation of biological processes in living animals. *Sensors*, 11(1) :180–206, Dec 2010.
- [90] Jean-René Martin, Kelly L. Rogers, Carine Chagneau, and Philippe Brûlet. In vivo bioluminescence imaging of ca<sup>2+</sup> signalling in the brain of drosophila. *PLOS ONE*, 2(3) :1–8, 03 2007.
- [91] Ho-Taek Song, Elaine K. Jordan, Bobbi K. Lewis, Wei Liu, Justin Ganjei, Brenda Klaunberg, Daryl Despres, Diane Palmieri, and Joseph A. Frank. Rat model of metastatic breast cancer monitored by mri at 3 tesla and bioluminescence imaging with histological correlation. *Journal of Translational Medicine*, 7(1) :88, 2009.
- [92] Eva A Naumann, Adam R Kampff, David A Prober, Alexander F Schier, and Florian Engert. Monitoring neural activity with bioluminescence during natural behavior. *Nat Neurosci*, 13 :513–510, 2010.
- [93] Spencer M. Steinberg, Edward J. Poziomek, William H. Engelmann, and Kim R. Rogers. A review of environmental applications of bioluminescence measurements. *Chemosphere*, 30(11) :2155 – 2197, 1995.
- [94] G. Fermi, M.F. Perutz, B. Shaanan, and R. Fourme. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *Journal of Molecular Biology*, 175(2) :159 – 174, 1984.
- [95] Ramu Anandkrishnan, Boris Aguilar, and Alexey V. Onufriev. H++ 3.0 : automating pk prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, 40(W1) :W537–W541, 2012.
- [96] Ke Tang, Jinfeng Zhang, and Jie Liang. Fast protein loop sampling and structure prediction using distance-guided sequential chain-growth monte carlo method. *PLoS Comput Biol*, 10(4) :e1003539, 04 2014.
- [97] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chemical Reviews*, 116(14) :7898–7936, 2016. PMID : 27333362.
- [98] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews : Computational Molecular Science*, 1(6) :932–942, 2011.

## BIBLIOGRAPHIE

- [99] Shankar Kumar, John M. Rosenberg, Djamel Bouzida, Robert H. Swendsen, and Peter A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry*, 16(11) :1339–1350, 1995.
- [100] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1) :275 – 282, 1995.
- [101] Mihaly Mezei Xuan-Yu Meng, Hong-Xing Zhang and Meng Cui. Molecular docking : A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, 7(2) :146–157, 2011.
- [102] Reed B. Jacob, Tim Andersen, and Owen M. McDougal. Accessible high-throughput virtual screening molecular docking software for students and educators. *PLOS Computational Biology*, 8(5) :1–5, 05 2012.
- [103] Erik Lindahl and Mark SP Sansom. Membrane proteins : molecular dynamics simulations. *Current Opinion in Structural Biology*, 18(4) :425 – 431, 2008. Membranes / Engineering and design.
- [104] Fatemeh Khalili-Araghi, James Gumbart, Po-Chao Wen, Marcos Sotomayor, Emad Tajkhorshid, and Klaus Schulten. Molecular dynamics simulations of membrane channels and transporters. *Current Opinion in Structural Biology*, 19(2) :128 – 137, 2009. Theory and simulation / Macromolecular assemblages.
- [105] Ken A. Dill and Justin L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338(6110) :1042–1046, 2012.
- [106] Christopher M. Dobson. Protein folding and misfolding. *Journal of Hazardous Materials*, Nature(426) :884–890, 2003.
- [107] Yuan-Yuan Cheng, Jia Zhu, and Ya-Jun Liu. Theoretical tuning of the firefly bioluminescence spectra by the modification of oxyluciferin. *Chemical Physics Letters*, 591 :156 – 160, 2014.
- [108] Shufeng Chen, Isabelle Navizet, Roland Lindh, Yajun Liu, and Nicolas Ferré. Hybrid qm/mm simulations of the obelin bioluminescence and fluorescence reveal an unexpected light emitter. *The Journal of Physical Chemistry B*, 118(11) :2896–2903, 2014. PMID : 24601757.
- [109] Adèle D. Laurent, Aymeric Blondel, and Denis Jacquemin. Choosing an atomic basis set for td-dft, soppa, adc(2), cis(d), cc2 and eom-ccsd calculations of low-lying excited states of organic dyes. *Theoretical Chemistry Accounts*, 134(6) :1–11, 2015.
- [110] J. Patrick Zobel, Juan J. Nogueira, and Leticia Gonzalez. The ipea dilemma in caspt2. *Chem. Sci.*, 8 :1482–1499, 2017.

BIBLIOGRAPHIE

- [111] Isabelle Navizet, Ya-Jun Liu, Nicolas Ferré, Hong-Yan Xiao, Wei-Hai Fang, and Roland Lindh. Color-tuning mechanism of firefly investigated by multi-configurational perturbation method. *Journal of the American Chemical Society*, 132(2) :706–712, 2010. PMID : 20014859.
- [112] Goran Kovačević and Valera Veryazov. Luscus : molecular viewer and editor for molcas. *Journal of Cheminformatics*, 7(1) :1–10, 2015.
- [113] Francesco Aquilante, Luca De Vico, Nicolas Ferré, Giovanni Ghigo, Per-Åke Malmqvist, Pavel Neogrády, Thomas Bondo Pedersen, Michal Pitoňák, Markus Reiher, Björn O. Roos, Luis Serrano-Andrés, Miroslav Urban, Valera Veryazov, and Roland Lindh. Molcas 7 : The next generation. *Journal of Computational Chemistry*, 31(1) :224–247, 2010.
- [114] R. S. Mulliken. Electronic population analysis on lcao–mo molecular wave functions. i. *The Journal of Chemical Physics*, 23(10) :1833–1840, 1955.
- [115] Jeng-Da Chai and Martin Head-Gordon. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10 :6615–6620, 2008.
- [116] Federico Melaccio, Massimo Olivucci, Roland Lindh, and Nicolas Ferré. Unique qm/mm potential energy surface exploration using microiterations. *International Journal of Quantum Chemistry*, 111(13) :3339–3346, 2011.
- [117] Shu-Feng Chen, Ya-Jun Liu, Isabelle Navizet, Nicolas Ferré, Wei-Hai Fang, and Roland Lindh. Systematic theoretical investigation on the light emitter of firefly. *Journal of Chemical Theory and Computation*, 7(3) :798–803, 2011.
- [118] Takashi Hirano, Yosuke Hasumi, Kazuhiro Ohtsuka, Shojiro Maki, Haruki Niwa, Minoru Yamaji, and Daisuke Hashizume. Spectroscopic studies of the light-color modulation mechanism of firefly (beetle) bioluminescence. *Journal of the American Chemical Society*, 131(6) :2385–2396, 2009. PMID : 19159303.
- [119] Chang-ik Song and Young Min Rhee. Dynamics on the electronically excited state surface of the bioluminescent firefly luciferase–oxyluciferin system. *Journal of the American Chemical Society*, 133(31) :12040–12049, 2011. PMID : 21728350.
- [120] Takashi Hirano, Yosuke Hasumi, Kazuhiro Ohtsuka, Shojiro Maki, Haruki Niwa, Minoru Yamaji, and Daisuke Hashizume. Spectroscopic studies of the light-color modulation mechanism of firefly (beetle) bioluminescence. *Journal of the American Chemical Society*, 131(6) :2385–2396, 2009. PMID : 19159303.
- [121] Luís Pinto da Silva and Joaquim C.G. Esteves da Silva. Computational studies of the luciferase light-emitting product : Oxyluciferin. *Journal of Chemical Theory and Computation*, 7(4) :809–817, 2011.

## BIBLIOGRAPHIE

- [122] Avisek Ghose, Mateusz Rebarz, Oleg V. Maltsev, Lukas Hintermann, Cyril Ruckebusch, Eduard Fron, Johan Hofkens, Yves Mély, Panče Naumov, Michel Sliwa, and Pascal Didier. Emission properties of oxyluciferin and its derivatives in water : Revealing the nature of the emissive species in firefly bioluminescence. *The Journal of Physical Chemistry B*, 119(6) :2638–2649, 2015. PMID : 25364813.
- [123] Yoshifumi Noguchi, Miyabi Hiyama, Motoyuki Shiga, Osamu Sugino, and Hidesumi Akiyama. Reverse stability of oxyluciferin isomers in aqueous solutions. *The Journal of Physical Chemistry B*, 120(34) :8776–8783, 2016. PMID : 27479583.
- [124] Panče Naumov, Yutaka Ozawa, Kei Ohkubo, and Shunichi Fukuzumi. Structure and spectroscopy of oxyluciferin, the light emitter of the firefly bioluminescence. *Journal of the American Chemical Society*, 131(32) :11590–11605, 2009. PMID : 19722653.
- [125] Joris J. Snellenburg, Sergey P. Laptinok, Richard J. DeSa, Panče Naumov, and Kyril M. Solntsev. Excited-state dynamics of oxyluciferin in firefly luciferase. *Journal of the American Chemical Society*, 138(50) :16252–16258, 2016. PMID : 27998082.
- [126] Saman Hosseinkhani. Molecular enigma of multicolor bioluminescence of firefly luciferase. *Cellular and Molecular Life Sciences*, 68(7) :1167–1182, Apr 2011.
- [127] Clotilde S. Cucinotta, Alice Ruini, Alessandra Catellani, and Andrés Stirling. Ab initio molecular dynamics study of the keto–enol tautomerism of acetone in solution. *ChemPhysChem*, 7(6) :1229–1234, 2006.
- [128] Giuliano Alagona and Caterina Ghio. Keto-enol tautomerism in linear and cyclic beta-diketones : A dft study in vacuo and in solution. *International Journal of Quantum Chemistry*, 108(10) :1840–1855, 2008.
- [129] Nicolas Cheron, Denis Jacquemin, and Paul Fleurat-Lessard. A qualitative failure of b3lyp for textbook organic reactions. *Phys. Chem. Chem. Phys.*, 14 :7170–7175, 2012.
- [130] Michael Page and James W. McIver Jr. On evaluating the reaction path hamiltonian. *The Journal of Chemical Physics*, 88(2) :922–935, 1988.
- [131] Michael Page, Charles Doubleday, and James W. McIver Jr. Following steepest descent reaction paths. the use of higher energy derivatives with ab initio electronic structure methods. *The Journal of Chemical Physics*, 93(8) :5634–5642, 1990.
- [132] Hrant P. Hratchian and H. Bernhard Schlegel. Accurate reaction paths using a hessian based predictor–corrector integrator. *The Journal of Chemical Physics*, 120(21) :9918–9924, 2004.

## BIBLIOGRAPHIE

- [133] H. P. Hratchian and H. B. Schlegel. Following reaction pathways using a damped classical trajectory algorithm. *The Journal of Physical Chemistry A*, 106(1) :165–169, 2002.
- [134] Isabelle Navizet, Ya-Jun Liu, Nicolas Ferré, Daniel Roca-Sanjuán, and Roland Lindh. The chemistry of bioluminescence : An analysis of chemical functionalities. *ChemPhysChem*, 12(17) :3064–3076, 2011.
- [135] Emil H. White, Eliezer Rapaport, Howard H. Seliger, and Thomas A. Hopkins. The chemi- and bioluminescence of firefly luciferin : An efficient chemical production of electronically excited states. *Bioorganic Chemistry*, 1(1) :92 – 122, 1971.
- [136] S. Fetzner. Oxygenases without requirement for cofactors or metal ions. *Applied Microbiology and Biotechnology*, 60(3) :243–257, Nov 2002.
- [137] Chun gang Min, Ai min Ren, Xiao na Li, Jing fu Guo, Lu yi Zou, Ying Sun, John D. Goddard, and Chia chung Sun. The formation and decomposition of firefly dioxetanone. *Chemical Physics Letters*, 506(4) :269 – 275, 2011.
- [138] Marlene DeLuca and Mary E. Dempsey. Mechanism of oxidation in firefly luminescence. *Biochemical and Biophysical Research Communications*, 40(1) :117 – 122, 1970.
- [139] Toru Nakatsu, Susumu Ichiyama, Jun Hiratake, Adrian Saldanha, Nobuyuki Kobashi, Kanzo Sakata, and Hiroaki Kato. Structural basis for the spectral difference in luciferase bioluminescence. *Nature*, 440(7082) :372–376, 03 2006.

## BIBLIOGRAPHIE



# **Annexes**



# Abréviations

Voici une liste des abréviations utilisées dans cette thèse :

ATP	Adénosine Tri-Phosphate
AMP	Adénosine Mono-Phosphate
BLI	BioLuminescence Imaging ou imagerie bioluminescente
BRET	Bioluminescence Resonance Energy Transfert
CASSCF	Complete Active Space Self Consistent Field
CASPT2	Complete Active Space Perturbation Theory second order
CI	Conical Intersection ou Intersection conique
CI(Jablonski)	Conversion interne
CIS(Jablonski)	Conversion intersystème
DFT	Density Functional Theory ou Théorie de la Fonctionnelle de la Densité
CGF	Contracted Gaussian Function ou Fonctions Gaussiennes Contractées
EE	Electrostatic Embedding
FC	Frank-Condon
GFP	Green Fluorescent Protein
GGA	Generalized Gradient Approximation
GS ou S0	Ground State ou état fondamental
HF	Hartree-Fock
HOMO	Highest Occupied Molecular Orbital ou plus Haute Orbitale Moléculaire Occupée
IC	Interaction de Configuration
<i>image</i>	image ou snapshot extrait d'une MD
iLH <sub>2</sub>	infraluciferin ou infraluciférine
IRC	Intrinsic Reaction Coordinate
KS	Kohn-Sham
LA	Link Atom ou Atome de Liaison
LCAO	Linear Combination of Atomic orbitals ou Combinaison Linéaire d'Orbitales Atomiques
LDA	Local Density Approximation
LH <sub>2</sub>	Luciferin ou luciférine
LQA	Local Quadratic Approximation
LSCF	Local Self Consistent Field ou champ autocohérent local

LUMO	Lowest Unoccupied Molecular Orbital ou Orbitale Moléculaire la plus Basse Vacante
MCSCF	Multiconfigurational SCF
MD	(Classical) Molecular Dynamics ou Dynamique moléculaire classique
ME	Mechanical Embedding
MM	Molecular Mechanics ou Mécanique Moléculaire
MS-CASPT2	Multi-State Complete Active Space Perturbation Theory second order
OxyLH <sub>2</sub>	Infraoxyluciferin ou infraoxyluciférine
OxyLH <sub>2</sub>	Oxyluciferin ou oxyluciférine
OM	Molecular Orbital ou Orbitale Moléculaire
PCM	Polarizable Continuum Model ou Modèle de Continuum Molarisable
PE	Polarizable Embedding
PMF	Potential Mean Force
QM	Quantum Mechanics ou Mécanique Quantique
QM/MM	méthode hybride couplant une partie QM et une partie MM
RMSD	Root Mean Square Deviation ou écart quadratique moyen
S1	First Singlet Excited State ou premier état excité singulet
SCF	Self Consistent Field ou champ autocohérent
SCRFF	Self Consistent Reaction Field ou champ de réaction autocohérent
STO	Slater Type Orbital ou Orbitale de Type Slatérienne
T <sub>e</sub>	Transition électronique entre S1 et S0
TS	Transition State ou état de transition
TD-DFT	Time Dependant DFT
vdW	van der Waals

# Liste des protéines

Le nombre de systèmes protéiques utilisés dans cette thèse étant importants, cette page regroupe une liste des protéines présentes dans cette thèse.

## Luciférases issues des lucioles :

Luciférase venant de *Photonus Pyralis*, la luciole nord-américaine :

- 4G3613.7.1 : protéine naturelle avec le domaine C-terminal fermé, assimilable à 2D1R, *Luciola cruciata*. Cette conformation représente le système à la fin de la réaction de bioluminescence. Elle est issue de la structure cristallographique 4G36.pdb<sup>2</sup>.
- 4G36-O213.7.1 : protéine dérivée de la 4G36 avec une molécule de dioxygène supplémentaire dans la protéine et le substrat Int présent dans la cavité13.1.
- 4G379.5.1 : protéine naturelle avec le domaine C-terminal bloqué dans une position semi-fermé. Cette conformation représente l'étape du mécanisme réactionnel où O<sub>2</sub> se coordine à la luciférine. Elle est issue de la structure cristallographique 4G37.pdb<sup>2</sup>.
- 4G37-S284T9.5.1 : protéine mutée dérivée de la 4G37 avec le résidu 284 muté de sérine à thréonine.
- 4G37-F250S10.2.3.2 : protéine mutée dérivée de la 4G37 avec le résidu 250 muté de phénylalanine à sérine.

Luciférase venant de *Luciola cruciata*, la luciole japonaise :

- 2D1R10.2.1 : protéine naturelle avec le domaine C-terminal fermé, assimilable à 4G36, *Photonus Pyralis*. Cette conformation représente le système à la fin de la réaction de bioluminescence. Elle est issue de la structure cristallographique 2D1R.pdb<sup>139</sup>. Pour plus d'information sur la modélisation de ce modèle voir la publication<sup>111</sup>.

Les deux structures cristallographiques suivantes sont issues de l'équipe de collaborateur de l'université de New York D'Abu Dhabi : Cesar A. C. Lopez, Wael Rabeh et Pance Naumov.

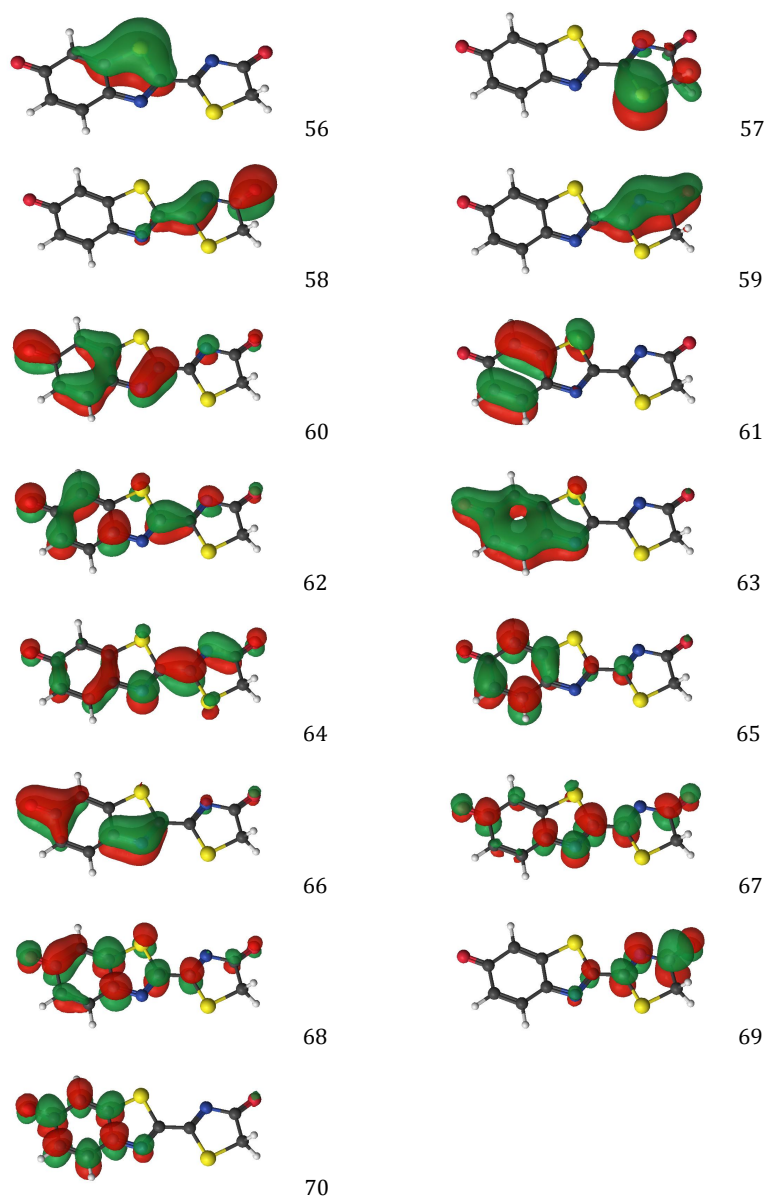
Luciférases issues de ver luisant *Amydetes viviani*<sup>17</sup> : GB<sub>Av</sub>.

- GB<sub>Av</sub>-A10.2.1 : protéine naturelle issu de la chaîne A et avec le domaine C-terminal ouvert.
- GB<sub>Av</sub>-B10.2.1 : protéine naturelle issu de la chaîne B et avec le domaine C-terminal ouvert.
- GB<sub>Av</sub>-B-closed\_w\_OxyLH<sub>2</sub>10.2.4 : protéine modélisée issu de la chaîne B avec le domaine C-terminal fermé lors d'une simulation d'umbrella sampling puis insertion de l'oxyluciférine et de l'AMPH dans la cavité.
- GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed10.2.4 : protéine modélisée issu de la chaîne B avec l'oxyluciférine et de l'AMPH dans la cavité. Puis le domaine C-terminal est fermé lors d'une simulation d'umbrella sampling.
- GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-R332L10.2.6.1 : protéine mutée et modélisée issu de GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed et avec le résidu 332 muté d'arginine à leucine.
- GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-I346L10.2.6.2 : protéine mutée et modélisée issu de GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed et avec le résidu 346 muté d'isoleucine à leucine.
- GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-insert-R351I10.2.6.3 : protéine mutée et modélisée issu de GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed et avec l'insertion d'une arginine à la position 351.

Luciférase issue du coléoptère *Phrixothrix hirtus* : RE<sub>Ph</sub>

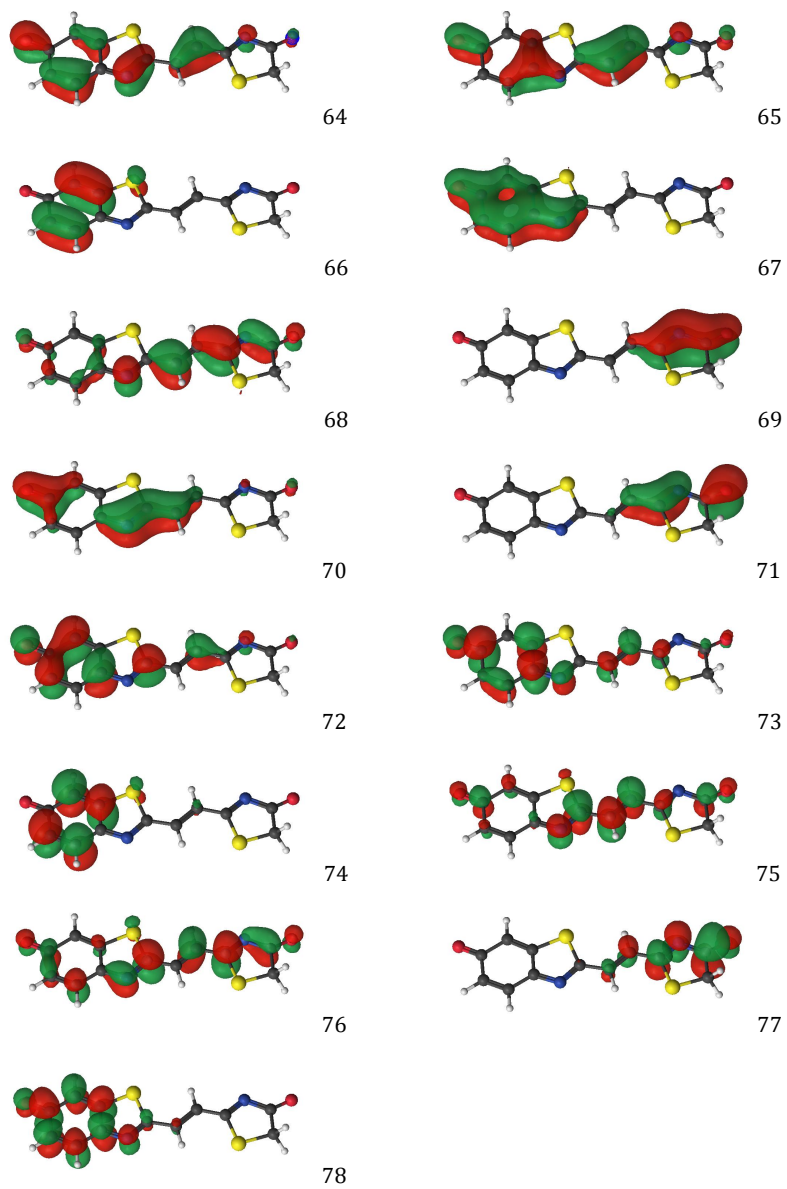
- RE<sub>Ph</sub>10.1 : protéine naturelle.

# Orbitales

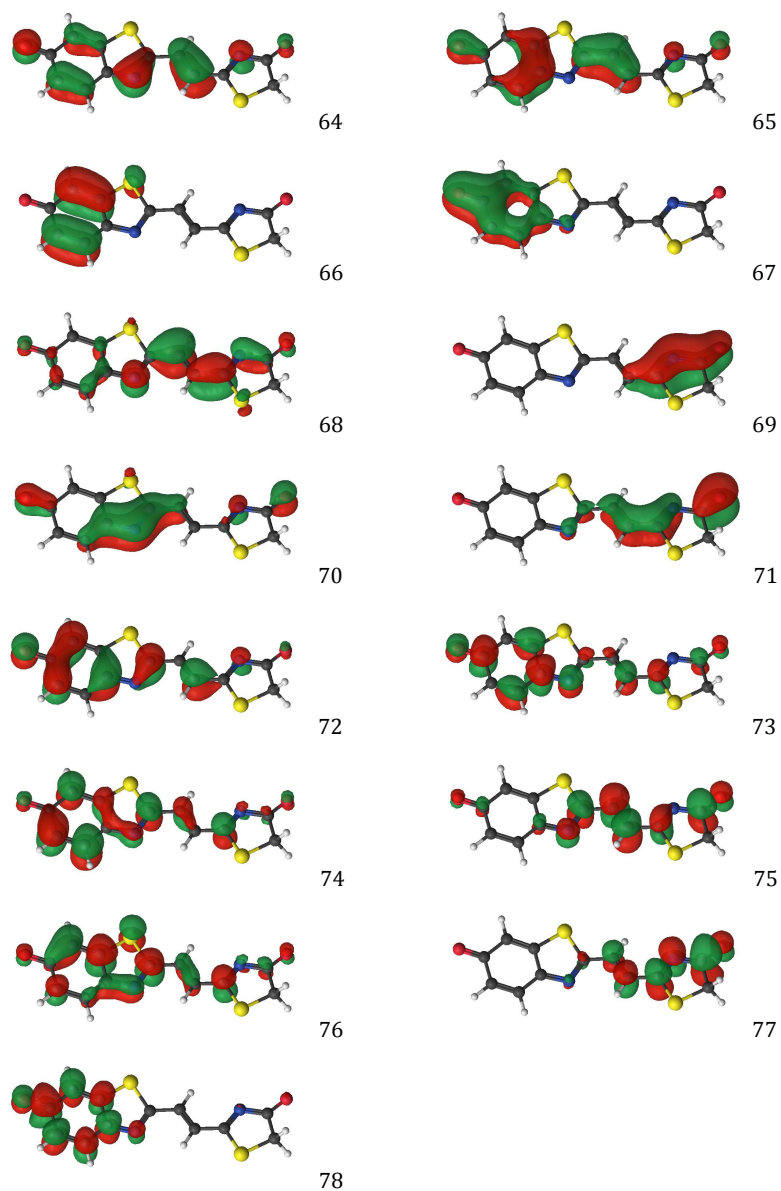


**Figure A1.** Les 15 orbitales actives sélectionnées en CASSCF pour OxyLH<sub>2</sub>, La HOMO correspond à la 62, la LUMO à la 64.



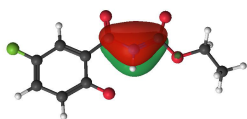


**Figure A2.** Les 15 orbitales actives sélectionnées en CASSCF pour OxyLH<sub>2</sub>-1, La HOMO correspond à la 72, la LUMO à la 68.

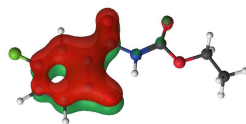


**Figure A3.** Les 15 orbitales actives sélectionnées en CASSCF pour OxyLH<sub>2</sub>-2, La HOMO correspond à la 72, la LUMO à la 68.

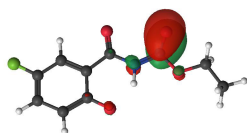
ANNEXES



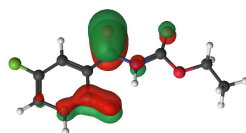
53 (2)



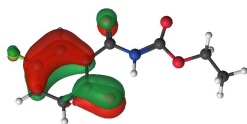
54 (1.97)



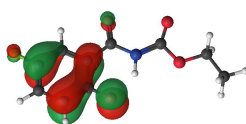
55 (1.96)



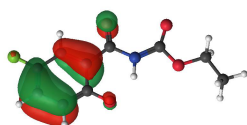
56 (1.92)



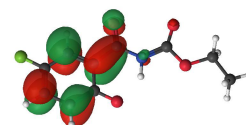
57 (1.95)



58 (1.36)

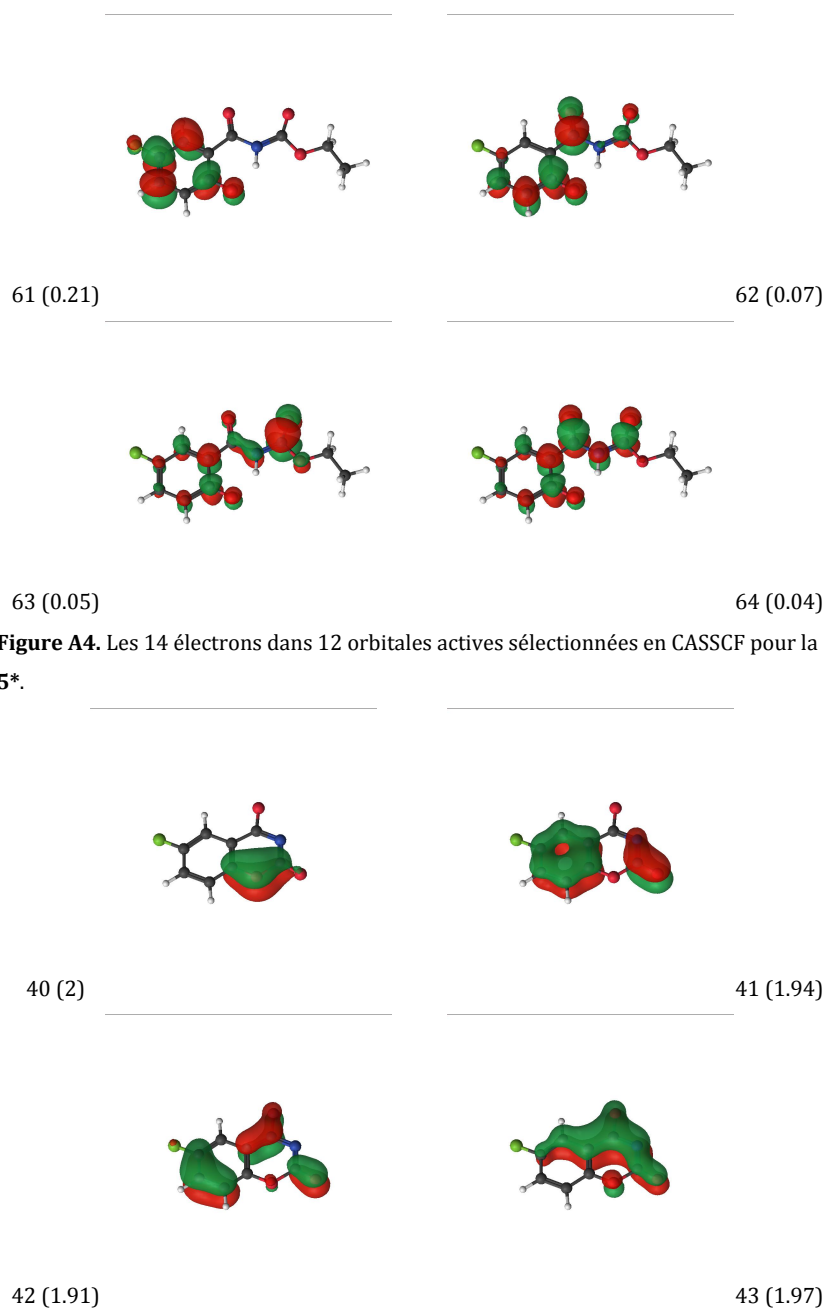


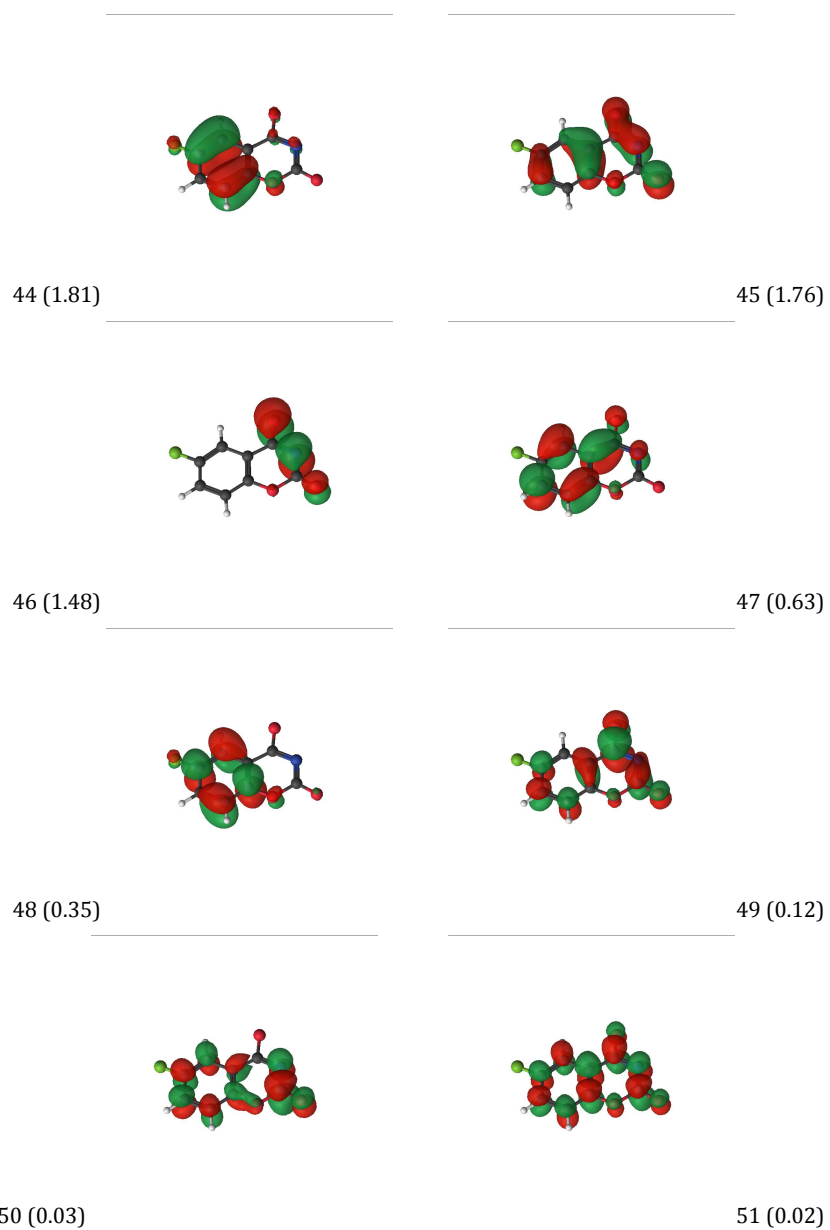
59 (1.79)



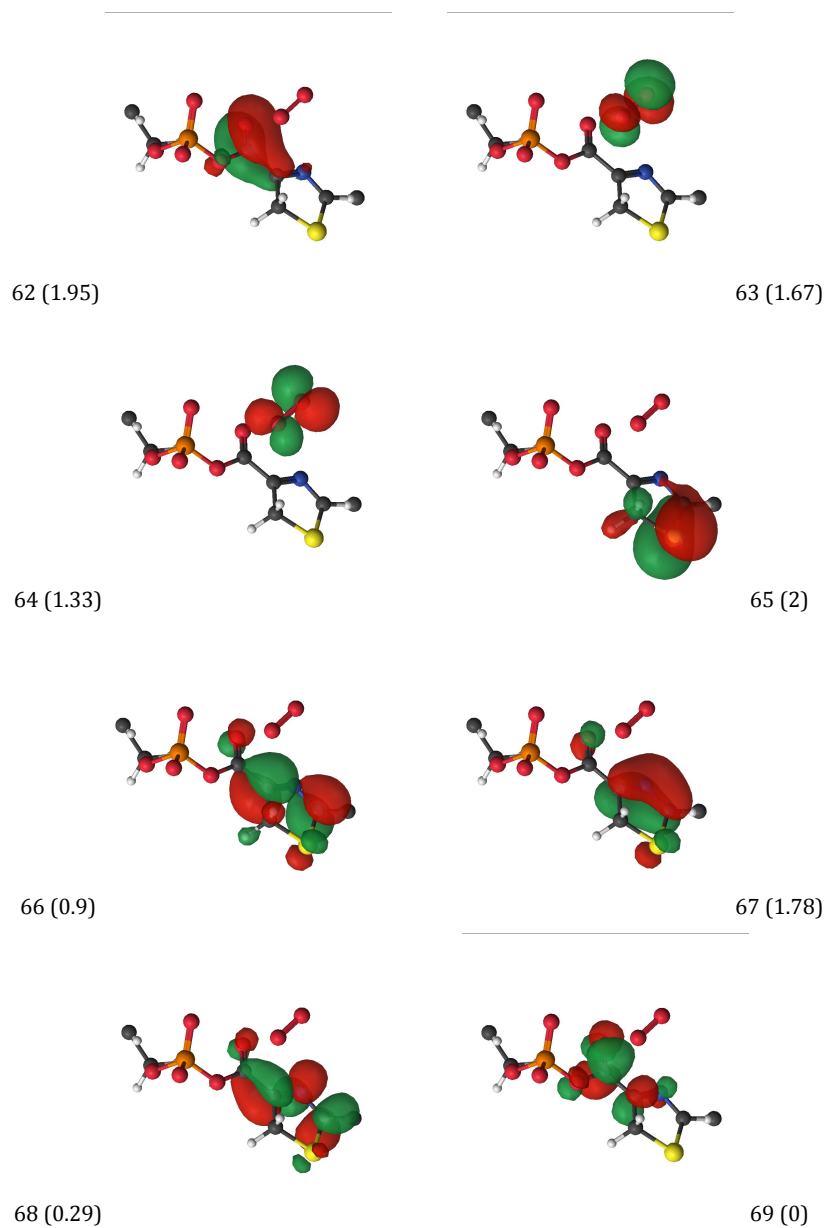
60 (0.7)

ANNEXES

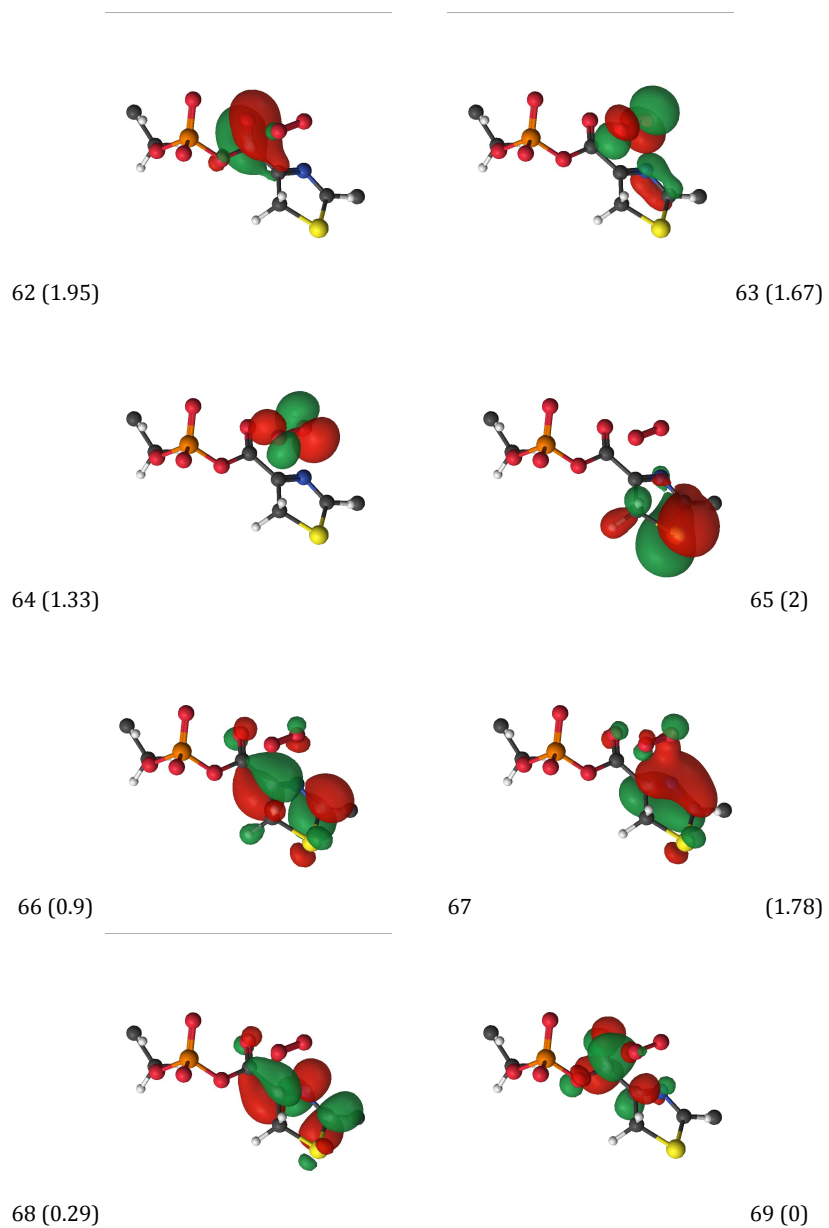




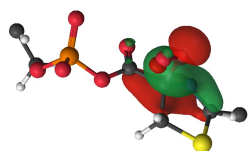
**Figure A5.** Les 14 électrons dans 12 orbitales actives sélectionnées en CASSCF pour la  $9^*$ .



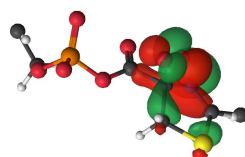
**Figure A6.** Les 8 orbitales actives sélectionnées en CASSCF pour la structure 4G36-O2-min.



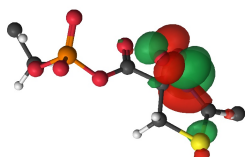
**Figure A7.** Les 8 orbitales actives sélectionnées en CASSCF pour la structure Int-O2 avec D-coord = 2.3 Å.



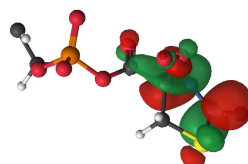
62 (1.95)



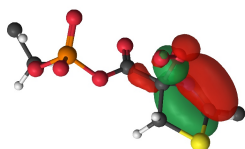
63 (1.67)



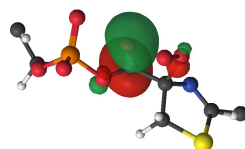
64 (1.67)



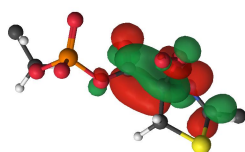
65 (0.67)



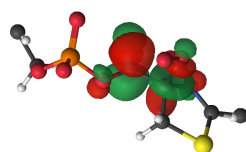
66 (1.91)



67 (1.94)



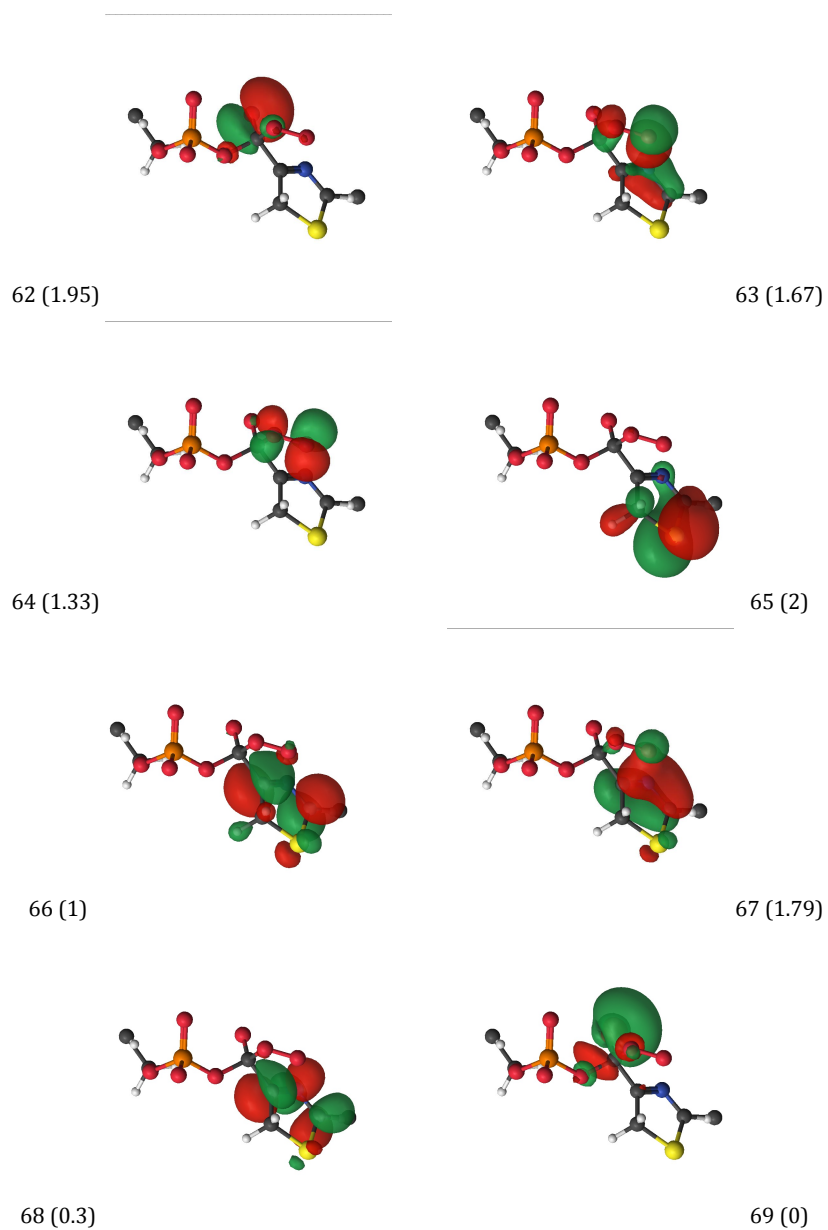
68 (0)



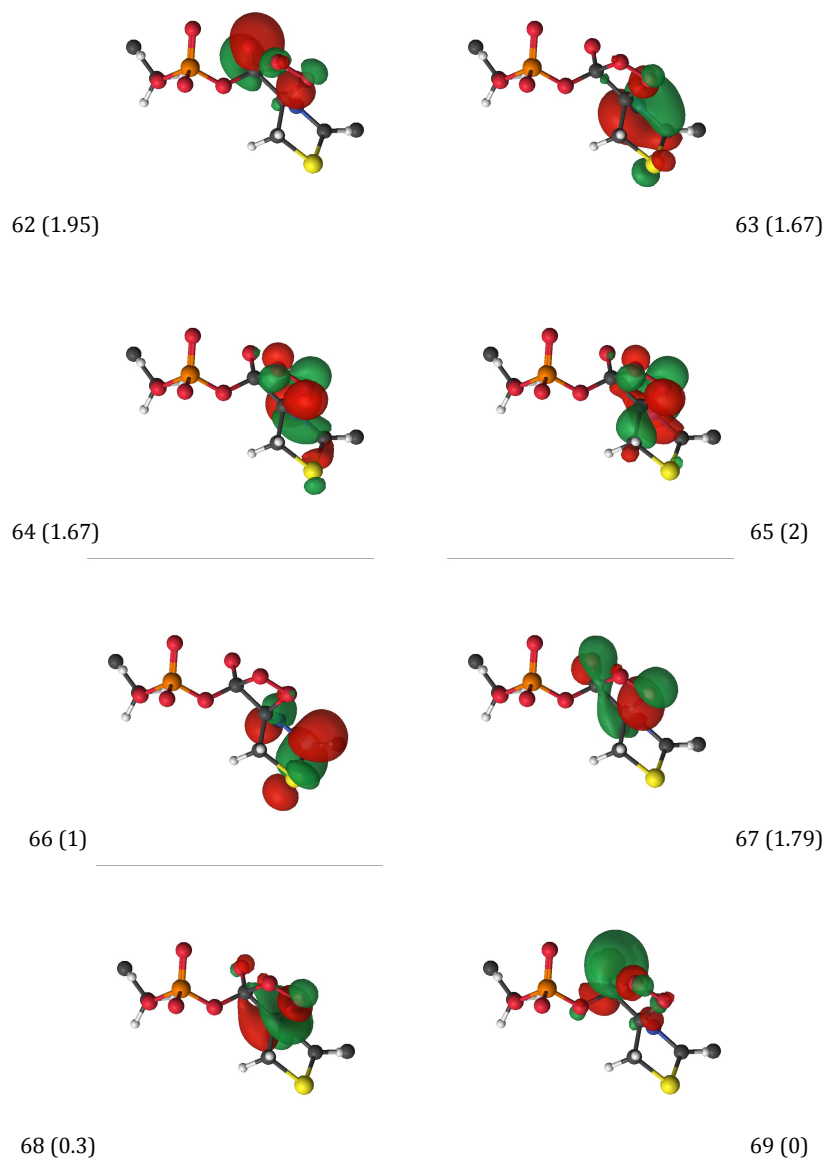
69 (0)

**Figure A8.** Les 8 orbitales actives sélectionnées en CASSCF pour la structure Int-s-O2 avec D-coord = 2.2 Å.

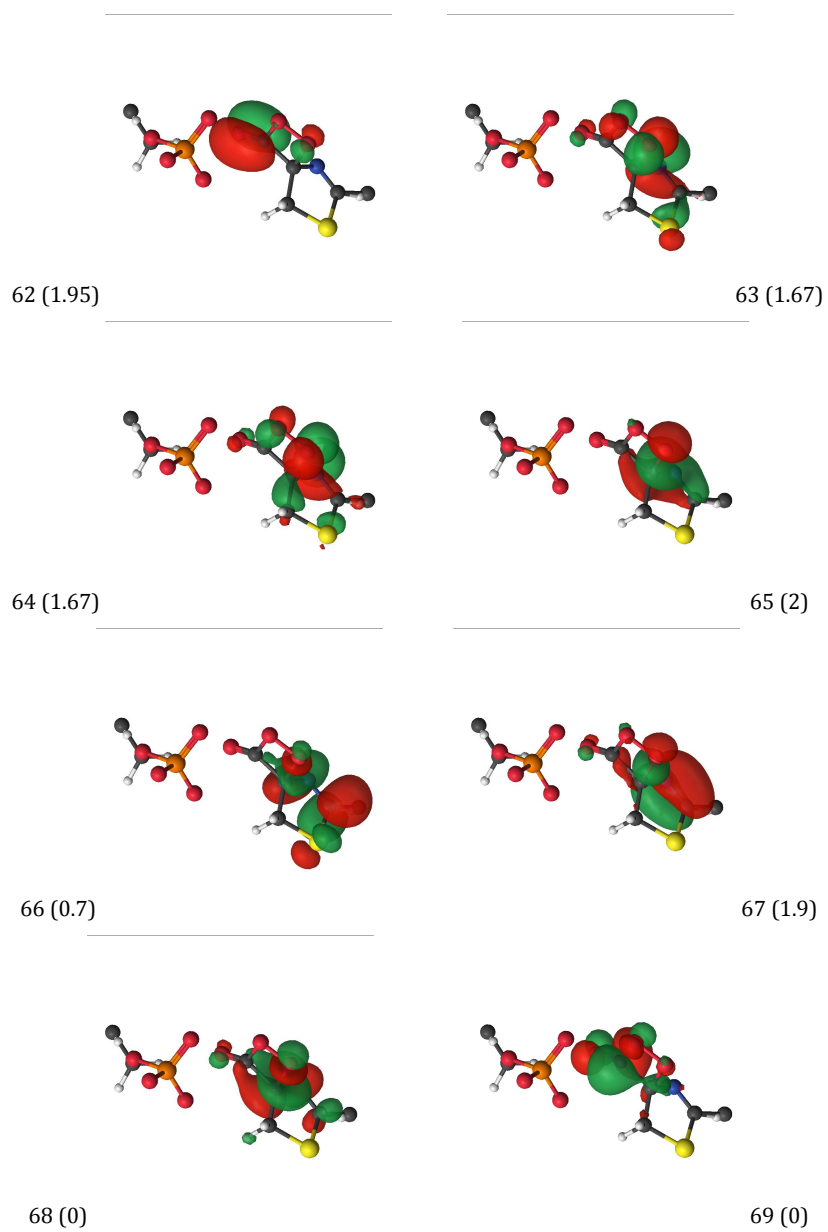




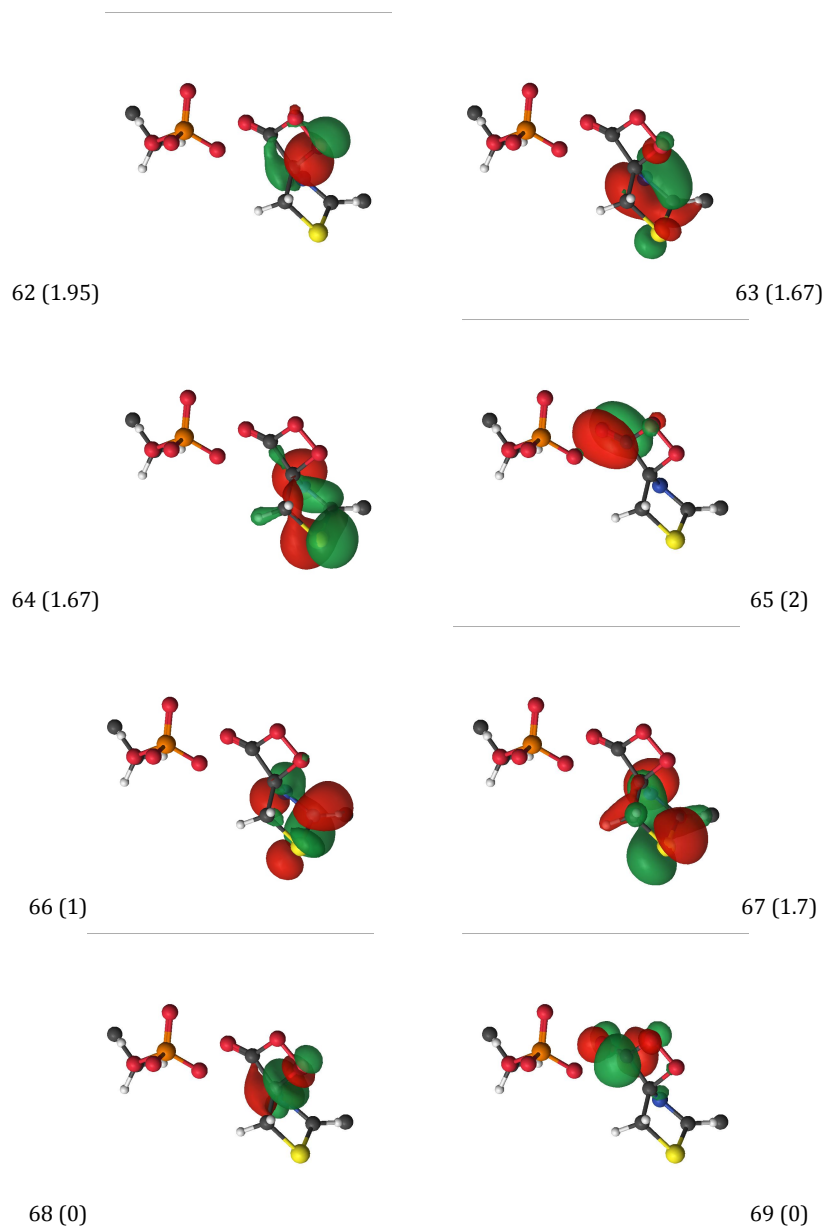
**Figure A9.** Les 8 orbitales actives sélectionnées en CASSCF pour la structure Int-t-O2 avec D-coord = 1.8 Å.



**Figure A10.** Les 8 orbitales actives sélectionnées en CASSCF pour la structure Int-t-O2 avec D-coord = 1.5 Å.



**Figure A11.** Les 8 orbitales actives sélectionnées en CASSCF pour la structure Int-B-s-dioxétanone.



**Figure A12.** Les 8 orbitales actives sélectionnées en CASSCF pour la structure Int-B-t-dioxétanone.

# **Publications**



Cite this: *Phys. Chem. Chem. Phys.*,  
2016, 18, 27460

Received 19th April 2016,  
Accepted 24th August 2016

DOI: 10.1039/c6cp02585d

www.rsc.org/pccp

## QM/MM calculations on a newly synthesised oxyluciferin substrate: new insights into the conformational effect†

Romain Berraud-Pache and Isabelle Navizet\*

In this publication we conduct calculations on a newly synthesised red-shifted emitter of luciferin in order to understand what are the main contributions to the colour-shifting emission. Indeed the bioluminescent system, especially from fireflies, is one of the main resources for medical imaging but its efficiency greatly 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 bioluminescence emission. We analysed the calculations and investigated the effects which play a non-negligible role in the emission, like the effect of the surroundings or the effect of the conformation of the emitter. Finally, in the absence of crystallographic structures, we proposed the most likely conformation for the emitter in the bioluminescence process.

### Introduction

For several years, modulation of the colour of 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 kinds 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. This 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 Fig. 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. Decomposition of this dioxetanone gives rise to an emissive molecule known as oxyluciferin (OxylH<sub>2</sub>) which releases energy by emitting light.<sup>16</sup>

In 2014, the group of Prof. 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 spectrum, the iLH<sub>2</sub> molecule leads to a strong red emission during the bioluminescence process. This range of emission in the red region of the visible spectrum allows for better penetration depths when used for medical imaging. By adding a double bond between the 2 cycles of 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

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

† 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<sub>c</sub> with the 6-31G(d,p) basis set. TD-DFT/MM results of T<sub>c</sub> with the 6-311G(2d,p) basis set for different functionals. TD-DFT/MM results of T<sub>c</sub> with the 6-311G(2d,p) basis set for different solvents. CAS orbitals of all emitters. MS-CASPT2/MM results of T<sub>c</sub> with 2 and 4 roots used in the calculation with an IPEA shift of 0.25. Comparison of MS-CASPT2/MM results of T<sub>c</sub> with 2 different IPEA shifts (0.25 and 0.1). Representation of the H-bond network in Ser<sub>284</sub>OxylH<sub>2</sub> before and after classical molecular dynamics. Graphical representation of the free energy of the system during the MD on Ser<sub>284</sub>OxylH<sub>2</sub>. Graphical representation of the dihedral angle N-CA-CB-OG in serine 284 during the MD of Ser<sub>284</sub>OxylH<sub>2</sub>. QM/MM results of T<sub>c</sub> for Ser<sub>284</sub>OxylH<sub>2</sub> before and after classical molecular dynamics. Representation of the H-bond network of the model Ser<sub>284</sub>OxylH<sub>2</sub>-2 and Ala<sub>284</sub>OxylH<sub>2</sub>-2. Comparison of the structures of 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<sub>c</sub> with the 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/c6cp02585d

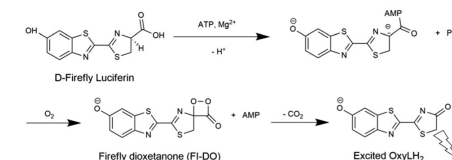


Fig. 1 Mechanism of firefly bioluminescence.

only one methyl group longer, emission for both the wild-type and the infra-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 American firefly and has been chemically engineered. It has been designed to reproduce the conformation of the protein when dioxygen binds to 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. Then the missing loop was added with the 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 Fig. S1, ESI†). The two SO<sub>4</sub><sup>-</sup> groups and the bridging residue XLX were removed.

The residues were protonated using Leap from the Amber14 suite of programs.<sup>21</sup> The contentious cases, especially for histidines, were resolved by computing their pK<sub>a</sub> with the H++ program.<sup>22</sup> We have chosen to double protonate 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 optimisation. 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 28 000 water molecules and 90 000 atoms in total. The AMBER99ff was used to model the residues of the protein. AMPH and the emitter (also called the 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 optimisation of the first singlet excited state, and frozen in its excited state conformation throughout the whole dynamic process. The system was heated from 100 K to 300 K in 20 ps. Then, under *NPT* conditions with *T* = 300 K and *P* = 1 atm, a 5 ns dynamic using periodic boundary conditions was realised with a 2 fs time step. During these simulations the 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 9 Å of the QM part. The microiterations technique<sup>30</sup> was used to converge the MM subsystem geometry for every QM minimisation 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) was first performed followed by calculation of the vertical difference of energies (*T<sub>v</sub>*) 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<sub>v</sub>* 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 follows. The first attempt was to try some TD-DFT calculations with the 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, ESI†). Thus as recommended in the literature,<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 ωB97XD<sup>38</sup> (see Table S2, ESI†). The chosen basis set was 6-311G(2d,p) as we did not use the basis sets with diffuse functions because these latter 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 0.02 Å.

For the rest of the paper, we have chosen to compare the 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 *in vacuo* and *in-solvent* calculations in a polar and non-polar solvent (see Table S3, ESI†). The solvent effects were taken into account using the Polarizable Continuum Model (PCM).<sup>39,40</sup>

TD-DFT results were compared with more precise calculations like multiconfigurational reference second-order perturbation theory. Complete-active-space second-order perturbation theory (CASPT2)<sup>41</sup> was also used to compute the *T<sub>v</sub>* 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 π-conjugated orbitals of OxyLH<sub>2</sub>. Then, 16-in-14 and a 14-in-13 active spaces were created by removing π conjugated orbitals centred on sulfurs. The 16-in-14 corresponds to the removal of the π sulfur centred orbital on the thiazolone ring while the 14-in-13 corresponds to the remove of both the π sulfur centred orbital on the thiazolone ring and on the benzothiazole ring. For OxyiLH<sub>2</sub>-1

and OxyiLH<sub>2</sub>-2 a 16-in-15 active space was chosen. This corresponds to all  $\pi$ -conjugated orbitals except those centred on both sulfurs due to computational limitations. The different active spaces are represented in Fig. S2 to S4 (ESI<sup>†</sup>). 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 shifts were tried, one at 0.25 and one at 0.1 (see Table S4 for the comparison between the 2 IPEA shifts and Table S5 for the results with 4 roots, ESI<sup>†</sup>).

In this paper we discuss several models. The first model designed represents a structure with the wild-type protein associated 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 the protein are available. Moreover, the experimenters<sup>17</sup> have only studied the stereoisomer with the double bond in the *trans* (*E*) position. We remained with the (*E*) stereoisomer in this study. Besides, there is not yet any experimental evidence of the chemical conformation of the stereoisomer (*E*) of OxyiLH<sub>2</sub>. Indeed, the surrounding protein prevents the substrate from moving through steric effects and thus from changing its conformation. Therefore each conformer has to be studied in details. One of the conformers seems, at first glance, to fit better into 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 Fig. 2).

The obtained model corresponding to the crystallographic form of the wild-type protein with oxyluciferin, as described above, is named Ser\_OxyLH<sub>2</sub> (substrate OxyLH<sub>2</sub> inside the protein with serine 284). For the other systems, modifications of either the emitter or the protein were made by hand. The model called Ser\_OxyiLH<sub>2</sub>-1 (Ser\_OxyiLH<sub>2</sub>-2) corresponds to the wild-type protein with one of the studied conformers of the OxyiLH<sub>2</sub> labelled OxyiLH<sub>2</sub>-1 emitter (OxyiLH<sub>2</sub>-2 emitter, respectively) (Fig. 2). The models called Thr\_OxyLH<sub>2</sub>, Thr\_OxyiLH<sub>2</sub>-1 and Thr\_OxyiLH<sub>2</sub>-2

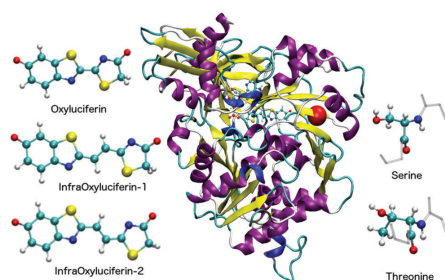


Fig. 2 Graphical representation of the different models used in this paper. On the left, the structures and the names of the 3 different substrates studied. Atoms are colour-coded as O = red, H = white, C = cyan, N = blue and S = yellow. On the right, the structures and the names of the wild type (serine)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

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

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

correspond to the mutated versions of the protein, S284T (serine mutated into threonine) with OxyLH<sub>2</sub>, OxyiLH<sub>2</sub>-1 or OxyiLH<sub>2</sub>-2, respectively. We have chosen to put the OH group of threonine in the direction of the emitter. The infraoxyluciferin structures were achieved by adding a double bond between the 2 cycles of the natural OxyLH<sub>2</sub>. Thus, the thiazolone ring was frozen and the remaining part of the molecule was added. All structures are described in Table 1 and summarised in Fig. 2.

## Results and discussion

For each emitter, OxyLH<sub>2</sub>, OxyiLH<sub>2</sub>-1 and OxyiLH<sub>2</sub>-2,  $T_e$  was calculated *in vacuo* and in-solvent. For each classical molecular dynamics (MD) simulation done on the models listed in 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 the protein.

### *In vacuo* 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 either *in vacuo* or 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 vacuo* with a RMSD lower than 0.02 Å. The augmentation of the polarity around the substrate leads to a smaller  $T_e$ . The values obtained taking into account the protein surrounding, with the QM/MM method (Table 3, ESI<sup>†</sup>) are closer to the ones in chloroform than in water or *in vacuo*.

### QM/MM

The utilisation of QM/MM calculations on the structures obtained after MD gives us good insight into 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 in 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).



## Paper

**Table 2** TD-DFT/MM for electronic transition between S1 and S0 ( $T_e$ ) with the 6-311G(2d,p) basis set and associated oscillator strength on the 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) $f = 0.47$	
<i>In vacuo</i> OxyLH <sub>2</sub>	2.52 (482) $f = 0.67$	2.27 (547) $f = 0.41$	
<i>In vacuo</i> OxyiLH <sub>2</sub> -1	2.36 (525) $f = 1.14$	2.13 (580) $f = 0.76$	
<i>In vacuo</i> OxyiLH <sub>2</sub> -2	2.26 (547) $f = 1.02$	2.03 (608) $f = 0.65$	

**Table 3** MS-CASPT2/MM for the electronic transition between S1 and S0 ( $T_e$ ) with the 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	2.24 (554)	2.22 (558)
	16-in-14	2.14 (579)	
Ser_OxyiLH <sub>2</sub> -1	16-in-15	2.19 (566)	1.85 (670)
	16-in-15	1.92 (645)	
Thr_OxyLH <sub>2</sub>	14-in-13	2.07 (598)	2.04 (605)
	16-in-14	2.15 (577)	
Thr_OxyiLH <sub>2</sub> -1	16-in-15	2.15 (578)	1.75 (706)
	16-in-15	1.93 (643)	
<i>In vacuo</i> OxyLH <sub>2</sub>	14-in-13	2.33 (532)	
	16-in-14	2.25 (550)	
<i>In vacuo</i> OxyiLH <sub>2</sub> -1	16-in-15	2.02 (614)	
	16-in-15	1.93 (642)	

MS-CASPT2/MM calculations were also computed and the obtained results are similar to that 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 (the default value 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, ESI†). Thus the  $\Delta T_e$  for the IPEA shift set at 0.25 ranges between 0.1 and 0.6 eV (see Table S5, ESI†). Other parameters can also influence the  $T_e$ , such as the size of the active space (see Table S5, ESI†) and the basis set.

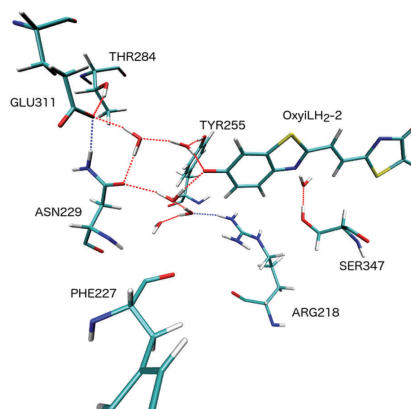
When the  $\pi$  conjugated system in the emitter is increased *i.e.* from OxyLH<sub>2</sub> to OxyiLH<sub>2</sub> the calculated emission of energy  $T_e$  decreases leading to red-shifting of the emission, as seen in the experimental results. Besides, for both emitters the emission is accompanied with 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 ring (see Fig. 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 value than the one from the OxyiLH<sub>2</sub>-1 structure. For example, Ser\_OxyiLH<sub>2</sub>-1 has a B3LYP  $T_e$  of 1.98 eV while 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, for the rest of this paper both structures and properties will be discussed with this level of theory.

### Classical molecular dynamics (MD)

During the MD, we were able to follow several elements involved in the interactions between the substrate and the environment (protein residues and water molecules). First, equilibration is attained after 1 ns for all MD simulations (see Fig. S6, ESI†). 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 in its initial S1 conformation. However, AMPH was not restrained but thanks to strong hydrogen bonds and steric effects the molecule does not move a lot (RMSD < 0.5 Å) in all the MD simulations. The snapshot used for the QM/MM calculations corresponds to residue 284 turned toward the cavity. A longer MD simulation shows that the dihedral angle N-CA-CB-OG is stable in this direction during the 5 first ns and returns to the same value after 8.5 ns (see Fig. S7, ESI†). In the cavity, an H-bond network interacts with the emitter and especially with the phenolate oxygen (O<sub>1</sub>) (see Fig. 3, 5 and Fig. S5, ESI†). The structures obtained after MD show that an H-bond network is created between the emitter and several residues especially residue 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 network of water molecules has already been observed in our previous publications.<sup>9</sup>



**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.

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

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

The effect of the creation of this H-bond network during the simulation is characterised by a blue-shift in the emission for all levels of theory computed (see Table S6, ESI<sup>†</sup>). In comparison, the PCM calculations in water give a red-shift emission compared to those carried out *in vacuo* (see Table S3, ESI<sup>†</sup>). However, in the PCM model, the solvent water is modelled 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 the HOMO and LUMO in Fig. 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 spectrum. This blue-shift induced by the presence of a water molecule near the phenolate group has already been 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 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 consistently as the red shift observed experimentally. 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 interactions with O<sub>1</sub> due to the mutation leads to destabilisation of the HOMO, resulting in a red shift.

For OxyiLH<sub>2</sub>, the experimental results also show a red shift with the mutation. However QM/MM energies  $T_c$  show that the mutation does not significantly change 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  $T_c$  is slightly blue shifted compared to for Ser\_OxyiLH<sub>2</sub>-2, which is the opposite of the experimental observation. Furthermore, to get more insight into the H-bond network we have mutated residue 284 from serine to alanine. Unlike the previous mutation, the mutation S284A induces a modification of the surroundings. Alanine is a non-polar residue, thus it can not be involved in an H-bond network (see Fig. S8, ESI<sup>†</sup>). The TD-DFT/MM B3LYP  $T_c$  (1.89 eV, 655 nm) obtained for Ala\_OxyiLH<sub>2</sub>-2 (*i.e.* OxyiLH<sub>2</sub>-2 with S284A luciferase) is similar to both  $T_c$  values obtained for 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 simulation. The use of many snapshots<sup>45–47</sup> may have given more precise information but are time-consuming when dealing with emission.

### Decomposition the contributions from the surroundings

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

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

We first looked at geometric constraints 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 collected some geometric parameters for the OxyiLH<sub>2</sub> substrates, optimised *in vacuo* or inside the mutated protein (see Table 4). There are no major differences on the C–O bond lengths and therefore these factors do not contribute. *In vacuo*, 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 angle between C=C and the thiazolone moiety ( $\beta_{11-8}$ ) (see Fig. 5). The twisting in OxyiLH<sub>2</sub>-1 is comparable with that in OxyLH<sub>2</sub> for both  $\beta$  values while for OxyiLH<sub>2</sub>-2 the twisting is located around  $\beta_{1-12}$  and has a higher value of 9° (see Table 4). Furthermore, Fig. 6 shows superposition of the optimised OxyiLH<sub>2</sub> structures. When OxyiLH<sub>2</sub>-1 shows a distortion of the benzothiazole plane in

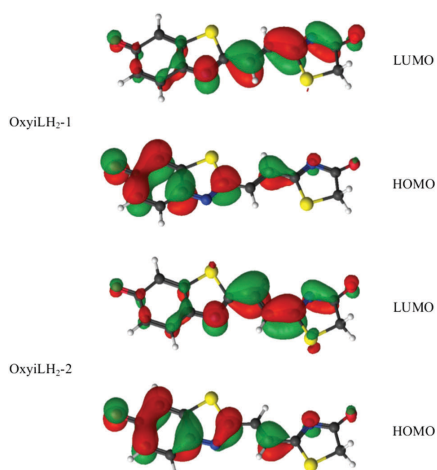
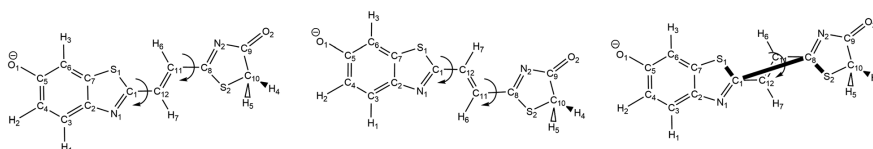


Fig. 4 Frontier orbitals of OxyiLH<sub>2</sub>-1 (top) and OxyiLH<sub>2</sub>-2 (bottom) obtained from CASSCF calculations of the emitter in the protein (displayed using the LUSCUS program<sup>48</sup>).

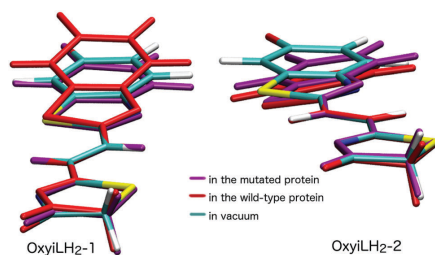
**Table 4** Geometric parameters of the OxyiLH<sub>2</sub> structures optimised inside the protein or *in vacuo* (optimisation with the B3LYP functional and 6-311G(2d,p) basis set). Single point calculations of the *in vacuo* energy  $T_e$  on these structures (B3LYP functional and 6-311G(2d,p) basis 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 <i>in vacuo</i>	OxyiLH <sub>2</sub> -2 optimised in wild protein	OxyiLH <sub>2</sub> -2 optimised in mutated protein	OxyiLH <sub>2</sub> -2 optimised <i>in vacuo</i>
C <sub>5</sub> -O <sub>1</sub> <sup>a</sup>	1.25	1.26	1.25	1.25	1.26	1.25
C <sub>9</sub> -O <sub>2</sub> <sup>a</sup>	1.24	1.24	1.22	1.22	1.23	1.22
$\beta_{1-12}$ <sup>b</sup>	1.52 <sup>c</sup>	2.84 <sup>c</sup>	0 <sup>e</sup>	-8.98 <sup>d</sup>	-8.62 <sup>d</sup>	0 <sup>d</sup>
$\beta_{11-8}$ <sup>b</sup>	-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>
$\beta_{2\text{ cycles}}$ <sup>b,g</sup>	0.53	2.14	0	20.68	18.17	0
$T_e^{\text{vac}}$ <sup>h</sup>	2.16	2.13	2.13	2.00	1.98	2.03

<sup>a</sup> Bond lengths in Å. <sup>b</sup> Angles in degrees. <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 calculations of  $T_e$  *in vacuo* in eV (to be compared with values 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  $\beta_{2\text{ cycles}}$ .



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

the wild-type protein, OxyiLH<sub>2</sub>-2 shows a rotation around the C1-C12 bond. Therefore, the torsion between the 2 cycles called  $\beta_{2\text{ cycles}}$  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 OxyiLH<sub>2</sub> is not as visible (Fig. S9, ESI†).

The constraint from 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 calculations of the  $T_e$  *in vacuo* ( $T_e^{\text{vac}}$ ) on both the *in vacuo* 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 follows: 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 greatest torsion is around C1-C12 at the edge of the electronic density of the LUMO which destabilizes it.

The electrostatic effect can be quantified by comparing two calculations: the QM/MM calculated  $T_e$  on the optimised structure taking into account the protein environment, and the single-point QM calculated  $T_e^{\text{vac}}$  of 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 for both proteins.

The results show that the presence of the protein environment reduces the gap between the ground and the first excited states of the emitter. For example, the electrostatic effect in Thr\_OxyiLH<sub>2</sub>-2 is obtained by calculating the difference between  $T_e^{\text{vac}}$  (1.98 eV, in Table 4) and  $T_e$  in the 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 in the wavelength by 44 nm).

In summary, the effect of the surroundings 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 OxyiLH<sub>2</sub>. The emission wavelength of the conformer OxyiLH<sub>2</sub>-2 is closer to the experimental value than the one of OxyiLH<sub>2</sub>-1 conformer. The *in vacuo* 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 difference of emission energies is increased further by 0.05 eV. Thus,

the difference in emission energy between these two conformers comes mainly from the nature of the conformer itself. Between these two studied conformers, 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 in the cavity of the protein. Some calculations were performed on these Z conformers and results are collected in the ESI† (see Fig. S10 and Table S7).

## Conclusion

In this paper we give theoretical insights on the red-emission of a modified firefly emitter. MD simulations followed by QM/MM calculations give fluorescence emissions which are close to experimental bioluminescence emissions, for both the wild-type and the modified emitters. We also investigate the influence of the surrounding protein on the emission, and thus find that two factors are mainly responsible for the colour-tuning. The presence of an H-bond network in the cavity of the emitter leads to a blue-shift in 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 the experimental values. In the end, without experimental evidence or a 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 into new questions about the modified light emitter iLH<sub>2</sub>: what 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 for experimental studies. For example, the mutation of polar (Glu 311 and Asn 229) and non-polar (Ile 351) residues present in the cavity could be explored both experimentally and theoretically in order to gain a better understanding of the H-bonding network. The presence of several possible emitters makes iLH<sub>2</sub> an interesting substrate from both an experimental and theoretical point of view in order to understand the colour-tuning of bioluminescence.

## Acknowledgements

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

## References

- 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.
- Y.-Y. Cheng, J. Zhu and Y.-J. Liu, *Chem. Phys. Lett.*, 2014, **591**, 156–160.
- 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.
- X. Q. Ran, X. Zhou and J. D. Goddard, *ChemPhysChem*, 2015, **16**, 396–402.
- L. P. da Silva and J. C. E. da Silva, *J. Chem. Theory Comput.*, 2011, **7**, 809–817.
- R. C. Steinhardt, J. M. O'Neill, C. M. Rathbun, D. C. McCutcheon, M. A. Paley and J. A. Prescher, *Chem. – Eur. J.*, 2016, **221**, 3671–3675.
- C. Miura, M. Kiyama, S. Iwano, K. Ito, R. Obata, T. Hirano, S. Maki and H. Niwa, *Tetrahedron*, 2013, **69**, 9726–9734.
- M. I. Koksharov and N. N. Ugarova, *Photochem. Photobiol. Sci.*, 2013, **12**, 2016–2027.
- I. Navizet, Y.-J. Liu, N. Ferré, H.-Y. Xiao, W.-H. Fang and R. Lindh, *J. Am. Chem. Soc.*, 2010, **132**, 706–712.
- 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.
- T. Hirano, Y. Hasumi, K. Ohtsuka, S. Maki, H. Niwa, M. Yamaji and D. Hashizume, *J. Am. Chem. Soc.*, 2009, **131**, 2385–2396.
- L. P. da Silva and J. C. E. da Silva, *J. Chem. Theory Comput.*, 2011, **7**, 809–817.
- S. Hosseinkhani, *Cell. Mol. Life Sci.*, 2010, **68**, 1167–1182.
- Y.-Y. Cheng and Y.-J. Liu, *J. Chem. Theory Comput.*, 2015, **11**, 5360–5370.
- S.-F. Chen, Y.-J. Liu, I. Navizet, N. Ferré, W.-H. Fang and R. Lindh, *J. Chem. Theory Comput.*, 2011, **7**, 798–803.
- I. Navizet, Y.-J. Liu, N. Ferré, D. Roca-Sanjuán and R. Lindh, *ChemPhysChem*, 2011, **12**, 3064–3076.
- A. P. Jathoul, H. Grounds, J. C. Anderson and M. A. Pule, *Angew. Chem., Int. Ed.*, 2014, **53**, 13059–13063.
- J. A. Sundlov, D. M. Fontaine, T. L. Southworth, B. R. Branchini and A. M. Gulick, *Biochemistry*, 2012, **51**, 6493–6495.
- B. R. Branchini, J. C. Rosenberg, D. M. Fontaine, T. L. Southworth, C. E. Behney and L. Uzasci, *J. Am. Chem. Soc.*, 2011, **133**, 11088–11091.
- K. Tang, J. Zhang and J. Liang, *PLoS Comput. Biol.*, 2014, **10**, e1003539.
- D. A. Case, *et al.*, *AMBER 14*, University of California, San Francisco, CA, 2015.
- R. Anandkrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537–W541.
- I. Navizet, D. Roca-Sanjuán, L. Yue, Y.-J. Liu, N. Ferré and R. Lindh, *Photochem. Photobiol.*, 2013, **89**, 319–325.
- H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

- 25 J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 26 N. Ferré and J. G. Ángyán, *Chem. Phys. Lett.*, 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, *Gaussian 09, Revision D.01*, Gaussian, Inc., Wallingford CT, 2009.
- 28 J. W. Ponder, *TINKER, Software Tools for Molecular Design, version 6.3*, Department of Biochemistry and Molecular Biophysics, Washington, University School of Medicine, St. Louis, MO, 2004.
- 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, *J. Comput. Chem.*, 2010, **31**, 224–247.
- 30 F. Melaccio, M. Olivucci, R. Lindh and N. Ferré, *Int. J. Quantum Chem.*, 2011, **111**, 3339–3346.
- 31 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
- 32 S. Chen, I. Navizet, R. Lindh, Y. Liu and N. Ferré, *J. Phys. Chem. B*, 2014, **118**, 2896–2903.
- 33 A. D. Laurent, A. Blondel and D. Jacquemin, *Theor. Chem. Acc.*, 2015, **134**, 1–11.
- 34 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 35 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.
- 36 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
- 37 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.
- 38 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- 39 B. Mennucci, E. Cancès and J. Tomasi, *J. Phys. Chem. B*, 1997, **101**, 10506–10517.
- 40 E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **107**, 3032–3041.
- 41 K. Andersson, P. Malmqvist and B. O. Roos, *J. Chem. Phys.*, 1992, **96**, 1218–1226.
- 42 B. O. Roos, R. Lindh, P.-A. Malmqvist, V. Veryazov and P.-O. Widmark, *J. Phys. Chem. A*, 2004, **108**, 2851–2858.
- 43 J. P. Zobel, J. J. Nogueira and L. González, 2016, in preparation.
- 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, *J. Am. Chem. Soc.*, 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, *J. Phys. Chem. B*, 2013, **117**, 2069–2080.
- 47 N. A. Murugan, J. M. H. Olsen, J. Kongsted, Z. Rinkevicius, K. Aidas and H. Ågren, *J. Phys. Chem. Lett.*, 2013, **4**, 70–77.
- 48 G. Kovačević and V. Veryazov, *J. Cheminf.*, 2015, **7**, 1–10.

## Chemiluminescence

# The Light Emitter of the 2-Coumaranone Chemiluminescence: Theoretical and Experimental Elucidation of a Possible Model for Bioluminescent Systems

Stefan Schramm,<sup>[a]</sup> Isabelle Navizet,<sup>\*[b]</sup> Panče Naumov,<sup>\*[c]</sup> Naba K. Nath,<sup>[c]</sup>  
 Romain Berraud-Pache,<sup>[b]</sup> Pascal Oesau,<sup>[a]</sup> Dieter Weiss,<sup>\*[a]</sup> and Rainer Beckert<sup>\*[a]</sup>

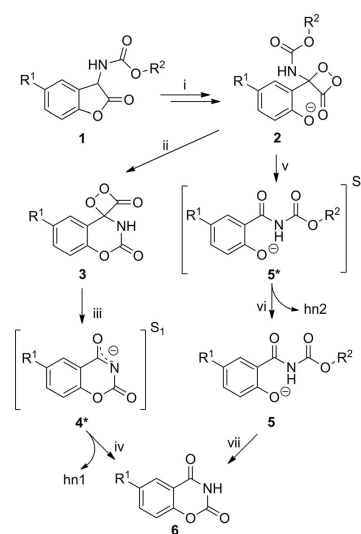
**Abstract:** The bioluminescence reaction of many biological subspecies, most notably fireflies, forms dioxetanone derivatives as high-energy intermediates. The thermal instability of dioxetanones complicates understanding of the transition from the ground state to the first excited state that leads to light emission. Herein, we report the reaction mechanism of 2-coumaranones, synthetically accessible strongly chemiluminescent materials that mimic the bioluminescence reaction. The pathways of chemiexcitation and photorelaxation are clarified on the basis of synthetic evidence and spectrochemical as well as computational mechanistic analysis.

aranones, synthetically accessible strongly chemiluminescent materials that mimic the bioluminescence reaction. The pathways of chemiexcitation and photorelaxation are clarified on the basis of synthetic evidence and spectrochemical as well as computational mechanistic analysis.

## Introduction

Chemiluminescence (CL) and bioluminescence (BL) are photochemical phenomena that lead to emission of light in the visible region of the electromagnetic spectrum by following a sequence of chemical reactions. In the case of BL, this excitation by chemical transformation occurs inside enzymes (luciferases). In CL, the reaction normally takes place in solution. The formation of a 1,2-dioxetanone moiety is common with many of the known CL and BL reactions.<sup>[1,2]</sup> This chemical group is formed if molecular oxygen reacts with the substrate, which lead to the formation of a peroxide bond (see compound **2** in Scheme 1). Decomposition of this intermediate leads to a luminophore (oxyluciferin in BL or light emitter) in its first singlet excited state ( $S_1$ ) (chemiexcitation), which relaxes by emission of light (photorelaxation).<sup>[3–5]</sup> Despite the increase in the use of BL and CL as probes, the comprehension of these systems remains incomplete. Indeed, the difficulty of getting RX structures from biological systems limits the number of available BL luminophores, whereas many of them are deemed very unstable in pure form, and their structures remain uncertain or are unknown. For in-

stance, although the BL of the Siberian glowworm (*Fridericia heliota*) and some fungal species have been known for centu-



**Scheme 1.** Two suggested pathways for the decomposition of 1,2-dioxetanone intermediate **2** in the chemiluminescence of 2-coumaranone derivatives **1**.  $R^1$  is a halogen and  $R^2$  is an alkyl or aryl group (examples are described in ref.<sup>[10]</sup>). Both closed-ring lactone **4\*** and salicylamide-like structure **5\*** are viable candidates for the excited-state structure of the light-emitting species. In one path (ii,iii,iv) the cyclization occurs before dioxetanone decomposition, and in the other path (v,vi,vii) the cyclization occurs after dioxetanone decomposition and light emission.

[a] Friedrich-Schiller-Universität Jena, Institut für Organische Chemie und Makromolekulare Chemie, Humboldtstr. 10, 07743 Jena, Germany  
 E-mail: dieter.weiss@uni-jena.de  
<http://www.agbeckert.uni-jena.de/>

[b] Université Paris-Est, Laboratoire Modélisation et Simulation, Multi Echelle, MSME UMR 8208 CNRS, 5 bd Descartes, 77454 Marne-la-Vallée, France  
 E-mail: isabelle.navizet@u-pem.fr

[c] New York University Abu Dhabi, P. O. Box 129188, Abu Dhabi, United Arab Emirates  
 E-mail: pance.naumov@nyu.edu  
[nyuad.nyu.edu/naumov-group](mailto:nyuad.nyu.edu/naumov-group)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/ejoc.201501515>.

ries, the structures of their substrates were characterized only recently.<sup>[6–8]</sup> The protocols for cases in which synthetic procedures are available<sup>[9,10]</sup> are laborious and cost ineffective, which greatly limits prospects for systematic studies of the related CL and BL mechanisms. The recent interest in new fluorescent probes<sup>[11–13]</sup> calls for the design and convenient access to diverse, chemically robust, stable, bright luminophores with tunable emission energies. All properties are crucial for their utility as labels in bioanalytical applications, typically with microscopic imaging techniques.

## Results and Discussion

To fulfill the demand for chemically variable luminophores, we recently prepared new derivatives of 2-coumaranones. This class of chemiluminescent compounds (**1** Scheme 1) first synthesized in 1979<sup>[14–18]</sup> are easily accessible and are strongly chemiluminescent.<sup>[19–21]</sup> The structures of the ethyl and methyl derivatives were confirmed by X-ray diffraction analysis (Figure 1; for details see the Supporting Information, Figures S6–S9). Upon treatment with base [i.e., 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU)] in polar aprotic solvents (e.g., MeCN, DMF, or acetone), these compounds emit bright blue light that is readily observable even with the naked eye in daylight (Figure 2, a). The CL also occurs after oxidation with a minimum amount of horse radish peroxidase (Figure 2, b). The energy released in this reaction can be further transferred to a fluorescence dye; the color of the secondary emission can be tuned across a wide range of the visible spectrum by selecting the dye (Figure 2, c–n). The 2-coumaranones own a 1,2-dioxetanone structure as an intermediate (see compound **2** in Scheme 1)<sup>[21]</sup> in common with many of the known CL and BL reactions.<sup>[1,2,22]</sup> Thus, they can be convenient analogues of BL precursors that provide instant access to this unstable high-energy intermediate, the structure of which has not yet been directly observed. Herein, the CL mechanism and the structure of the light emitter of 2-coumaranones were delineated to obtain fundamental insight into the molecular mechanism of their light emission. This required a combination of quantum chemical calculations and spectrochemical and structural analysis. The reaction pathways that have been advanced for CL are shown in Scheme 1. Until now, studies on these compounds have not distinguished between these two pathways.<sup>[14,19,20]</sup> Reaction of **1** with molecular oxygen (path i) initially leads to high-energy 1,2-dioxetanone intermediate **2**. Decomposition of **2** can occur by cyclization (path ii), whereby spirocyclic 1,2-dioxetanone **3** is generated. Spirocyclic **3** decomposes by releasing CO<sub>2</sub> (path iii) to afford product lactone **4** in its excited state (**4\***). After light emission (path iv), the lactone relaxes to its ground state (GS) **6**. As an

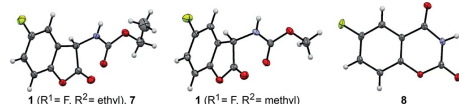


Figure 1. Crystal structures of compound **7** in Scheme 2 (**1** with R<sup>1</sup> = F and R<sup>2</sup> = Et in Scheme 1), **1** (with R<sup>1</sup> = F and R<sup>2</sup> = Me in Scheme 1) and **8**.

alternative pathway, 1,2-dioxetanone **2** can decompose directly by releasing CO<sub>2</sub> (path v) to afford the excited state (i.e., **5\***) of compound **5**, which subsequently relaxes and cyclizes (path vii) to product **6**.

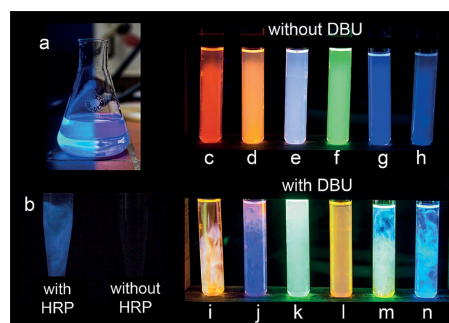
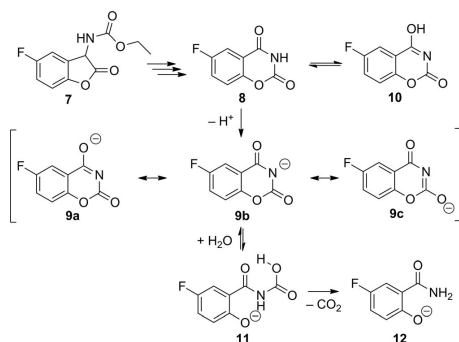


Figure 2. Chemiluminescence and fluorescence of 2-coumaranones. (a) The chemiluminescence is visible in daylight. (b) Chemiluminescence is induced by horseradish peroxidase (HRP). (c–n) Fluorescence of organic dyes induced by chemiluminescent light from the 2-coumaranones in DMF before (c–h) and after (i–n) the addition of DBU as the base. Organic dyes: (c,i) a thiazole dye (for the structure of the dye, see the Supporting Information), (d,j) Rhodamine 6G, (e,k) sodium fluoresceinate, (f,l) 9,10-bis(phenylethynyl)anthracene, (g,m) pyranine. Panels h and n show control experiments (no dye added).

To decipher the reaction pathways for the CL of 2-coumaranones and to identify the light emitter in the CL reaction, we focused on compound **7** (Scheme 2, or **1** with R<sup>1</sup> = F and R<sup>2</sup> = Et in Scheme 1). The chemical product of the CL of **7**, compound **8**, was isolated from the reaction mixture and its structure was confirmed by X-ray diffraction (for details of the structure analysis, see the Supporting Information). This result confirmed that **8** was indeed the product of CL, at least after isolation as described earlier.<sup>[15]</sup> The maximum wavelength of the CL emission of **7** was  $\lambda_{\text{em,CL}} = 433$  nm with a quantum yield of  $(4.5 \pm 0.2) \times 10^2$  E mol<sup>-1</sup>. The fluorescence maximum ( $\lambda_{\text{em,fl}} = 434$  nm) of the reaction mixture after any observable CL had ceased ("spent solution"), excited at  $\lambda = 366$  nm, was (within spectral resolution) identical to that of CL emission. However, if product **8** was excited in MeCN at the maximum absorption in the absence of base, isolated N-protonated product **8** showed only very weak, hardly detectable fluorescence with  $\lambda_{\text{em,fl}} = 328$  nm. This observation indicated that the protonated form of **8** is neither the emitter in the CL reaction nor emitting in the spent solution. This prompted us to consider other protonated states as candidates, including tautomeric form **10** and its conjugate base **9**, the resonance structures of which, **9a**, **9b**, and **9c**, could be generated by deprotonation of **8** with a strong base such as DBU (Scheme 2).  $\alpha$ -Azacoumarinolactone anion **9** could hydrolyze in strong basic solution to open-chain carbamic acid **11**. Indeed, free carbamic acids such as **11** are well known to be thermally labile, even at room temperature;<sup>[23]</sup> decomposition by release of CO<sub>2</sub> would lead to the phenolate of 4-fluorosalicylamide **12**. All of these reactions are viable pathways and, thus, were examined as possible routes for the

ring opening before (light emitter) and after the emission (in the spent solution). We first focused our attention on the simplest product, 4-fluorosallylamide **12**. Salicylamides are known to exhibit intense blue fluorescence.<sup>[24]</sup> We prepared **12** and analyzed its absorption and fluorescence spectra (Figure 3; for details, see the Supporting Information and Table S1).



Scheme 2. Tautomerization of chemiluminescent product **8** to  $\alpha$ -azacoumarin **10** and deprotonation of chemiluminescent end-product **8** to  $\alpha$ -azacoumarinate anion **9** (resonance structures **9a**, **9b**, and **9c**). Hydrolysis of lactone **9b** leads to open-chain carbamic acid **11**, which decomposes to 4-fluorosallylamide **12**. Structures are the most stable calculated ones (see Table S3).

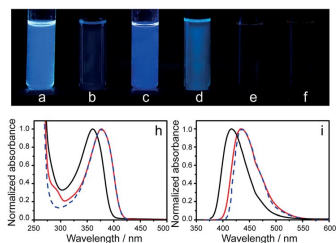


Figure 3. Fluorescence of potential emitters in DMF. (a) Compound **5** with DBU, (b) compound **5** without DBU, (c) fluorosallylamide **12** with DBU, (d) fluorosallylamide **12** without DBU, (e) isolated product **8** with DBU, and (f) isolated product **8** without DBU. (h) Absorption and (i) fluorescence spectra of 4-fluorosallylamide **12** with DBU (black), compound **5** with DBU (blue), and the spent solution after decay of the chemiluminescence (red). The solutions were excited at their specific maximum absorption wavelength.

Except for a spectral shift, the absorption and fluorescence spectra of solution of **12** and the spent solution had similar appearance and identical Stokes shifts. These results were used to establish and calibrate a quantum calculation protocol for compound **12** (see Computations Section in the Supporting Information). The absorption and emission transitions of different isomers of **12** in MeCN were computed at the DFT B3LYP/6-31+G(d,p) level of theory. The strongly basic conditions in the experiments imply that the luminophore should exist as an anion. Besides, the theoretical results for the neutral conformers do not match the experimental data (Table S2). Calculations of

the charged (–) compound resulted in most stable anion form **12m** (Table S2), with  $\lambda_{\text{abs}} = 340$  nm and  $\lambda_{\text{em}} = 405$  nm, both values deviating by only 20 to 10 nm from the respective experimental values. The small magnitude of the blueshift (<0.1 eV) confirmed the reliability of the used DFT method for this system. The components in the solution after the CL decay were established by a combined experimental and theoretical approach. O-Protonated tautomer **10** was excluded on the basis of 2D-NOESY NMR spectroscopy (see the Supporting Information). The calculations were consistent with this result on the basis of the prohibitively high energy gap between **8** and **10** ( $\Delta E = 73.4$  kJ mol<sup>-1</sup> in vacuo). Deprotonation of **8**, facilitated by base, results in **9**. Electronic calculations on deprotonated forms **9** showed that the negative charge is completely delocalized over the heteroatoms of the molecule (Figure S3). Three resonance structures are drawn in Scheme 2. Compound **10** and deprotonated basic form **9** are hetero-analogous to 4-hydroxycoumarin. These, in turn, are well known for their intense blue fluorescence, especially at basic pH values.<sup>[25]</sup> Therefore, **9** is a viable candidate as a fluorescent species. However, DFT calculations were consistent with the fact that the experimental fluorescence of anion **9** is nondetectable (Table S5, Figure 3, e).

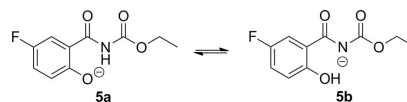
Furthermore, we considered the possibility that the fluorescence was generated by open-chain carbamic acid **11**. As mentioned above, this compound is not very stable and decomposes rapidly to **12** or cyclizes to **9**. The computed Gibbs free energies of these two reactions are in favor of decomposition to **12** ( $\Delta_r G = -50$  kJ mol<sup>-1</sup>) and **9** ( $\Delta_r G = -25$  kJ mol<sup>-1</sup>). The DFT- and CASPT2-calculated absorption and emission wavelengths of **11** are remarkably close to the experimental results obtained on the spent solution (Table 1). Therefore, open-form **11**, which coexists with nondetectable fluorescent (Figure 2, e,f) cyclic form **9**, is capable of fluorescence.

Table 1. Experimental and calculated wavelength maxima for the most stable conformations of compounds **5** (conformer **5a**) and **11**.

	Spent solution <sup>[a]</sup>	<b>5</b> (DFT) <sup>[b]</sup>	<b>5</b> (CASPT2) <sup>[c]</sup>	<b>11</b> (DFT) <sup>[b]</sup>
		$\lambda$ [nm]		
Absorption	378	367	374	369
Fluorescence	434	429	423	434

[a] For experimental details, see the Experimental Section in the Supporting Information. [b] B3LYP/6-31+G(d,p) with polarized continuum model (PCM) MeCN solvation. [c] CASPT2/ANO-RCC-VTZP in vacuo.

Compound **5**\* (with R<sup>1</sup> = F and R<sup>2</sup> = Et in Scheme 1, Table S5, and excited ester form of **11**) appears as the most viable emitter in the CL reaction. The O-deprotonated form **5a** is stabilized by about 9.65 kJ mol<sup>-1</sup> relative to N-deprotonated form **5b** (Scheme 3, Table S6, **5a** is the protonated structure drawn in Scheme 1). The spectral wavelength of this structure at the DFT and CASPT2 levels is reported in Table 1. We then focused on



Scheme 3. Different structures and isomers of compound **5**.



the synthesis of compound **5**. Initial attempts indicated that synthetic **5** is unstable in solution and decomposes quickly to final product **8**. Therefore, all subsequent procedures were performed below 5 °C and in the absence of light. The recorded absorption and emission spectra of deprotonated synthetic **5** correspond to those recorded from a spent solution (Figure 2) and are consistent with the predicted theoretical results. This finding is in agreement with the reaction path:  $v \rightarrow vi \rightarrow vii$ .

On the basis of these two arguments, we conclude that the emitter of the CL reaction of 2-coumaranone **7** is form **5a** of compound **5**.

How can this study be a model for bioluminescent systems? First, the present chemiluminescent system presents the formation and decomposition of the dioxetanone moiety, as in most bioluminescent systems, and all insight into such a decomposition can be relevant for bioluminescent systems. Second, emitter **5** may also be structurally related to the newly discovered *Fridericia heliota* oxyluciferin.<sup>[7]</sup> The latter has two moieties involved in bioluminescence: the lysine part, at which formation of dioxetanone and release of CO<sub>2</sub> occur, and a chromophore (CompX). The conjugated system of CompX is very similar to the herein-studied 2-coumaranone systems (see Figure S10). Thus, the present chemiluminescence system may be used to understand the role of the bioluminescent CompX moiety.

## Conclusions

In summary, by using a combined experimental and computational approach, a fundamental step in the mechanism of a class of strongly chemiluminescent 2-coumaranones was herein clarified. The key intermediates were identified on the basis of direct comparison with their synthetic analogues, and conclusions were aided by computational assessment of the viable isomers. Both the experimental and computational results were consistent and showed that the open form was the light emitter, not the closed form, in the chemiluminescence of 2-coumaranones. The open form may be structurally related to the chromophore (CompX) of the *Fridericia heliota* bioluminescence system and provides a convenient model for the latter, which is difficult to access by experimental methods. The results presented herein highlight the successful utility of complementary approaches in resolving nontrivial mechanistic pathways. Further calculations on the formation and opening of the dioxetanone will provide additional information on this fundamental reaction in bio- and chemiluminescent systems.

## Acknowledgments

This work was financially supported by New York University Abu Dhabi. The authors thank the service team of the IOMC and IAAC in Jena and in particular Dr. W. Günther for NMR spectro-

scopy experiments. S.S. acknowledges the Friedrich-Ebert Stiftung for financial support and Prof. L. F. M. L. Ciscato, S. Habenschicht, Prof. S. Gräfe, and Dr. S. Kupfer for their advice. This research was in part performed by using Core Technology Platform resources at New York University Abu Dhabi.

**Keywords:** Luminescence · Bioluminescence · Photochemistry · Reaction mechanisms · Oxygen heterocycles · Natural products

- [1] I. Navizet, Y.-J. Liu, N. Ferré, D. Roca-Sanjuán, R. Lindh, *ChemPhysChem* **2011**, *12*, 3064–3076.
- [2] B.-W. Ding, P. Naumov, Y.-J. Liu, *J. Chem. Theory Comput.* **2015**, *11*, 591–599.
- [3] T. Wilson, J. W. Hastings, *Annu. Rev. Cell Dev. Biol.* **1998**, *14*, 197–230.
- [4] J. Vieira, L. Pinto da Silva, J. C. G. Esteves da Silva, *J. Photochem. Photobiol. B* **2012**, *117*, 33–39.
- [5] L. Pinto da Silva, J. C. G. Esteves da Silva, *ChemPhysChem* **2012**, *13*, 2257–2262.5a.
- [6] V. N. Petushkov, M. A. Dubinnyi, A. S. Tsarkova, N. S. Rodionova, M. S. Baranov, V. S. Kublitski, O. Shimomura, I. V. Yampolsky, *Angew. Chem. Int. Ed.* **2014**, *53*, 5566–5568; *Angew. Chem.* **2014**, *126*, 5672–5674.
- [7] M. A. Dubinnyi, Z. M. Kaskova, N. S. Rodionova, M. S. Baranov, A. Y. Gorokhovatsky, A. Kotlobay, K. M. Solntsev, A. S. Tsarkova, V. N. Petushkov, I. V. Yampolsky, *Angew. Chem. Int. Ed.* **2015**, *54*, 7065–7067; *Angew. Chem.* **2015**, *127*, 7171–7173.
- [8] K. V. Purtov, V. N. Petushkov, M. S. Baranov, K. S. Mineev, N. S. Rodionova, Z. M. Kaskova, A. S. Tsarkova, A. I. Petunin, V. S. Bondar, E. K. Rodicheva, S. E. Medvedeva, Y. Oba, Y. Oba, A. S. Arseniev, S. Lukyanov, J. I. Gitelson, I. V. Yampolsky, *Angew. Chem. Int. Ed.* **2015**, *54*, 8124–8128; *Angew. Chem.* **2015**, *127*, 8242–8246.
- [9] H. Würfel, D. Weiss, R. Beckert, A. Güther, *J. Sulfur Chem.* **2012**, *33*, 9–16.
- [10] T. B. Shrestha, D. L. Troyer, S. H. Bossmann, *Synthesis* **2014**, *46*, 646–652.
- [11] A. Roda, M. Mirasoli, E. Michelini, M. Di Fusco, M. Zangheri, L. Cevenini, B. Roda, P. Simoni, *Biosens. Bioelectron.* **2016**, *76*, 164–179.
- [12] M. Zangheri, L. Cevenini, L. Anfossi, C. Baggiani, P. Simoni, F. Di Nardo, A. Roda, *Biosens. Bioelectron.* **2015**, *64*, 63–68.
- [13] L. J. Kricka, *Anal. Chim. Acta* **2003**, *500*, 279–286.
- [14] G. J. Lofthouse, H. Suschitzky, B. J. Wakefield, R. A. Whittaker, B. Tuck, *J. Chem. Soc. Perkin Trans. 1* **1979**, 1634–1639.
- [15] B. Matuszczak, *Monatsh. Chem.* **1996**, *127*, 1291–1303.
- [16] B. Matuszczak, *Die Pharmazie* **1996**, *51*, 862–865.
- [17] B. Matuszczak, *Monatsh. Chem.* **1997**, *128*, 945–951.
- [18] B. Matuszczak, *J. Prakt. Chem./Chem.-Ztg.* **1998**, *340*, 20–25.
- [19] S. Schramm, L. F. M. L. Ciscato, P. Oesau, R. Krieg, F. Richter, I. Navizet, D. Roca-Sanjuána, D. Weiß, R. Beckert, *ARKIVOC (Gainesville, FL, U.S.)* **2015**, 4, 44–59.
- [20] S. Schramm, D. Weiß, H. Brandl, R. Beckert, H. Görls, D. Roca-Sanjuána, I. Navizet, *ARKIVOC (Gainesville, FL, U.S.)* **2013**, 3, 174–188.
- [21] L. F. M. L. Ciscato, F. H. Bartoloni, A. S. Colavite, D. Weiss, R. Beckert, S. Schramm, *Photochem. Photobiol. Sci.* **2014**, *13*, 32–37.
- [22] F. H. Bartoloni, M. A. de Oliveira, L. F. M. L. Ciscato, F. A. Augusto, E. L. Bastos, W. J. Baader, *J. Org. Chem.* **2015**, *80*, 3745–3751.
- [23] M. M. Sidky, A. A. El-kateb, M. R. Mahran, I. T. Hennawy, H. A. A. El-Malek, *Phosphorus Sulfur Silicon Relat. Elem.* **1987**, *29*, 11–15.
- [24] G. J. Woolfe, P. J. Thistlethwaite, *J. Am. Chem. Soc.* **1980**, *102*, 6917–6923.
- [25] R. Poláček, P. Májek, K. Hroboňová, J. Sádecká, *J. Fluoresc.* **2015**, *25*, 297–293.

Received: December 2, 2015

Published Online: January 12, 2016

# Table des figures

1	Mécanisme de la bioluminescence chez les lucioles et structure chimique du substrat oxyluciférine (OxyLH <sub>2</sub> ) dans la conformation phénolate-kéto. . . .	2
1.1	Représentation schématique de l'approche CASSCF sur un système moléculaire donné. L'espace actif CAS contient 6 électrons dans 4 orbitales actives et est alors noté CAS(6-in-4). Les orbitales inactives et virtuelles correspondent respectivement aux orbitales doublement occupées et inoccupées situées hors de l'espace actif. . . . .	13
1.2	Représentation schématique des doubles excitations incluses dans V <sub>SD</sub> . L'espace actif est inclus à l'intérieur de la boîte dessinée, sous cette boîte se trouve les orbitales inactives et au-dessus les orbitales virtuelles. Ces doubles excitations sont nommées selon le nombre de trou (hole, h) dans l'espace inactif et particule (p) dans l'espace virtuel créés par ces doubles excitations. Ainsi la double excitation 2h1p signifie que cette dernière crée 2 trous dans les orbitales inactives et ajoute 1 particule dans les orbitales virtuelles. . . .	15
1.3	Comparaison de la qualité d'une fonction de Slater 1s par rapport à des fonctions STO-1G, STO-2G et STO-3G. Cette figure est extraite du livre <sup>23</sup> , page 158.	21
2.1	Schéma représentant les 3 types d'histidines, Hie, His et Hip que l'on peut trouver dans un système biologiques. Les lettres et chiffres situés en indices des symboles chimiques correspondent à la représentation de cet atome dans le champ de force ff99. . . . .	26
2.2	Termes liés de la fonction d'énergie potentielle. En haut à gauche, terme de liaison de longueur l, en haut à droite, terme d'angle de valeur $\theta$ , en bas à gauche terme d'angle dièdre de valeur $\omega$ et à droite terme d'angle impropre $\theta$ .	27
2.3	Schéma d'une sphère avec un rayon de cutoff de la taille d'une demi-boîte et de centre la particule noire. Les particules grisées sont prises en compte dans le calcul des interactions. . . . .	32
3.1	Schéma d'une partition QM/MM. . . . .	34
3.2	Représentation des deux différentes coupures entre la partie QM et la partie MM. A gauche, l'utilisation d'un link-atom noté HLA. A droite l'utilisation des orbitales localisées. Les atomes notés Q <sub>x</sub> correspondent aux atomes de la partie QM et les atomes notés E <sub>x</sub> à ceux de la partie MM. La coupure QM/MM s'effectue entre l'atome Q <sub>1</sub> et l'atome E <sub>1</sub> . . . . .	37

TABLE DES FIGURES

4.1	Définition des différentes surfaces (en trait épais). a) Surface de van-der-Waals (vdWS) b) Surface Accessible au solvant (SAS) c) Surface Excluant le Solvant (SES). Appropriate credit for the requested material should be given as follows : "Adapted with permission from (Darrin M. York and Martin Karplus. A smooth solvation potential based on the conductor-like screening model. The Journal of Physical Chemistry A, 103(50) :11060-11079, 1999) Copyright (2017) American Chemical Society . . . . .	40
5.1	Schéma des principales réactions photochimiques sur une surface d'énergie potentielle. . . . .	45
6.1	Le diagramme de Perrin-Jablonski. . . . .	51
6.2	A gauche, structure d'une 2-coumaranone. R <sup>1</sup> est un halogène et R <sup>2</sup> un aryle ou un alkyle. A droite, montage montrant l'émission de chimiluminescence des 2-coumaranones dans un pièce obscure (demi-image de gauche) et en plein jour (demi-image de droite). . . . .	54
6.3	Images de plusieurs espèces bioluminescentes : a) <i>Luciola cruciata</i> , Crédit : Aflo/naturepl.com. b) <i>Dinoflagellata</i> , Crédit : Doug Perrine/naturepl.com. c) <i>Photonus Pyralis</i> , Crédit : By art farmer from evansville indiana, usa - firefly 8823, CC BY-SA 2.0. d) <i>Panellus stipticus</i> , Crédit : By Ylem - Own work . . . .	56
6.4	Image de différentes applications utilisant la bioluminescentes : a) BLI de cellules cancéreuse du sein chez la souris <sup>90</sup> . b) Bâtiment éclairé par la bioluminescence issue de bactérie symbiotiques du calamar, Crédit : Glowee <a href="http://www.glowee.fr/">http://www.glowee.fr/</a> c) BLI issue d'un cerveau de drosophile après un influx nerveux <sup>91</sup> . . . . .	57
7.1	L-acide aminé. . . . .	63
7.2	Structures et abréviations des 20 acides aminés standard. La chaîne latérale (groupement R) est représentée en bleu. . . . .	64
7.3	Structure d'un polypeptide d'une protéine selon la convention standard. Le résidu 1 est appelé le résidu N-terminal et le résidu 4 est appelé le résidu C-terminal. . . . .	65
7.4	A gauche, structure en ruban d'une hélice $\alpha$ et représentation en tout atome des résidus la composant. A droite, structure en ruban de deux feuillets $\beta$ antiparallèles en jaune ainsi que d'une boucle "épingle a cheveu" reliant ces deux feuillets. . . . .	65
7.5	Structure des domaines de la luciférase. En bleu, le domaine N-terminal et en violet le domaine C-terminal. . . . .	66
7.6	Structure cristallographique déprotonée de l'hémoglobine $\alpha$ . Les différentes sous-unités sont représentées en ruban et d'une couleur différente. Les quatre hèmes sont représentés en tout atome. La structure est issue du PDB 3hhb.pdb <sup>94</sup> . . . . .	68
7.7	Mouvement des micro-filaments d'actines lors de la contraction musculaire. . . . .	68

TABLE DES FIGURES

8.1	Représentation en tout-atome d'un tripeptide ainsi que les modèles gros-grain correspondant <sup>97</sup> . Les modèles présentés sont le Rosetta centroid mode (CEN), le CABS, le UNRES, le SICHO et le modèle Levitt et Warshel. Le code couleur est le suivant : O rouge, H noir, C cyan, N bleu, pseudo-atomes orange et bleu foncé. Les pseudo-liaisons de distances variables sont représentées sous forme de ressort. . . . .	72
8.2	Schéma des différentes étapes d'un umbrella sampling. La coordonnée de réaction est séparée en plusieurs fenêtres, les valeurs obtenues sont alors regroupées pour former un histogramme pondéré permettant d'obtenir le PMF.	73
8.3	Schéma du fonctionnement d'un screening virtuel grâce au docking <sup>102</sup> . . . . .	74
9.1	Représentation graphique de la luciférine LH <sub>2</sub> à gauche et de l'infra-luciférine iLH <sub>2</sub> à droite. . . . .	81
9.2	Représentation graphique des différents substrats issus de OxyiLH <sub>2</sub> étudiés dans ce chapitre. . . . .	82
9.3	Représentation graphique des différents modèles principalement utilisés pour cette étude. Sur la gauche, les structures et noms des 3 principaux substrats étudiés. Sur la droite, la structure et nom du résidu 284 dans la protéine naturelle (Sérine) et dans la protéine mutée (Thréonine). Au milieu, la représentation de la protéine luciférase avec OxyiLH <sub>2</sub> -E-1. La protéine est représentée sous forme de ruban avec les hélices alpha en violet et les feuillets bêta en jaune. La position du résidu 284 est indiquée en rouge. . . . .	83
9.4	Orbitales frontières de OxyiLH <sub>2</sub> -E-1 (en haut) and OxyiLH <sub>2</sub> -E-2 (en bas) obtenues par un calcul CASSCF/MM dans la protéine et dessinées avec le logiciel Luscus <sup>112</sup> . . . . .	88
9.5	Représentation du réseau de liaisons hydrogène dans le modèle Ser_OxyLH <sub>2</sub> avant la dynamique moléculaire classique (MD) (à gauche) et "après MD" (à droite). . . . .	88
9.6	Structure de OxyiLH <sub>2</sub> -E-1 (à gauche) définissant les dièdres $\beta_{1-12}$ et $\beta_{11-8}$ ; Structure de OxyiLH <sub>2</sub> -E-2 (au milieu) définissant les dièdres $\beta_{1-12}$ et $\beta_{11-8}$ ; Structure de OxyiLH <sub>2</sub> -E-1 (à droite) définissant le dièdre $\beta_{2cycles}$ . . . . .	90
9.7	Superposition géométrique des structures OxyiLH <sub>2</sub> -E-1 (à gauche) et OxyiLH <sub>2</sub> -E-2 (à droite) dans le vide et dans les protéines 4G37 et 4G37-S284T. La superposition est réalisée le long des atomes C11 et C12 du pont entre les deux cycles. Le code couleur est le suivant : géométrie obtenue dans le vide cyan ; dans la protéine 4G37 rouge ; dans la protéine 4G37-S284T violet. . . . .	91
10.1	Photos de l'espèce bioluminescente <i>Amydetes viviani</i> à gauche et de l'espèce bioluminescente <i>Phrixothrix hirtus</i> à droite. Crédits photos : Prof. Vadim Viviani. . . . .	99
10.2	Représentation de la conformation des domaines C-terminaux de la structure GB <sub>Av</sub> -B et 2D1R à gauche et de la structure GB <sub>Av</sub> -B et 4G37 à droite. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines. La protéine GB <sub>Av</sub> -B est dessinée en ruban bleu, la 2D1R en vert et la 4G37 en orange. Les résidus colorés en rouge représentent les derniers résidus de la chaîne polypeptidique. . . . .	100

TABLE DES FIGURES

10.3	Représentation des conformations des domaines C-terminaux de l' <i>image</i> obtenue à 16 ns de la MD sur GB <sub>Av</sub> -B et la structure cristallographique GB <sub>Av</sub> -A partielle. La protéine GB <sub>Av</sub> -B est dessinée en ruban bleu, et la GB <sub>Av</sub> -A partielle en violet. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines. . . . .	101
10.4	Représentation des conformations des domaines C-terminaux de la structure construite de la GB <sub>Av</sub> -A-construite et la structure cristallographique GB <sub>Av</sub> -A complète. La structure construite de la GB <sub>Av</sub> -A-construite est dessinée en ruban bleu de prusse, et la GB <sub>Av</sub> -A en violet. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines. . . . .	102
10.5	Structure du site actif de la GB <sub>Av</sub> -B-OxyLH <sub>2</sub> de l' <i>image</i> (2) du tableau 10.1, superposée avec la structure du système de la luciole 2D1R. Les structures sont comparées en superposant leur domaines N-terminaux. L'OxyLH <sub>2</sub> dans l' <i>image</i> (2) du tableau 10.1 est représentée en tout atomes. De la structure modèle 2D1R, seule OxyLH <sub>2</sub> est représentée en vert pour plus de clarté. Le résidu Ser 245, représenté en CPK correspond à une substitution entre 2D1R (Phe) et GB <sub>Av</sub> (Ser). . . . .	102
10.6	Structure du site actif de la GB <sub>Av</sub> -B-OxyLH <sub>2</sub> dans les différents <i>images</i> (1 à 5). Les domaines N-terminaux ont été superposés. Le substrat OxyLH <sub>2</sub> est représenté dans une couleur dépendant de l' <i>image</i> référencée dans le tableau 10.1. En bleu, le substrat de l' <i>image</i> (1), en vert le (2), en jaune le (3), en orange le (4) et en rouge le (5). Le résidu Ser245 est mis en valeur par une représentation tout atome. . . . .	104
10.7	Structure du site actif de la GB <sub>Av</sub> -B-OxyLH <sub>2</sub> +H <sub>2</sub> O, superposée avec la structure du système GB <sub>Av</sub> -B-OxyLH <sub>2</sub> . La comparaison entre les conformations de OxyLH <sub>2</sub> dans l' <i>image</i> (2) du tableau 10.2, est représentée en tout atomes et OxyLH <sub>2</sub> issu de l' <i>image</i> (3) du tableau 10.1 dans le modèle GB <sub>Av</sub> -B-OxyLH <sub>2</sub> en jaune. Les structures sont comparées en superposant leur domaines N-terminaux. Le résidu Ser 245 et la molécule d'eau ajoutée artificiellement sont représentés en CPK. . . . .	104
10.8	Comparaison entre les domaines C-terminaux de la structure GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed et 2D1R à gauche et de la structure GB <sub>Av</sub> -B-closed_w_OxyLH <sub>2</sub> et 2D1R à droite. Les structures sont comparées en superposant les domaines N-terminaux, représentés en bleu cyan, des deux protéines. La protéine GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed est dessinée en ruban bleu foncé, la 2D1R en vert et la GB <sub>Av</sub> -B-closed_w_OxyLH <sub>2</sub> en bleu roi. . . . .	106
10.9	Structure du site actif de la GB <sub>Av</sub> -B-closed_w_OxyLH <sub>2</sub> de l' <i>image</i> (2) du tableau 10.3, superposée avec la structure du système de la luciole 2D1R. Les structures sont comparées en superposant leur domaines N-terminaux. L'OxyLH <sub>2</sub> dans l' <i>image</i> (2) du tableau 10.3 est représentée en tout atomes. De la structure modèle 2D1R, seule OxyLH <sub>2</sub> est représentée en vert pour plus de clarté. Le résidu Ser 245, représenté en CPK correspond à une substitution entre 2D1R (Phe) et GB <sub>Av</sub> (Ser). . . . .	107

TABLE DES FIGURES

10.10	Structure du site actif de la GB <sub>Av</sub> -B-closed_w_OxyLH <sub>2</sub> dans les différentes <i>images</i> (1 à 2). Les domaines N-terminaux ont été superposés. Le substrat OxyLH <sub>2</sub> est représenté dans une couleur différente dépendant de l' <i>image</i> référencée dans le tableau 10.3. En bleu, le substrat de l' <i>image</i> (1), en vert le (2). . . . .	108
10.11	Structure du site actif de la GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed de l' <i>image</i> (2) du tableau 10.4, superposée avec la structure du système de la luciole 2D1R. Les structures sont comparées en superposant leur domaines N-terminaux. L'OxyLH <sub>2</sub> dans l' <i>image</i> (2) du tableau 10.4 est représentée en tout atomes. De la structure modèle 2D1R, seule OxyLH <sub>2</sub> est représentée en vert pour plus de clarté. Le résidu Ser 245, représenté en CPK correspond à une substitution entre 2D1R (Phe) et GB <sub>Av</sub> (Ser). . . . .	109
10.12	Structure du site actif de la GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed dans les différents <i>images</i> (1 à 5). Les domaines N-terminaux ont été superposés. Le substrat OxyLH <sub>2</sub> est représenté dans une couleur différente dépendant de l' <i>image</i> référencée dans le tableau 10.4. En bleu, celui de l' <i>image</i> (1), en vert le (2), en jaune le (3), en orange le (4) et en rouge le (5). . . . .	110
10.13	Structure du site actif dans la GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed avec les principaux résidus importants dans l'étude des mutations. . . . .	111
10.14	Structure du site actif dans la GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L. L' <i>image</i> représentée est la (2) et est issue du tableau 10.5. . . . .	113
10.15	Orbitales frontières HOMO et LUMO dans OxyLH <sub>2</sub> . . . . .	116
10.16	Schéma de l'influence des liaisons hydrogène sur la longueur d'onde d'émission de l'oxyluciférine . . . . .	117
11.1	Les deux chemins possibles de la décomposition de l'intermédiaire <b>2</b> dans la chimiluminescence des 2-coumaranones <b>1</b> . R <sub>1</sub> est un atome de fluor et R <sub>2</sub> un groupement éthyl. . . . .	126
11.2	Le chemin de décomposition de <b>6</b> en <b>8</b> avec la structure intermédiaire fluorescente <b>7</b> . . . . .	127
11.3	Schéma des 3 trois structures cibles. . . . .	128
12.1	Schéma des 6 formes chimiques de l'oxyluciférine. . . . .	129
12.2	Représentation graphique du système étudié dans cette partie. Les traits en pointillés représentent de possibles liaisons hydrogène. . . . .	130
12.3	Structure kéto choisie comme structure de départ de l'étude de la tautomérisation entre les deux conformères de l'oxyluciférine kéto-énol. Cette structure est issue de la simulation MD numéro 1. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . .	130
12.4	Zoom sur la structure de l'état de transition obtenue en PCM dans l'état fondamental S <sub>0</sub> . L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . .	132
12.5	Zoom sur la structure de l'état de transition obtenue en PCM dans l'état excité S <sub>1</sub> . L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . .	132

TABLE DES FIGURES

- 12.6 En haut, graphique du profil énergétique en fonction de l'IRC en solvant implicite (PCM) dans l'eau pour S0. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0. En marron, la courbe du profil énergétique vers la forme kéto dans S0; en bleu la courbe du profil énergétique vers la forme énol dans S0. En bas, graphique du profil énergétique en fonction de l'IRC en solvant implicite (PCM) dans l'eau pour S1. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS). . . . . 133
- 12.7 Graphique du profil énergétique en fonction de l'IRC en solvant implicite (PCM) dans l'eau. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0. En marron, la courbe du profil énergétique vers la forme kéto dans S0; en bleu la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS). 134
- 12.8 Représentation de l'état de transition appelé TS1 et dans l'état excité S1. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . . 137
- 12.9 Représentation de l'état de transition appelé TS2 et dans l'état excité S1. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . . 137
- 12.10 Graphique du profil énergétique en fonction de l'IRC dans la protéine pour le TS1. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0 dans la protéine. En marron, la courbe du profil énergétique vers la forme kéto dans S0; en bleu la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS). 139

TABLE DES FIGURES

12.11	Représentation du tautomère phénolate-énol obtenu pour le dernier point du profil énergétique vers la forme énol pour le TS1 et le TS2, dans l'état excité S1 et dans la protéine. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . .	140
12.12	Graphique du profil énergétique en fonction de l'IRC dans la protéine pour le TS2. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0 dans la protéine. En marron, la courbe du profil énergétique vers la forme kéto dans S0; en bleu la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1 (S1GS). En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1 (S1GS).	141
12.13	Représentation du tautomère phénolate-kéto obtenu pour le dernier point du profil énergétique vers la forme kéto pour le TS1 et le TS2, dans l'état excité S1 et dans la protéine. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . .	141
12.14	Graphique du scan réalisé autour du dièdre O4-P-O5-C4. A gauche la structure issue du dernier point de l'IRC vers le phénolate-kéto, à partir de l'état de transition figure 12.13. A droite, l'énergie de la structure après la rotation de 60° du dièdre. Cette structure correspond à la structure de gauche dans 12.15 et est similaire à la structure de départ. . . . .	143
12.15	A gauche, représentation du tautomère phénolate-kéto après la rotation de 60° du dièdre O4-P-O5-C4 dans la protéine et dans S0. A droite, représentation du tautomère phénolate-kéto dans la structure de départ à l'intérieur de la protéine et dans S0. L'oxyluciférine, l'AMPH et la molécule d'eau sont représentés sous forme de bâtons. . . . .	143



TABLE DES FIGURES

12.16	Graphique du profil énergétique en fonction de l'IRC dans la protéine pour le TS1 et le TS2. Le tautomère phénolate-kéto est formé vers les coordonnées de réaction positives et le tautomère phénolate-énol est formé vers les coordonnées de réaction négative. Le point à 0 eV correspond au dernier point du chemin réactionnel calculé dans le cas du phénolate-kéto dans S0 dans la protéine. Pour le TS1, en ligne rouge, la courbe du profil énergétique vers la forme kéto dans S0 ; en vert la courbe du profil énergétique vers la forme énol dans S0. En croix rouge opaque, la courbe du profil énergétique vers la forme kéto dans S1 ; en croix rouge transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1. En croix verte opaque, la courbe du profil énergétique vers la forme énol dans S1 ; en croix verte transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1. Pour le TS2, en ligne jaune, la courbe du profil énergétique vers la forme kéto dans S0 ; en vert clair la courbe du profil énergétique vers la forme énol dans S0. En croix jaune opaque, la courbe du profil énergétique vers la forme kéto dans S1 ; en croix jaune transparente, la courbe du profil énergétique vers la forme kéto dans S0 calculée à partir de S1. En croix verte claire opaque, la courbe du profil énergétique vers la forme énol dans S1 ; en croix verte claire transparente, la courbe du profil énergétique vers la forme énol dans S0 calculée à partir de S1. . . . .	145
13.1	Représentation graphique du substrat Int et de la molécule de dioxygène. Les coupures entre la partie QM et les parties MM de ce substrat sont indiqués. .	151
13.2	Représentation des deux hypothèses proposées pour la coordination de O <sub>2</sub> sur le substrat Int. A gauche via un ion intermédiaire superoxyde (mécanisme étudié dans ce chapitre). A droite via un ion intermédiaire hydroperoxyde. .	152
13.3	Représentation du système avant la simulation umbrella sampling et appelé 4G36-O <sub>2</sub> . Le substrat Int et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons. O <sub>2</sub> est placé dans un canal joignant le solvant et le site actif de la protéine. La protéine est représentée en dessinant la surface de cette dernière avec un rayon de 1.4 Ået par la couleur grise. . . . .	152
13.4	Profil de l'énergie libre obtenue lors la simulation umbrella sampling. . . .	153
13.5	Représentation du système après la simulation umbrella sampling et après la minimisation au niveau MM. Le substrat Int et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par le modèle CPK. . . . .	154
13.6	Représentation du système après la simulation umbrella sampling et la minimisation en DFT/MM. Le substrat Int-dioxétanone et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. . . . .	155
13.7	Schéma représentant les différentes étapes et approches utilisées lors de l'approche de O <sub>2</sub> vers le substrat Int pour les calculs SA-CASSCF/MM. . . . .	156

TABLE DES FIGURES

13.8	Graphique représentant l'énergie du système en eV en fonction de la valeur de la distance contrainte. Le point à zéro eV correspond à l'énergie de la structure pour D-coord valant 3.5 Å et nommée 4G36-O <sub>2</sub> -min. Le point nommé Int-A-t-dioxétanone correspond au point obtenu lors de la formation du cycle dioxétanone après l'application de la contrainte D-ester sur la structure Int-t-O <sub>2</sub> avec D-coord = 1.5 Å. En rouge correspond au chemin réactionnel de l'état S <sub>0</sub> , en orange S <sub>1</sub> , en vert l'état T <sub>0</sub> et en bleu l'état T <sub>1</sub> . . . .	156
13.9	Représentation du système après la simulation umbrella sampling et pour le modèle Int-O <sub>2</sub> avec D-coord valant 2.3 Å. Le substrat Int et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord. . . . .	157
13.10	Représentation du système après la simulation umbrella sampling et pour Int-s-O <sub>2</sub> avec D-coord valant 2.2 Å pour les états singulets. Le substrat Int et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord. . . . .	158
13.11	Représentation du système après la simulation umbrella sampling et pour Int-s-O <sub>2</sub> avec D-coord valant 1.6 Å pour les états singulets. Le substrat Int et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord. . . . .	159
13.12	Représentation du système après la MD umbrella sampling et pour Int-t-O <sub>2</sub> avec D-coord valant 1.8 Å pour les états triplets. Le substrat Int et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de D-coord. . . . .	159
13.13	Représentation du système après la simulation umbrella sampling et pour Int-t-O <sub>2</sub> avec D-coord valant 1.5 Å pour les états triplets. Le substrat Int-dioxétanone et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS. . . . .	160
13.14	Représentation du système après la simulation umbrella sampling, pour la structure Int-B-s, la structure obtenue après une contrainte D-ester entre C2 et O4 de 1.8 Å sur la structure Int-O <sub>2</sub> avec D-coord = 2.3 Å et pour les états singulets. Le substrat Int et la molécule de O <sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4. . . . .	162

TABLE DES FIGURES

13.15 Graphique représentant l'énergie du système en eV pour 4 points lors de l'approche de O<sub>2</sub>. Le point à zéro eV correspond à l'énergie de la structure pour D-coord valant 3.5 Å. Le point nommé Int-B-s/t correspond au point obtenu après l'application de la contrainte D-ester sur la structure Int-O2 avec D-coord = 2.3 Å. Le point nommé Int-B-s/t-dioxétanone correspond au point obtenu après l'application de la contrainte D-coord = 1.5 Å suivi d'une minimisation sur la structure Int-B-s/t. En rouge correspond l'état S0, en orange S1, en vert l'état T0 et en bleu l'état T1. . . . . 163

13.16 Représentation du système après la simulation umbrella sampling, pour la structure Int-B-s-dioxétanone, après une contrainte D-coord valant 1.5 Å suivi d'une minimisation sur la structure Int-B-s pour les états singulets. Le substrat Int-dioxétanone et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4. . . . . 164

13.17 Représentation du système après la simulation umbrella sampling, pour la structure Int-B-t, la structure obtenue après une contrainte D-ester entre C2 et O4 de 1.8 Å sur la structure Int-O2 avec D-coord = 2.3 Å et pour les états triplets. Le substrat Int et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et la molécule d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4. . . . . 165

13.18 Représentation du système après la simulation umbrella sampling, pour la structure Int-B-t-dioxétanone, après une contrainte D-coord valant 1.5 Å suivi d'une minimisation sur la structure Int-B-t pour les états triplets. Le substrat Int-dioxétanone et la molécule de O<sub>2</sub> sont représentés sous forme de bâtons, le résidu Hip 245 et les molécules d'eau par la représentation CPK. Les traits pointillés représente les distances C1-O1 et C2-O2 intervenant dans le calcul de Dis-CAS ainsi que la distance C2-O4. . . . . 165

# Liste des tableaux

9.1	Description des 10 modèles utilisés dans ce chapitre. . . . .	83
9.2	Transitions électroniques entre S1 and S0 ( $T_e$ ) (et force d'oscillateur associée) obtenues en QM/MM sur la structure 4G37 pour différentes fonctionnelles. TD-DFT/MM avec une base 6-311G(2d,p) sur une structure optimisée dans l'état S1 avec la fonctionnelle indiquée. . . . .	84
9.3	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en MS-CASPT2/MM entre différents espaces actifs et 2 valeurs différentes du shift IPEA. Les calculs sont réalisés avec une base ANO-RCC-VTZP, en tenant compte de 2 états électroniques et avec un level shift de 0.1. . . . .	85
9.4	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. MS-CASPT2/MM obtenu avec une base ANO-RCC-VTZP, un espace actif 16-in-14 avec 2 états électroniques, un level shift de 0.1 et un IPEA shift de 0.1. TD-DFT/MM obtenu avec une base 6-311G(2d,p) sur une structure optimisée dans l'état S1 avec la fonctionnelle B3LYP. . . . .	86
9.5	Transitions électroniques entre S1 and S0 $T_e$ (et force d'oscillateur associée) obtenues en TD-DFT/MM avec la fonctionnelle B3LYP et la base 6-311G(2d,p) pour différents environnements et sur une structure optimisée dans l'état S1 avec la même fonctionnelle dans les mêmes conditions. . . . .	86
9.6	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur une structure issue avant et durant une dynamique moléculaire classique. MS-CASPT2/MM obtenu avec une base ANO-RCC-VTZP, un espace actif 16-in-14 avec 2 états électroniques, un level shift de 0.1 et un IPEA shift de 0.1. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la fonctionnelle B3LYP. . . . .	88
9.7	Contribution de l'effet électrostatique sur les substrats OxyiLH <sub>2</sub> -E-1 et OxyiLH <sub>2</sub> -E-2 dans la protéine 4G37 et la protéine 4G37-S284T. <i>Single point</i> TD-DFT obtenu avec une base 6-311G(2d,p) dans le vide sur un structure optimisée géométriquement en TD-DFT/MM avec la même fonctionnelle et base. . . . .	90
9.8	Paramètres géométriques des deux substrats OxyiLH <sub>2</sub> -E-1 et OxyiLH <sub>2</sub> -E-2 optimisés dans la protéine ou dans le vide (optimisation réalisée avec la fonctionnelle B3LYP et la base 6-311G(2d,p)). $T_e^{vac}$ réalisée sur ces structures (fonctionnelle B3LYP et la base 6-311G(2d,p)). . . . .	92
9.9	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. TD-DFT/MM avec une base 6-311G(2d,p) sur une structure optimisée dans l'état S1 avec la fonctionnelle indiquée. . . . .	93

LISTE DES TABLEAUX

10.1 Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B-open. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé dans 4G37 est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que les valeurs expérimentales issues des références<sup>17,2</sup>. . . . . 103

10.2 Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B-OxyLH<sub>2</sub>+H<sub>2</sub>O et 4G37-OxyLH<sub>2</sub>-F250S. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé dans 4G37 est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que les valeurs expérimentales issues des références<sup>17,2</sup>. . . 105

10.3 Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B-closed\_w\_OxyLH<sub>2</sub>. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé dans 4G37 est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que les valeurs expérimentales issues des références<sup>17,2</sup>. . . . . 108

10.4 Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence<sup>17</sup>. . . . . 110

10.5 Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-R332L. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé pour le modèle GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence<sup>17</sup>. . . . . 113

10.6 Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed-I346L. Les numéros entre parenthèses correspondent au numéro de l'*image* extraite de la MD. La valeur obtenue selon le même procédé pour le modèle GB<sub>Av</sub>-B+OxyLH<sub>2</sub>-closed est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de  $T_e^{vac}$  pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence<sup>17</sup>. . . . . 114

LISTE DES TABLEAUX

10.7	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-insert-R351. Les numéros entre parenthèses correspondent au numéro de l' <i>image</i> extraite de la MD. La valeur obtenue selon le même procédé pour le modèle GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed est donnée pour comparaison. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée dans l'état S1 avec la même fonctionnelle. La valeur de $T_e^{vac}$ pour chaque structure est donnée ainsi que la valeur expérimentale issue de la référence <sup>17</sup> . . . . .	115
10.8	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM sur la structure GB <sub>Av</sub> -B+OxyLH <sub>2</sub> -closed-R332L optimisée dans S1 et comparaison avec les <i>single points</i> obtenus en mettant les charges des atomes de l'arginine 332 égales à zéro, appelé "off" dans le tableau. Les numéros entre parenthèses correspondent aux numéros de l' <i>image</i> extraite de la MD. Les valeurs sont obtenues par TD-DFT/MM avec une base 6-311G(2d,p) sur un structure optimisée géométriquement avec la même fonctionnelle. . . . .	117
10.9	Tableau des différents modèles utilisés dans ce chapitre ainsi que les numéros des résidus des histidines doublement protonés dans ces derniers. . . .	119
11.1	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues sur 2 possibles émetteurs de la chimiluminescence chez les 2-coumaranones. TD-DFT obtenu avec une base 6-31+G(d,p) sur un structure optimisée géométriquement avec la même fonctionnelle. MS-CASPT2 obtenu avec un espace actif de 14 électrons dans 12 orbitales en tenant compte de 4 états électroniques et avec une base ANO-RCC-VTZP. Les calculs sont effectués pour les structures 4* et 5* . .	126
12.1	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. La valeur de référence provient de calculs TD-DFT/MM effectués sur une <i>image</i> extraite d'une MD sur un modèle contenant le substrat phénolate-kéto et sur un modèle contenant le substrat phénolate-énol. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure obtenue dans l'état S1 avec la même fonctionnelle. . . . .	140
12.2	Transitions électroniques entre S1 and S0 ( $T_e$ ) obtenues en QM/MM. La valeur de référence provient de calculs TD-DFT/MM effectués sur une <i>image</i> extraite d'une MD sur un modèle contenant le substrat phénolate-kéto et sur un modèle contenant le substrat phénolate-énol. TD-DFT/MM avec une base 6-311G(2d,p) sur un structure obtenue dans l'état S1 avec la même fonctionnelle. . . . .	142







## Résumé

La bioluminescence est un processus complexe dans lequel la réaction chimique d'un substrat, catalysée par une protéine, entraîne l'apparition d'une émission de lumière dans le spectre visible. Dans le cas des lucioles, un insecte émettant dans le domaine du jaune-vert, le substrat se nomme luciférine et la protéine luciférase. Cependant, la taille et la complexité de ce système chimique limite sa compréhension, notamment celle du mécanisme réactionnel. L'apport de la chimie théorique dans ce domaine est essentiel et a prouvé son utilité dans de nombreux cas. L'utilisation de la méthode QM/MM, méthode hybride couplant la mécanique quantique et la mécanique moléculaire permet de modéliser et d'étudier ces systèmes biologiques. Cette thèse se focalise sur deux approches différentes de l'étude de la bioluminescence chez les lucioles. La première consiste à étudier l'effet de certaines modifications chimiques sur la couleur de la bioluminescence. On s'intéresse plus particulièrement à un analogue de la luciférine et à certaines luciférases issues d'autres systèmes bioluminescents. Par cette étude on cherche à rationaliser et à prédire l'effet ainsi que l'impact de ces changements sur l'émission. Le deuxième sujet explore deux étapes du mécanisme réactionnel de la bioluminescence : d'une part, la coordination du dioxygène sur un intermédiaire de la réaction, une étape encore non étudiée et d'autre part la réaction de tautomérisation dans l'état excité et au sein de la protéine entre deux formes émissives possibles de la luciférine.

Mots clefs : Bioluminescence, états excités, Chimie théorique, QM/MM, MD, réaction enzymatique

## Abstract

The bioluminescence is a complex process that involves the reaction of a substrate, catalysed by an enzyme that sheds light in the visible spectra. In fireflies, the light emitted has a yellow-green tone thanks to the interaction between the substrate luciferin and the protein luciferase. However the size and the complexity of the system prevent its comprehension especially when dealing with the reaction mechanism. The use of computational chemistry is key to understand and improve the comprehension of the bioluminescence. The hybrid QM/MM method that combines quantum mechanics with molecular mechanics is a great tool to model and study bioluminescent systems. This thesis deals with two different approaches of the bioluminescence in fireflies. The first one is related to the study of chemical modifications that tune the emission colour. We will discuss about one analogue of the luciferin and on new luciferases from others bioluminescent species. The goals of this study are to rationalise and predict both the effect and the impact of these modifications on the emission. The second subject deals with two different steps of the bioluminescent mechanism. The first one discusses the binding of the dioxygen to the bioluminescent intermediate, which was so far unstudied and the second one about the tautomerization in the excited state and in the protein of two possible emissive forms of the luciferin.

Key words : Excited States, Bioluminescence, Theoretical Chemistry, QM/MM, MD, Enzymatic reaction