



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Thèse en vue de l'obtention du titre de

Docteur de l'Université Henri Poincaré
(Spécialité Chimie Informatique et Théorique)

**Étude de phénomènes électroniques de macromolécules à
l'aide de méthodes hybrides QM-MM.**

Adèle D. LAURENT

École Doctorale Lorraine de Chimie et de Physique Moléculaires (SESAMES).

Cette thèse est présentée et soutenue le 1er octobre 2010.

Le jury sera constitué de :

<i>Rapporteurs :</i>	Mme Chantal DANIEL	Directeur de recherche, Université de Strasbourg, France.
	Mme Isabelle DEMACHY	Professeur, Université de Paris-Sud 11, France.
<i>Examineurs :</i>	Mr Denis JACQUEMIN	Professeur, Université de Nantes, France.
	Mlle Élise DUMONT	Maître de conférence, École Normale Supérieure de Lyon, France.
	Mr Xavier ASSFELD	Professeur, Université de Nancy, France.

Thèse en vue de l'obtention du titre de

Docteur de l'Université Henri Poincaré
(Spécialité Chimie Informatique et Théorique)

**Étude de phénomènes électroniques de macromolécules à
l'aide de méthodes hybrides QM-MM.**

Adèle D. LAURENT

École Doctorale Lorraine de Chimie et de Physique Moléculaires (SESAMES).

Cette thèse est présentée et soutenue le 1er octobre 2010.

Le jury sera constitué de :

<i>Rapporteurs :</i>	Mme Chantal DANIEL	Directeur de recherche, Université de Strasbourg, France.
	Mme Isabelle DEMACHY	Professeur, Université de Paris-Sud 11, France.
<i>Examineurs :</i>	Mr Denis JACQUEMIN	Professeur, Université de Nantes, France.
	Mlle Élise DUMONT	Maître de conférence, École Normale Supérieure de Lyon, France.
	Mr Xavier ASSFELD	Professeur, Université de Nancy, France.

Résumé

Les méthodes hybrides alliant la mécanique quantique et la mécanique moléculaire (QM/MM) sont des outils adéquats pour traiter des systèmes biologiques. Les phénomènes électroniques souvent étudiés sur des petites molécules ont, dès lors pu être envisagés dans des environnements macromoléculaires. Ce travail explore trois phénomènes électroniques en présence d'un environnement protéique : l'absorption, la capture électronique et les ionisations de cœur. Nous avons employé les derniers développements de la méthode QM/MM *Local Self-Consistent Field* (LSCF) pour traiter la jonction covalente entre la partie QM et la partie MM ainsi que le couplage QM/MM avec le PCM. En premier lieu, nous nous sommes focalisés sur les spectres d'absorption de chromophores présents dans des macromolécules. Nous avons mis au point un couplage entre les méthodes LSCF/MM et PCM pour prendre en compte la polarisation électronique de l'environnement suite à l'absorption d'un photon par le chromophore. Ce modèle, le LSCF/MM:ERS, a été testé et validé sur le spectre d'absorption du complexe de squaraine-tétralactame. Une étude plus poussée a ensuite été réalisée sur une protéine fluorescente en décomposant la longueur d'onde d'absorption maximale en trois contributions physiques. Les effets de la substitution du chromophore ont aussi été évalués. Dans une seconde partie, nous avons étudié la capture électronique par un cyclotide contenant trois ponts disulfures, qui, après irradiation, forment des demi-liaisons caractéristiques (2c-3e). La dernière partie est consacrée à la validation de l'approche de la projection asymptotique. Elle a été réalisée dans le cadre d'une étude sur les ionisations de cœur d'un ensemble de molécules pour, par la suite, étudier les ionisations de cœur de la glycine présentes dans des systèmes de plus en plus complexes jusqu'à la Sérum-Albumine Humaine.

Mots-clé : absorption UV/visible, capture électronique, ionisation de cœur, QM/MM, LSCF.

Abstract

Hybrid methods that combine quantum mechanics and molecular mechanics (QM/MM) provide a near-ideal treatment of biological system reactivity and spectroscopy. Many electronic phenomena often studied on small systems can be now foreseen in macromolecular surroundings. This work considers the treatment with QM/MM tools of three electronic phenomena in biosystem : absorption, electronic attachment and core ionization. Latest developments of the Local Self-Consistent Field formalism (LSCF) have been used to treat delicate covalent junctions between the QM part and the MM part and the coupling of QM/MM and PCM methods. Firstly we have focussed on absorption spectra of chromophores embedded in macromolecules. The combined LSCF/MM and PCM approach have been employed to account for the electronic polarization when the chromophore absorbs one photon. This new method, called LSCF/MM:ERS has been tested and validated with the study of the absorption spectra on the squarain-tetralactam complex. Then, we have considered a fluorescent protein and decomposed the maximum absorption wavelength into several physical contributions. We have also analyzed the substitution effect of the chromophore. Secondly, the electronic capture have been studied on a cyclotide containing three disulfide bonds, which forms characteristic hemi-bond (2c-3e) after irradiation. The last electronic phenomena studied is the core ionization tackled within the framework of the asymptotic projection approach. The latter has been implemented, then tested and validated on a set of molecules. This method have been employed to analyze the specific core ionization of glycine-containing systems of increasing complexity, up to the Human Serum Albumin.

Keywords : UV/visible absorption, electronic capture, core ionization, QM/MM, LSCF.

Remerciements

Au terme de ces trois ans de thèse au sein du laboratoire de Chimie et Biochimie Théoriques (CBT) de Nancy, je tiens à adresser des remerciements à plusieurs personnes.

Je remercie le directeur de l'équipe CBT, monsieur Manuel Ruiz-Lopez (Manolo), pour son accueil, l'ancien directeur de l'école doctorale Lorraine de Chimie et de Physique Moléculaires, monsieur Canet, et les membres du conseil de l'école doctorale qui m'ont autorisée à faire cette thèse.

Il y a quatre ans, j'ai fait un « mémoire » (stage de M2) dans le laboratoire de monsieur Jean-Marie André à Namur. J'ai fait la connaissance du milieu de la recherche et de plusieurs personnes : Eric Perpète (Promoteur), Denis Jacquemin (Co-promoteur), Julien Prémat, Catherine Michaux, Valérie Wathelet et enfin, de Xavier Assfeld lors d'un *workshop* à Barcelone.

Ma venue en France a étonné beaucoup de monde à ce moment-là (famille et ami(e)s). La question du « Pourquoi ? » est revenue régulièrement, et pourtant la réponse était simple : un *challenge* supplémentaire sur le plan scientifique. Avec trois ans de recul, je peux dire, à titre personnel, que c'était aussi pour satisfaire un besoin d'indépendance. Je remercie donc Eric et Denis de m'avoir donné cette opportunité et je remercie particulièrement Xavier de m'avoir aidée à réaliser cette démarche qui était autant scientifique qu'humaine. Xavier Assfeld, en tant que directeur de Thèse, a facilité mon intégration dans l'équipe et m'a apporté le soutien scientifique dont j'avais besoin. En outre, j'ai pu profiter de son expérience et de ses réseaux dans le milieu de la chimie quantique. À l'occasion de conférences et de séminaires, y compris à l'étranger, j'ai rencontré de nombreux chercheurs qui ont contribué à compléter mes connaissances.

Au cours de ces trois ans, il y a eu beaucoup de rencontres avec d'autres thésards, post-doctorants, ATER et maîtres de conférence. J'aimerais particulièrement remercier Pierre-François Loos (T2) et Elise Dumont qui étaient à Nancy durant ma première année de thèse et qui ont

énormément contribué à mon travail scientifique. Je garde de très bons souvenirs d'eux. J'ai gardé d'excellents contacts avec Elise Dumont avec qui j'ai continué à travailler et à échanger presque quotidiennement tant sur le sujet de la thèse que sur ses propres thématiques lyonnaises.

Dans l'équipe CBT, j'ai passé de très bons moments avec Delphine Bas, Eddy Thiriot (TD en Pharmacie), Jenifer Thomas et Marius Retegan mais aussi avec l'équipe de Dynamique des Assemblages Membranaires : Christophe Chipot, François Dehez, Lucie Delemotte, Eva-Maria Krammer, Valeria Nibali, Pepe de la Rosa et enfin Fabien Archambault. Beaucoup d'entre eux sont devenus d'excellents amis, je vous remercie donc tous pour vos conseils et votre convivialité.

Quelques moments inoubliables évocateurs : la semaine à Strasbourg pour le master CIT (2007), la semaine à Paris au 1er Thetis, un séjour en Malaisie avec Laurence et Juan-Luca, mon premier RCTF avec une chouette ambiance à Dinard, cette semaine inoubliable à Barcelone chez Carlos Aleman avec T2 et Julien Préat, la collaboration avec Vitaly Glushkov et les 10 jours ukrainiens avec Xavier, les COTAW 2008 et 2009, ...

Enfin, je souhaiterais remercier toutes les personnes que j'ai rencontrées hors cadre scientifique, c'est-dire celles de la cafet' du second cycle (Maité et Philou), ses adeptes (Christian, Steph, Nico et Ludo). J'ai apprécié les moments de détente autour d'un billard ou d'une bière.

Finalement, j'aimerais remercier tous les membres de ma famille pour leur soutien moral et les relectures attentives de cette thèse.

Merci pour tous ces bons moments passés à Nancy.

Table des matières

Introduction générale	1
I Méthodologie	7
1 Les méthodes <i>ab initio</i> indépendantes du temps	11
1.1 La méthode Hartree Fock	11
1.2 La théorie des perturbations : Møller-Plesset	15
1.3 La théorie de la fonctionnelle de la densité (DFT)	15
1.3.1 Les différents types de fonctionnelles d'échange-corrélation	17
1.3.1.A L'approximation de la densité locale	18
1.3.1.B L'approximation du gradient généralisé	18
1.3.1.C La connexion adiabatique et les fonctionnelles hybrides	19
1.3.2 La prise en compte des effets de dispersion	21
2 Évaluation des transitions électroniques	25
2.1 Les méthodes post Hartree-Fock	25
2.1.1 Interaction de configuration	25
2.1.2 Les méthodes de champ auto-cohérent multi-configurationnel	27
2.2 Les méthodes dépendantes du temps	28
2.2.1 Hartree-Fock dépendant du temps (TD-HF)	28
2.2.2 Théorie de la fonctionnelle de la densité (TD-DFT)	33
2.3 Les effets de la solvatation	36
2.3.1 Principes de base des modèles de continuum	37
2.3.1.A La description de la cavité	38
2.3.1.B La résolution du problème électrostatique	39

2.3.1.B.a	La résolution numérique, ASC	40
2.3.1.B.b	La résolution analytique, MPE	41
2.3.1.C	Les termes non-électrostatiques	42
2.3.1.C.a	Le terme de cavitation	42
2.3.1.C.b	Le terme de dispersion	43
2.3.1.C.c	Le terme de répulsion	43
2.3.2	L'état de non-équilibre	43
2.3.2.A	Les transitions verticales	46
2.3.2.A.a	Le modèle PCM	47
2.3.2.A.b	Le modèle NCM	47
II	Le traitement de systèmes biologiques	50
3	La mécanique moléculaire	52
3.1	La fonction d'énergie potentielle	53
3.2	Les champs de forces classiques	55
3.3	Les champs de forces polarisables	56
4	Les méthodes hybrides QM-MM	58
4.1	Les expressions de l'énergie QM-MM	61
4.1.1	La contribution électrostatique	62
4.1.1.A	Le <i>mechanical embedding</i> (ME)	62
4.1.1.B	L' <i>electrostatic embedding</i> (EE)	62
4.1.1.C	Le <i>polarizable embedding</i> (PE)	63
4.1.2	Les contributions de van der Waals dans la région frontière	63
4.2	Les différentes coupures en QM/MM	64
4.2.1	Les approches de type Link Atom	65
4.2.1.A	L'approche ONIOM	66
4.2.2	Les approches des orbitales localisées	67
4.3	Le Local Self-Consistent Field/MM (LSCF/MM)	68
4.3.1	La modification des équations de Roothaan et l'algorithme	70
4.3.2	Le traitement des orbitales gelées en LSCF	71

III	Les phénomènes électroniques appliqués aux macromolécules	80
5	Absorption UV/visible	84
5.1	La réponse électronique de l'environnement	88
5.1.1	Le couplage LSCF/MM :SCRF	89
5.2	Le complexe de squaraine-tétralactame	91
5.2.1	Méthodologie utilisée	91
5.2.2	Résultats et discussion	93
5.2.2.A	Analyse de la géométrie	93
5.2.2.B	Le spectre d'absorption	95
5.2.3	Conclusion	98
5.3	Les protéines fluorescentes	98
5.3.1	La protéine ECFP	101
5.3.2	Méthode utilisée pour les FP	104
5.3.2.A	Décomposition des propriétés optiques	106
5.3.3	Le chromophore de la ECFP isolé	107
5.3.3.A	L'effet de la coupure en LSCF/MM	108
5.3.4	Spectre d'absorption des deux conformations de la ECFP	111
5.3.4.A	Et les molécules d'eau dans tout cela...	115
5.3.5	La mutation du chromophore	116
5.3.5.A	Comparaison des résultats	117
5.3.5.B	Les différentes contributions de la longueur d'onde d'absorption	120
5.3.6	Conclusion générale sur l'étude des protéines fluorescentes	123
6	La capture électronique	136
6.1	Description de kalata B1 (kB1)	141
6.2	Méthodologie employée pour traiter les cyclotides	143
6.3	Première capture électronique	146
6.3.1	Décomposition de l'AEA en terme d'effets géométriques et électrostatiques	147
6.4	Capture d'un second électron	151
6.5	Conclusions du travail sur kalata B1	156

7 Les ionisations de cœur	158
7.1 L'approche de la projection asymptotique	162
7.1.1 L'implantation	165
7.2 Validation de la projection asymptotique	165
7.2.1 Ionisation de cœur de molécules organiques	166
7.2.2 Application aux acides aminés	171
7.3 Les effets de l'environnement sur la Glycine	175
7.3.1 La mono-, di- et triglycine	175
7.3.2 La Sérum-Albumine Humaine (HSA)	179
7.4 Conclusions	189
Conclusions et perspectives	192
Appendices	201
A Quelques fonctionnelles usuelles	202
A.1 LDA et LSDA	202
A.2 Les GGA	202
B Les acides aminés	204
C Les ionisations de cœur	207
C.1 La Sérum-Albumine Humaine	207
Liste des abréviations	212
Bibliographie	217

Introduction générale

De la chimie vers la biologie

A l'heure actuelle, le domaine de la biologie est en plein essor. Les outils expérimentaux peuvent caractériser des structures de plus en plus complexes et avec une résolution de plus en plus fine. Le prix Nobel de chimie en 2009 a d'ailleurs été attribué à trois chercheurs biologistes¹ parce qu'ils ont établi la structure des ribosomes et leur fonctionnement à l'échelle atomique. L'étude de machinerie de plusieurs milliers d'atomes est devenue commune. Les méthodes cristallographiques permettent actuellement de résoudre des structures de plus en plus complexes, ce qui entraîne une amélioration de la connaissance des fonctions biologiques via la structure atomique. Par conséquent, les méthodes théoriques telles que la dynamique moléculaire (MD) deviennent des outils cruciaux pour évaluer l'activité des protéines dans leur environnement physiologique. Elles permettent de décrire des systèmes énormes et ce, pour un coût calculatoire parfaitement accessible. La MD est basée sur le principe de la mécanique newtonienne et des champs de forces de mécanique moléculaire (MM) décrivant les interactions entre atomes. Cependant, la description explicite des électrons des atomes n'est pas prise en compte et des phénomènes électroniques tels que des transferts de charges, des excitations électroniques ne peuvent être considérés avec de telles méthodes. Celles-ci rencontrent aussi des limites quant à la formation ou la rupture de liaisons, souvent présentes dans le site réactionnel de protéines (transfert de protons, ...).

Pour avoir une description précise de phénomènes électroniques, il est donc nécessaire de passer par des outils quantiques nettement plus coûteux en temps de calcul. Malgré les avancées de ces techniques, il n'est pas encore possible de traiter des systèmes de plus d'une centaine d'atomes, voire un peu plus avec des méthodes semi-empiriques ou DFTB. Une méthode permettant d'allier les avantages de chacun des outils de la chimie théorique est d'utiliser des méthodes hybrides QM-MM (notation générale) alliant la mécanique quantique (QM) et la mécanique mo-

¹Venkatraman Ramakrishnan (Cambridge), Thomas A. Steitz (Yale) et Ada E. Yonath (Israël)

léculaire. Ces méthodes sont apparues il y a 35 ans [1] et n'ont cessé de se développer au fil des années. Actuellement, des simulations Born-Oppenheimer ou Car-Parinello permettent de fournir des trajectoires de systèmes décrits en QM-MM. Les temps de simulation de ces méthodes sont cependant encore assez limités en comparaison aux temps atteints par MD (une dizaine de picosecondes contre quelques micro ou millisecondes). Plus ces échelles de temps augmenteront, plus la dynamique des phénomènes biochimiques sera comprise. De grands efforts sont réalisés pour augmenter les performances et/ou la rapidité des programmes de simulation en les rendant parallélisables.

Dans cette thèse, nous nous sommes intéressés à trois phénomènes électroniques en présence d'un environnement protéique : l'absorption, la capture électronique et les ionisations de cœur.. Ceux-ci seront décrits par des méthodes hybrides QM-MM. La méthode principalement employée est celle développée à Nancy depuis les années quatre-vingt-dix : le Local Self-Consistent Field (LSCF) [2–10]. Nous avons utilisé les derniers développements mis au point par Pierre-François Loos et Nicolas Ferré (cf. Thèses [11, 12]).

Ce manuscrit est composé de trois parties regroupant plusieurs chapitres. La première partie présente les divers outils de la mécanique quantique. Son premier chapitre reprend les méthodes *ab initio* dont nous avons eu besoin telles que la théorie de la fonctionnelle de la densité pour évaluer les spectres d'absorption UV/visible ou la théorie des perturbations employée pour estimer les affinités électroniques adiabatiques. Ce chapitre comprend aussi les fondements de base de la méthode Hartree-Fock. Le second chapitre est dédié aux approches permettant d'évaluer des transitions électroniques, allant des méthodes post Hartree-Fock (interaction de configuration (CI) et des méthodes de champ auto-cohérent multiconfigurationnel (MCSCF)) aux méthodes liées à la réponse linéaire. Les équations des méthodes TD-HF et TD-DFT en ont été dérivées puisque la TD-DFT nous a permis de calculer les longueurs d'onde d'absorption. Finalement, les principes de base des outils théoriques qui prennent en compte la solvatation (PCM et NCM) sont explicités afin de montrer comment ces modèles de solvatation réagissent dans des situations de non-équilibre (par exemple, les transitions électroniques).

La seconde partie concerne le traitement des systèmes biologiques d'un point de vue théorique. Le premier chapitre reprend brièvement les notions principales de la mécanique classique ainsi que les champs de forces que nous avons employés pour décrire les différents systèmes. Le

second chapitre est dédié aux méthodes hybrides alliant la mécanique quantique et la mécanique moléculaire (QM-MM notation générale) et reprend les principes des approches Link Atom (LA) et des approches des orbitales gelées. Le troisième chapitre contient une description détaillée de l'approche LSCF/MM (*Local Self-Consistent Field*) développée au laboratoire de Nancy. Le traitement de la coupure délicate d'une liaison covalente entre la partie d'intérêt (QM) et l'environnement (MM) y est présenté de manière pointilleuse parce que ces récents développements ont été employés lors du traitement du spectre d'absorption des protéines fluorescentes ou encore de l'évaluation des énergies d'ionisation des glycines de l'albumine humaine.

La troisième partie de cette thèse concerne les résultats obtenus au cours de ces trois dernières années avec comme thème principal les phénomènes électroniques de macromolécules.

Un premier chapitre est consacré aux calculs de spectre d'absorption. Afin de prendre en compte la réponse électronique de l'environnement du chromophore, nous avons proposé une méthode simple couplant le QM-MM et un continuum (LSCF/MM:ERS). Ce modèle a été utilisé sur deux systèmes macromoléculaires différents, d'une part, sur le complexe de squaraine-tétralactame et d'autre part, sur les protéines fluorescentes.

Le second chapitre est dédié à la capture électronique de macrocycles contenant des liaisons disulfures. L'affinité électronique adiabatique a été étudiée sur un cyclotide suite à l'addition d'un premier et ensuite d'un second électron.

Finalement, le dernier chapitre (chapitre 7) contient le développement de l'approche de la projection asymptotique (AP) et son implantation dans le code gaussian modifié pour le LSCF/MM. L'AP a ensuite été testée et validée sur un ensemble de molécules organiques. Elle nous a permis d'évaluer les ionisations de cœur des acides aminés et des glycines de la Sérum-Albumine Humaine.

La dernière partie de ce manuscrit présentera les perspectives liées au LSCF ou à la méthodologie employée pour traiter les systèmes biologiques.

Dans la première annexe, nous reprendrons quelques fonctionnelles de la DFT, la seconde annexe contient une table listant tous les acides aminés et leur abréviation (code à trois lettres et code à une lettre). La troisième annexe reprend une table des ionisations de cœur obtenues sur le carbone α de la glycine de l'albumine humaine.

Première partie

Méthodologie

Préambule

CETTE section est inspirée de plusieurs ouvrages de chimie quantique de référence [13–16]. Nous ne donnerons ici qu’un aperçu des techniques théoriques indépendantes du temps et nous nous attarderons plus amplement sur les méthodes dépendantes du temps que nous avons utilisées au cours de cette thèse. Le système d’unités atomiques a été employé tout au long du manuscrit. Une liste des acronymes utilisés se trouve à la fin du manuscrit.

En chimie quantique, il existe trois postulats de base. Le premier est lié au principe d’incertitude de Heisenberg qui stipule qu’il est impossible de déterminer simultanément la position et la vitesse d’une particule. La notion de trajectoire, chère à la dynamique moléculaire, n’a donc plus de signification et est remplacée par la notion de fonction d’onde $\Psi(\mathbf{r}, t)$. La fonction d’onde n’a pas de signification physique ; par contre son carré représente la probabilité de densité de particules. Selon le second postulat, la fonction d’onde est définie au temps t par l’équation de Schrödinger dépendante du temps :

$$H\Psi(\mathbf{r}, t) = i\frac{\partial\Psi(\mathbf{r}, t)}{\partial t} \quad (0.1)$$

Dans cette équation H est l’opérateur Hamiltonien polyélectronique. Cette équation est employée pour le calcul de propriétés moléculaires et est primordiale pour résoudre des problèmes dépendants du temps. Qu’il s’agisse de la résolution de problèmes sur des systèmes statiques, comme le calcul de spectres d’absorption ou sur des systèmes de dynamique *ab initio* tels la dynamique d’Ehrenfest [17], de Born-Oppenheimer [18] ou de Car-Parrinello [19–21].

Pour un système stationnaire, l’équation de Schrödinger s’écrit :

$$H\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \quad (0.2)$$

où E est l’énergie d’un état stationnaire. Le Hamiltonien s’écrit comme la somme des opérateurs énergie cinétique et énergie potentielle. Ceux-ci sont définis dans l’équation ci-dessous où intervient le nombre d’électrons (N) et de noyaux (M) du système, la masse du noyau A (M_A) de

nombre atomique Z_A et les distances entre électrons i et j (r_{ij}), entre électron i et noyau A (r_{iA}) et entre noyaux A et B (r_{AB}).

$$H = - \sum_{i=1}^N \frac{\nabla_i^2}{2} - \sum_{A=1}^M \frac{\nabla_A^2}{2M_A} - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{r_{AB}} \quad (0.3)$$

Les deux premiers termes décrivent les opérateurs d'énergie cinétique des électrons et celui des noyaux. Ensuite, sont décrites respectivement l'attraction coulombienne entre électrons et noyaux, la répulsion électronique et la répulsion nucléaire.

Dans un premier temps, nous présenterons brièvement les méthodes indépendantes du temps telles Hartree-Fock (HF), la théorie des perturbations de Møller Plesset (MP) et la théorie de la fonctionnelle de la densité (*Density Functional Theory*, DFT). Bien que très connus, nous rappellerons les fondements de HF afin de les comparer aux modifications qu'il a été nécessaire d'introduire dans le formalisme LSCF (*Local Self-Consistent Field*) développé à Nancy [2,4,5,12].

Dans le second chapitre, nous développerons de manière plus détaillée les méthodes qui permettent d'évaluer l'énergie de transition électronique. En commençant par les approches CIS ou CASSCF, nous présenterons de manière plus détaillée les équations Hartree-Fock dépendantes du temps (TD-HF) et celles issues de la théorie de la fonctionnelle de la densité dépendante du temps (TD-DFT) que nous avons employée. Les principes de base des approches développées pour prendre en compte la réponse électronique de milieu solvatoé seront présentés afin de considérer ensuite toute la problématique du milieu protéique.

Chapitre 1

Les méthodes *ab initio* indépendantes du temps

1.1 La méthode Hartree Fock

D'APRÈS l'approximation de Born-Oppenheimer, reposant sur la différence de masse d'un électron et d'un noyau, le Hamiltonien polyélectronique de l'Eq. (0.3) peut être simplifié : le second terme peut être négligé et le dernier terme décrivant la répulsion entre noyaux considéré comme constant pour une géométrie donnée¹. Dans l'équation fondamentale de la théorie de Hartree-Fock (HF), chaque électron ressent l'ensemble des autres électrons comme un champ moyen. Le Hamiltonien polyélectronique de l'équation de Schrödinger devient par conséquent un jeu de Hamiltoniens monoélectroniques (h_i),

$$h_i = -\frac{\nabla_i^2}{2} - \sum_A^{\text{noyaux}} \frac{Z_A}{r_{iA}} + \sum_j^{N/2} (2J_j(r_i) - K_j(r_i)) \quad (1.1)$$

Les deux termes monoélectroniques comprenant l'énergie cinétique et l'interaction électrons-noyaux sont souvent repris dans un terme général nommé, le Hamiltonien de cœur h_i^c . Les deux derniers termes de l'Eq. (1.1) sont des termes biélectroniques qui expriment le potentiel ou champ moyen que subit un électron de la part de ses congénères. Cette équation tient compte de

¹La masse d'un électron étant 1836 fois plus légère que la masse d'un noyau, les électrons se déplaceront nettement plus vite que les noyaux. Une bonne approximation est de considérer les électrons d'une molécule comme se déplaçant dans un champ de noyaux fixes.

la répulsion coulombienne électronique (J_j) et du principe de Pauli (K_j , appelé terme d'échange).

$$J_j(1)\phi_i(1) = \int \frac{\phi_j^*(2)\phi_j(2)}{r_{ij}}\phi_i(1)d\tau \quad (1.2)$$

$$K_j(1)\phi_i(1) = \int \frac{\phi_j^*(2)\phi_i(2)}{r_{ij}}\phi_j(1)d\tau \quad (1.3)$$

Seules les interactions répulsives et d'échange « moyennes » entre la particule et le champ de ses congénères sont considérées. Ainsi, la probabilité de trouver simultanément deux électrons de spins opposés au même endroit de l'espace est non nulle ; cette erreur dite de corrélation peut être minimisée par différentes stratégies donnant lieu aux méthodes post Hartree-Fock et celles s'appuyant sur la théorie de la fonctionnelle de la densité (*Density Functional Theory*, DFT).

L'énergie du système est donnée par la relation :

$$E_0 = \langle \Psi_0 | H_0 | \Psi_0 \rangle \quad (1.4)$$

$$|\Psi_0\rangle = -|\psi_1\psi_2\psi_2\dots\rangle \quad (1.5)$$

où, suite à l'approximation orbitale, la fonction d'onde multiélectronique du système est représentée comme le produit de fonctions d'onde monoélectroniques ψ_i . Par le biais du Fockien, il suffit de résoudre le système linéaire suivant *via* les équations de Roothaan pour déterminer l'énergie propre :

$$h_i\psi_i = \epsilon_i\psi_i \quad (1.6)$$

Dans la suite, nous ne considérerons qu'un système à couches fermées et nous utiliserons, par conséquent, le formalisme restreint ou RHF (*Restricted Hartree-Fock*), pour lequel deux électrons de spin différents d'une même orbitale moléculaire (OM) possèdent une partie d'espace identique ².

Conformément à l'approximation LCAO (*Linear Combination of Atomic Orbitals*), nous allons introduire une base de K orbitales atomiques (OA), définie comme $\{\phi_\mu\}_{1\leq\mu\leq K}$:

$$\psi_i = \sum_{\mu}^N c_{\mu i}\phi_{\mu} \quad (1.7)$$

²Dans le cas de systèmes à couches ouvertes, le formalisme UHF (*Unrestricted Hartree-Fock*) peut être utilisé.

Sur cette base d'OA, l'élément $\mu\nu$ de la matrice de Fock F s'exprime :

$$F_{\mu\nu} = \int \phi_{\mu}^*(1)h(1)\phi_{\nu}(1)d\mathbf{r}_1 \quad (1.8)$$

$$= \int \phi_{\mu}^*(1)h^c(1)\phi_{\nu}(1)d\mathbf{r}_1 + \sum_a^{N/2} \int \phi_{\mu}^*(1) [2J_a(1) - K_a(1)] \phi_{\nu}(1)d\mathbf{r}_1 \quad (1.9)$$

$$= H_{\mu\nu}^c + \sum_a^{N/2} 2(\mu\nu|aa) - (\mu a|a\nu) \quad (1.10)$$

En insérant l'approximation de la combinaison linéaire, nous obtenons :

$$F_{\mu\nu} = H_{\mu\nu}^c + \sum_a^{N/2} \sum_{\lambda\sigma} c_{\lambda a} c_{\sigma a}^* [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] \quad (1.11)$$

$$= H_{\mu\nu}^c + \sum_{\lambda\sigma} P_{\lambda\sigma}^T \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu) \right] \quad (1.12)$$

$$= H_{\mu\nu}^c + G_{\mu\nu} \quad (1.13)$$

avec les intégrales biélectroniques :

$$(\mu\nu|\lambda\sigma) = \iint \phi_{\mu}^*(1)\phi_{\nu}(1)\frac{1}{r_{12}}\phi_{\lambda}^*(2)\phi_{\sigma}(2)d\mathbf{r}_1d\mathbf{r}_2 \quad (1.14)$$

$$(\mu\sigma|\lambda\nu) = \iint \phi_{\mu}^*(1)\phi_{\nu}(1)\frac{1}{r_{12}}\phi_{\lambda}^*(2)\phi_{\sigma}(2)d\mathbf{r}_1d\mathbf{r}_2 \quad (1.15)$$

$$(1.16)$$

L'élément $\mu\nu$ de la matrice densité totale du système \mathbf{P}^T dans le formalisme RHF est défini comme :

$$P_{\mu\nu}^T = 2 \sum_i^N c_{\mu i} c_{\nu i}^* \quad (1.17)$$

La condition de stationnarité de l'énergie par rapport aux coefficients $\{c_{\mu i}\}_{\substack{1 \leq \mu \leq N \\ 1 \leq i \leq N}}$

$$\frac{\delta E}{\delta c_{\mu i}} = 0 \quad (1.18)$$

conduit aux équations de Roothaan :

$$\sum_{\nu}^N F_{\mu\nu} c_{\nu i} = \sum_{\nu}^N S_{\mu\nu} c_{\nu i} \varepsilon_i \quad (1.19)$$

ou sous forme matricielle,

$$\mathbf{F} \cdot \mathbf{C} = \mathbf{S} \cdot \mathbf{C} \cdot \mathbf{E} \quad (1.20)$$

La matrice \mathbf{S} représente la matrice de recouvrement telle que $S_{\mu\nu} = \int \phi_\mu^*(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)d\mathbf{r}_1$. La matrice \mathbf{C} regroupe les coefficients des OA dans les OM tel que : $(\mathbf{C})_{\mu i} = c_{\mu i}$. L'énergie électronique du système s'écrit donc, sur la base des AOs :

$$E = \sum_{\mu\nu}^N P_{\mu\nu}^T H_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma}^N P_{\mu\nu}^T P_{\lambda\sigma}^T G_{\mu\nu\lambda\sigma} \quad (1.21)$$

Les fonctions de base ne sont pas orthogonales entre elles (mais bien normales). Il est donc nécessaire d'orthogonaliser les fonctions de base afin que la résolution des équations de Roothaan devienne un problème aux valeurs propres. On introduit alors une matrice d'orthogonalisation de la base d'OA \mathbf{X} telle que :

$$\mathbf{X}^\dagger \cdot \mathbf{S} \cdot \mathbf{X} = \mathbb{I} \quad (1.22)$$

où \mathbb{I} est la matrice identité. Dans le formalisme SCF, la matrice d'orthogonalisation est souvent celle de Löwdin : $\mathbf{X} = \mathbf{S}^{-1/2}$. En substituant \mathbf{C} par $\mathbf{X}\mathbf{C}'$ et en multipliant par la gauche de l'Eq. (1.20) par \mathbf{X}^\dagger (matrice adjointe de la matrice \mathbf{X}), cela mène à un problème standard aux valeurs propres hermitiques :

$$\mathbf{F}' \cdot \mathbf{C}' = \mathbf{C}' \cdot \mathbf{E} \quad (1.23)$$

avec

$$\mathbf{F}' = \mathbf{X}^\dagger \cdot \mathbf{F} \cdot \mathbf{X} \quad (1.24)$$

La résolution de l'Eq. (1.23) s'effectue par la diagonalisation de la matrice \mathbf{F}' dépendante des coefficients recherchés \mathbf{C} . La méthode de résolution auto-cohérente ou SCF (*Self-Consistent Field*) est alors employée. Son algorithme est résumé ci-dessous :

1. Calcul des intégrales $S_{\mu\nu}$, $H_{\mu\nu}^c$ et $(\mu\nu|\lambda\sigma)$
2. Estimation de la matrice densité \mathbf{P} .
3. Construction de la matrice de Fock $\mathbf{F} = \mathbf{H}^c + \mathbf{G}$.
4. Expression de \mathbf{F} dans la base orthogonale : $\mathbf{F}' = \mathbf{X}^\dagger \cdot \mathbf{F} \cdot \mathbf{X}$ ($\mathbf{X} = \mathbf{S}^{-1/2}$).
5. Diagonalisation de \mathbf{F}' .
6. Transformation des MOs dans la base non-orthogonale : $\mathbf{C} = \mathbf{X} \cdot \mathbf{C}'$.
7. Calcul de la nouvelle matrice densité \mathbf{P} et calcul de l'énergie $E = \frac{1}{2}Tr [\mathbf{P}(\mathbf{H}^c + \mathbf{F})]$.
8. Test de convergence sur E et/ou \mathbf{P} . Si non satisfait, retour à 3.

1.2 La théorie des perturbations : Møller-Plesset

UNE première méthode post HF pour prendre en compte la corrélation électronique est la théorie des perturbations de Møller Plesset. Elle est basée sur la théorie des perturbations de Rayleigh-Schrödinger qui consiste à développer en série de Taylor l'énergie du système. L'opérateur Hamiltonien du système H est décrit comme la somme d'un hamiltonien de référence $H^{(0)}$ et d'un opérateur de perturbation V .

$$H = H^{(0)} + V \quad (1.25)$$

Dans la théorie des perturbations de Møller-Plesset à l'ordre n (MP n), l'opérateur hamiltonien de référence est donné par la relation [22] :

$$H^{(0)} = \sum_i^N h(i) \quad (1.26)$$

où N est le nombre de fonctions de base et h_i est défini par l'Eq. (1.1). En se limitant à l'ordre 2 dans le développement (MP2), on aboutit à l'Eq. (1.27)

$$E^{\text{MP2}} = E_{(0)} + E^{(2)} \quad (1.27)$$

avec $E_{(0)} = E^{\text{HF}}$ et $E^{(2)}$ la correction de l'énergie à l'ordre 2 qui est donnée par la relation suivante :

$$E^{(2)} = \langle \Psi^{(0)} | V | \Psi^{(1)} \rangle \quad (1.28)$$

$$= \frac{1}{4} \sum_{i,j>i}^{\text{occ}} \sum_{a,b>a}^{\text{vir}} \frac{[(ij|ab) - (ia|jb)]^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad (1.29)$$

Ainsi, l'énergie électronique totale MP2 est donnée par l'énergie de HF et par la correction de second ordre MP : $E^{\text{MP2}} = E^{\text{HF}} + E^{(2)}$. La solution de l'équation MP à l'ordre un donne l'énergie de Hartree-Fock. La première correction de perturbation au-delà du traitement Hartree-Fock qui ne s'annule pas est le terme d'énergie de second ordre (par le théorème de Brillouin).

1.3 La théorie de la fonctionnelle de la densité (DFT)

CETTE section reprend les principes de base de la théorie de la fonctionnelle de la densité (DFT), méthode introduisant la corrélation *via* une fonctionnelle d'échange et de corrélation. Comme son nom l'indique, la DFT est basée sur la densité des particules $[\rho(\mathbf{r})]$, et non sur la

fonction d'onde pour décrire l'état fondamental du système. Toutefois, le formalisme mis en place *in fine* est proche de celui de Hartree-Fock. L'énergie totale peut être divisée en trois parties, l'énergie cinétique T , l'attraction entre noyaux et électrons, V_{ne} , et la répulsion électron-électron, V_{ee} . Similairement à la théorie de Hartree-Fock, la répulsion électron-électron est séparée en deux termes : le terme de Coulomb $J[\rho(\mathbf{r})]$ et le terme d'échange $K[\rho(\mathbf{r})]$. L'énergie d'attraction entre les électrons et le noyau peut être exprimée en fonction de la densité électronique :

$$V_{ne}[\rho(\mathbf{r})] = \sum_A^{\text{noyaux}} \int \frac{-Z_A}{|\mathbf{r} - \mathbf{r}_A|} \rho(\mathbf{r}) d\mathbf{r} \quad (1.30)$$

et de la même manière, le terme de répulsion coulombienne devient :

$$V_{ee} = J[\rho(\mathbf{r})] = \frac{1}{2} \iint \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (1.31)$$

La DFT est construite sur les deux théorèmes de Hohenberg et Kohn. Le premier stipule que dans l'état fondamental, une relation directe entre la densité électronique et le potentiel nucléaire est établie, ainsi qu'entre l'opérateur Hamiltonien et l'énergie. Le second *prédit* la densité d'un système. Dans celui-ci, Hohenberg et Kohn montrent que la densité obéit au principe variationnel : $E[\rho'(\mathbf{r})] \geq E[\rho_0(\mathbf{r})]$, avec $\rho_0(\mathbf{r})$ la densité exacte de l'état fondamental et $\rho'(\mathbf{r})$ une densité d'essai à laquelle correspond un potentiel V'_{ext} , une seule fonction d'onde ψ' et un Hamiltonien H' . Ces deux théorèmes indiquent qu'à partir de la densité, nous pouvons déterminer le Hamiltonien, la fonction d'onde et donc l'énergie ; mais aucune suggestion n'est faite quant à l'utilisation de la densité pour déterminer l'énergie et ce, sans avoir recours à la fonction d'onde.

Une première approche solutionnant ces limites est apparue en 1965 avec le modèle de Kohn et Sham. Par analogie à la théorie HF, les électrons sont traités comme des particules indépendantes se mouvant dans le champ moyen créé par les autres électrons ; cependant, cette fois la corrélation électronique est incluse à l'aide d'une fonctionnelle d'échange-corrélation. La DFT est donc une théorie exacte ; et qui le reste à condition que la fonctionnelle d'échange-corrélation soit aussi définie de manière exacte. La fonctionnelle de Hohenberg-Kohn est alors introduite :

$$F_{\text{HK}}[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + J[\rho(\mathbf{r})] + E_{\text{xc}}[\rho(\mathbf{r})] \quad (1.32)$$

T_s est l'énergie cinétique d'un système sans interaction, $J[\rho(\mathbf{r})]$ est le terme habituel de répulsion électronique. E_{xc} évalue la différence entre la répulsion électrostatique classique et « quantique », mais aussi la différence d'énergie cinétique entre un système fictif sans interaction et un système

réel.

$$E_{xc}[\rho(\mathbf{r})] = \Delta T[\rho(\mathbf{r})] + \Delta J[\rho(\mathbf{r})] \quad (1.33)$$

Notons qu'en pratique, la plupart des fonctionnelles modernes ne calculent pas explicitement cette correction cinétique. En effet, beaucoup incluent des paramètres empiriques et introduisent une correction de l'énergie cinétique basée sur des résultats expérimentaux. En considérant la fonctionnelle de Hohenberg et Kohn, l'équation de l'énergie totale électronique peut être exprimée comme suit :

$$E[\rho(\mathbf{r})] = \int \rho(\mathbf{r})V_{ne}(\mathbf{r})d\mathbf{r} + F_{HK}[\rho(\mathbf{r})] \quad (1.34)$$

Cette équation justifie le second théorème de Hohenberg-Kohn qui stipule que la densité obéit au théorème des variations. Kohn et Sham considèrent un système fictif de particules sans interaction, supposant ainsi que ce système a la même densité et la même énergie que le système réel. Pour s'en assurer, les particules se meuvent dans un potentiel externe $v_{eff}[\rho(\mathbf{r})]$, sans toutefois interagir. Dès lors, l'expression de l'énergie est simplifiée :

$$E[\rho(\mathbf{r})] = T_s[\rho(\mathbf{r})] + \int v_{eff}[\rho(\mathbf{r})]\rho(\mathbf{r})d\mathbf{r} \quad (1.35)$$

Cette dernière équation définit le potentiel effectif qui est tiré de la dérivée de la fonctionnelle de Hohenberg-Kohn. Le potentiel effectif est défini par :

$$v_{eff}[\rho(\mathbf{r})] = V_{ne}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}' + V_{xc} \quad (1.36)$$

$$V_{xc} = \frac{\partial E_{xc}[\rho(\mathbf{r})]}{\partial \rho(\mathbf{r})} \quad (1.37)$$

où V_{xc} est le potentiel d'échange-corrélation associé à l'énergie E_{xc} , exprimée comme :

$$E_{xc}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\epsilon_{xc}[\rho(\mathbf{r})]d\mathbf{r} \quad (1.38)$$

Dans l'approche originelle de Kohn et Sham, l'énergie d'échange-corrélation par électron ou par particule (ϵ_{xc}) est fonction de la densité dans un gaz uniforme d'électron. Elle est définie comme la somme des contributions d'échange et de corrélation : $\epsilon_{xc} = \epsilon_x + \epsilon_c$

1.3.1 Les différents types de fonctionnelles d'échange-corrélation

Jusqu'ici, nous avons présenté le principe de la DFT, méthode exacte. En pratique, on approche l'énergie d'échange-corrélation par différentes fonctionnelles que nous allons expliciter brièvement dans cette section.

1.3.1.A L'approximation de la densité locale

La méthode de la LDA (*Local Density Approximation*) fait partie de la première « génération » de fonctionnelles. Cette méthode est basée sur l'utilisation du modèle du gaz uniforme d'électrons où la densité électronique est constante dans tout l'espace. Elle postule qu'en chaque point \mathbf{r} d'une distribution électronique inhomogène où la densité est $\rho(\mathbf{r})$, les valeurs de E_x et de $\epsilon_x[\rho(\mathbf{r})]$ se comportent identiquement au gaz uniforme d'électrons. Dès lors, la densité électronique au voisinage de \mathbf{r} est remplacée par une densité électronique constante qui a la même valeur qu'en \mathbf{r} . Cependant, cette densité électronique est différente en tout point de l'espace. L'énergie d'échange est donnée par la formule de Dirac. Par particule, elle vaut :

$$\epsilon_x^{\text{LDA}}[\rho(\mathbf{r})] = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \rho^{1/3}(\mathbf{r}) = -C_x \rho^{1/3}(\mathbf{r}) \quad (1.39)$$

La méthode X_α de Slater peut être considérée comme une méthode LDA où l'énergie de corrélation a été négligée et où l'énergie d'échange est ajustée par un paramètre α ($\alpha = \frac{2}{3}$ (LDA), 1 (Slater) ou $\frac{3}{4}$ (Gáspár)). La LDA se justifie pour deux cas limites : lorsque la densité électronique varie lentement ou lorsque la densité est très élevée. Les résultats du test G2 (calcul d'énergies d'atomisation de 50 petites molécules comportant des éléments de la seconde et de la troisième période du tableau de Mendeleïev) donnent un écart moyen de +36 kcal.mol⁻¹ par rapport aux données expérimentales [23]. Les énergies de liaison calculées par la méthode de la LDA sont trop importantes mais cet écart représente une grande amélioration par rapport aux résultats HF, entâchés eux d'une erreur de -78 kcal.mol⁻¹ ! Avec la LDA, les molécules sont modélisées par un nuage d'électrons uniforme, ce qui conduit en pratique à une surestimation de la délocalisation.

Remarquons que si l'on tient compte de la polarisation de spin, la LDA devient LSDA pour *Local Spin Density Approximation*. Les fonctionnelles de Slater, de la méthode du X_α et celle décrite par Vosko, Wilk et Nusair (VWN(5)) [24] sont des exemples de fonctionnelles LDA ou LSDA (voir Annexe A.1).

1.3.1.B L'approximation du gradient généralisé

Dans un système moléculaire, la densité électronique n'est pas spatialement uniforme, ce qui limite évidemment l'usage de la LDA/LSDA. Une façon de corriger cet effet est de rendre les fonctionnelles d'échange et de corrélation dépendantes du gradient de la densité locale : on obtient des fonctionnelles de seconde « génération » GGA (*Generalized Gradient Approximation*)

qui reflètent une certaine anisotropie de l’enveloppe de Slater. La plupart de ces fonctionnelles GGA sont construites en ajoutant un terme correctif à la fonctionnelle LDA :

$$\epsilon_{x/c}^{\text{GGA}}[\rho(\mathbf{r})] = \epsilon_{x/c}^{\text{LSDA}}[\rho(\mathbf{r})] + \Delta\epsilon_{x/c} \left[\frac{|\nabla\rho(\mathbf{r})|}{\rho^{3/4}(\mathbf{r})} \right] \quad (1.40)$$

Les corrections les plus répandues sont celles de A. D. Becke en 1988 [25] pour l’échange et celles de J. P. Perdew [26, 27] ou de Lee-Yang-Parr [28] pour la corrélation. A titre d’illustration, nous donnons quelques expressions pour chacune de ces corrections en Annexe A.2.

Dans la pratique, les fonctionnelles GGA donnent de meilleurs résultats que les fonctionnelles LDA mais ne conduisent pas à une description précise de toutes les propriétés des molécules. C’est notamment le cas des spectres UV/visible pour lesquels les énergies de transition électronique restent sous-estimées. Il existe également des fonctionnelles *meta*-GGA, où la fonctionnelle d’échange-corrélation dépend, en plus de la densité et de son gradient, de la dérivée seconde de la densité électronique (notamment). Des exemples de ce type de *meta*-GGA sont celles élaborées par van Voorhis et Scuseria : la fonctionnelle VSXC [29], celle de Handy [30] et de TPSS [31]. Cette dernière fonctionnelle donne de très bons résultats pour décrire la liaison hydrogène mais ne décrit pas correctement les angles de torsion des molécules présentant des systèmes π conjugués [32].

1.3.1.C La connexion adiabatique et les fonctionnelles hybrides

Depuis les années 90, une nouvelle approche est apparue, fournissant des énergies, des structures et des propriétés moléculaires en meilleur accord avec l’expérience que les LDA et les GGA. Cette approche combine les traitements Hartree-Fock et ceux de la DFT sur les effets d’échange (et de corrélation) aux travers des fonctionnelles hybrides. La connexion adiabatique est un changement qui convertit un système de référence « sans interaction » en un système avec interactions. On peut montrer que l’énergie d’échange et de corrélation peut être déterminée comme :

$$E_{xc} = \int_0^1 \langle \psi(\lambda) | V_{xc}(\lambda) | \psi(\lambda) \rangle d\lambda \quad (1.41)$$

où λ décrit l’ampleur de l’interaction électronique, la force de couplage électronique qui varie entre deux cas limites. Lorsque $\lambda = 0$, l’équation correspond à la valeur de l’énergie d’échange HF du système sans aucune interaction entre les électrons mais, calculée avec les orbitales de Kohn-Sham (c’est ce que l’on nomme souvent l’échange « exact »). Il n’y a pas, par conséquent,

d'énergie de corrélation. Pour $\lambda = 1$, nous avons un système réel en interaction complète. La totalité de l'échange-corrélation est décrite par une fonctionnelle DFT. L'intégration revient à introduire une part de l'échange exact dans la fonctionnelle de l'énergie afin de remédier au défaut de la correction GGA du modèle du gaz uniforme d'électrons.

L'intégration entre les deux systèmes limites se passe à densité constante et à configuration électronique fixe, ce qui est à l'origine du terme « connexion adiabatique ». L'Eq. (1.41) est le fondement des fonctionnelles dites hybrides de type ACM*i* (*Adiabatic Connexion Model*) avec *i* le nombre de paramètres empiriques introduits. Les hybrides ACM1 utilisent un seul paramètre pour corriger la GGA.

$$E_{xc}^{\text{ACM1}} = (1 - \lambda_i) E_x^{\text{DFT}} - \lambda_i E_x^{\text{HF}} + E_c^{\text{DFT}} \quad (1.42)$$

Le paramètre λ_i est souvent semi-empirique : il est ajusté par les auteurs des différentes hybrides pour que ces dernières s'accordent au mieux avec l'expérience. On citera comme exemple B1PW91, B1LYP, PBE0 [33, 34], ... Dans la dernière, le pourcentage d'échange exact (HF) est de 25% et ce paramètre de mélange entre les parties DFT et HF est fixé uniquement par des considérations théoriques : on parle d'ACM0.

$$E_{xc}^{\text{PBE0}} = \frac{3}{4} E_x^{\text{PBE}} + \frac{1}{4} E_x^{\text{HF}} + E_c^{\text{PBE}} \quad (1.43)$$

Pour les hybrides ACM3, les corrections GGA sur l'échange et sur la corrélation de fonctionnelles LDA apparaissent clairement. Ainsi pour déterminer la valeur de l'énergie d'échange-corrélation, on ajoute une correction GGA (λ_2 et λ_3) aux énergies d'échange et de corrélation LDA et un certain pourcentage d'énergie d'échange exacte (λ_1), les paramètres empiriques, λ_i , étant choisis pour optimiser la méthode :

$$E_{xc} = E_{xc}^{\text{LDA}} + \lambda_1 (E_x^{\text{HF}} - E_x^{\text{LDA}}) + \lambda_2 \Delta E_x^{\text{GGA}} + \lambda_3 \Delta E_c^{\text{GGA}} \quad (1.44)$$

Becke, en 1993, [35] a été le premier à inclure trois paramètres et à définir une fonctionnelle du type ACM3. Il a utilisé sa fonctionnelle GGA de 1988, Eq. (A.2) en annexe, pour la partie d'échange et la fonctionnelle de Perdew-Wang [27] pour la partie de corrélation. Les valeurs de λ_1 , λ_2 et λ_3 valent respectivement 0.20, 0.72 et 0.81. Ces paramètres sont tirés de résultats expérimentaux (chaleur d'atomisation, électroaffinité, potentiel d'ionisation, ...) et optimisés par la méthode des moindres carrés. La fonctionnelle la plus utilisée actuellement est une autre ACM3 nommée B3LYP où la fonctionnelle de Perdew-Wang, introduite dans l'ACM3 de Becke en 1993,

est remplacée par la fonctionnelle de Lee-Yang-Parr (Annexe A.4) pour la partie de corrélation. Généralement, le paramètre essentiel reste celui qui indique le pourcentage d'échange exact HF dans la fonctionnelle, c'est-à-dire λ_1 .

Récemment, une nouvelle sous-classe de fonctionnelles hybrides a été développée par différents groupes ; elles sont dites à portées séparées (*Range-Separated Hybrids*, RSH) [36]. En effet, dans certains cas spécifiques (transfert de charges) les hybrides conventionnelles ne parviennent pas à traiter correctement les propriétés moléculaires. Cette nouvelle série d'hybrides inclut ainsi un pourcentage d'échange exact qui n'est plus constant sur la distance inter-électronique. Pour cela, l'opérateur de répulsion électronique, $\frac{1}{r_{12}}$, a été scindé en deux parties à l'aide de la fonction erreur [37–39],

$$\frac{1}{r_{12}} = \frac{1 - \operatorname{erf}(\mu r_{12})}{r_{12}} + \frac{\operatorname{erf}(\mu r_{12})}{r_{12}} \quad (1.45)$$

où μ est un paramètre d'atténuation. Une généralisation de l'Eq. (1.45) a été proposée par Yanai, Tew et Handy [40,41], qui ont introduit deux nouveaux paramètres : a et b .

$$\frac{1}{r_{12}} = \frac{1 - [a + b \operatorname{erf}(\mu r_{12})]}{r_{12}} + \frac{a + b \operatorname{erf}(\mu r_{12})}{r_{12}} \quad (1.46)$$

La Figure (1.1) décrit plus explicitement les différences entre les fonctionnelles hybrides traditionnelles et les hybrides à longue portée. Lorsque les paramètres de l'Eq. (1.46) sont $a = 0.19$, $b = 0.46$ et $\mu = 0.33$, cette équation est celle décrivant la fonctionnelle CAM-B3LYP (*Coulomb attenuated Model-B3LYP*). À courte distance, l'échange exact est décrit de la même manière que dans B3LYP ; par contre, à longue distance, le comportement asymptotique n'est pas respecté [40–42]. Hirao et ses collaborateurs ont proposé LC-PBE [43] qui peut s'appliquer à toutes les fonctionnelles hybrides pures (combinaisons de GGA pour l'échange et la corrélation 1 :1). Elle est caractérisée par un μ de 0.33. La fonctionnelle LC- ω PBE de Vydrov et Scuseria repose sur un paramètre d'atténuation plus important ($\mu = 0.40$) [44]. Ces deux dernières hybrides à longue portée reproduisent correctement le comportement asymptotique pour $r_{12} \rightarrow \infty$. De nombreux papiers ont été publiés récemment pour tester ces nouvelles fonctionnelles, nous invitons le lecteur à les consulter [45–51].

1.3.2 La prise en compte des effets de dispersion

Les forces intermoléculaires, souvent nommées forces de van der Waals, correspondent à des interactions entre molécules à courtes distances (d'environ 3 à 8 Å) qui ne modifient pas la nature

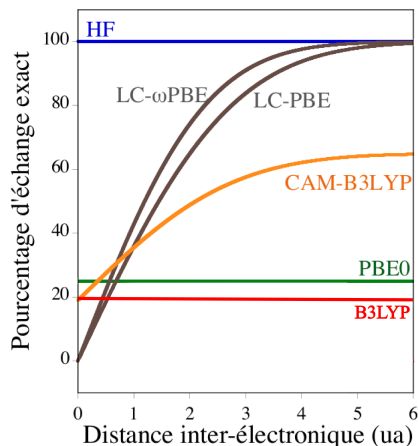


Figure 1.1: Comparaison des fonctionnelles hybrides conventionnelles et des hybrides modifiées à longue portée.

des molécules, mais qui produisent une force totale d'attraction intermoléculaire responsable d'une stabilité non-négligable.

Les forces de London ou effet de dispersion sont des forces faibles intermoléculaires créées par des dipôles instantanés : le dipôle instantané d'une molécule A induit un dipôle instantané sur une molécule B et interagit avec lui. Ces forces représentent en général la plus importante composante des forces de Van der Waals (entre 0.5 et 40 kcal.mole⁻¹). Elles apparaissent lorsque la densité électronique d'une molécule n'est pas équitablement répartie autour de celle-ci, ce qui crée ainsi un léger moment dipolaire. Ces moments dipolaires instantanés varient très rapidement au cours du temps. Ainsi à chaque distribution inhomogène se crée un moment dipolaire instantané qui peut interagir avec les moments dipolaires qu'il induit sur les molécules voisines et une force, dite de London, s'exerce alors entre les molécules.

Les forces de Keesom résultent quant à elle d'une interaction intermoléculaire entre dipôles permanents. Et, enfin lorsqu'une molécule A possède un dipôle permanent, elle peut induire un dipôle instantané sur une molécule B, cette interaction est due aux forces de Debye.

Les forces de London, étant la contribution majoritaire dans les forces de vdW, jouent un rôle important au niveau atomique (repliement de protéines, empilement des bases de l'ADN, reconnaissance moléculaire, ...) dans les systèmes biologiques (ADN, protéine) mais aussi dans

des systèmes chimiques *via* des interactions $\pi - \pi$, T-shape, . . . Théoriquement, les contributions électrostatiques et d'échange peuvent être décrites, comme nous l'avons vu, à l'aide d'un champ moyen (HF) tandis que la dispersion est un effet purement lié à la corrélation électronique entre deux molécules. Dans les méthodes dites de fonction d'onde, la dispersion est prise en compte par le CI (Interaction de configuration) via les di-excitations (Théorème de Nesbet). Des méthodes comme MP2 peuvent décrire cette dispersion mais elles surestiment systématiquement les énergies d'interactions et sous-estiment les distances entre molécules de systèmes $\pi - \pi$. Quant à la DFT, elle échoue complètement pour décrire correctement les interactions dispersives de dimères [52–54].

Récemment, la dispersion a été introduite dans la DFT par Grimme [55] et Head-Gordon dans MP2 [56, 57] en corrigeant l'énergie totale. Ces approches traitent la dispersion à l'aide d'une correction empirique de l'énergie en $C_6^{ij} R_{ij}^{-6}$ où R_{ij} est la distance interatomique et C_6^{ij} sont les coefficients de dispersion. L'énergie totale en DFT, par exemple, devient :

$$E_{\text{DFT-D}} = E_{\text{DFT}} + E_{\text{disp}} \quad (1.47)$$

où l'énergie de dispersion est définie empiriquement par la loi :

$$E_{\text{disp}} = -S_6 \sum_i^{N-1} \sum_{j=i+1}^N \frac{C_6^{ij}}{R_{ij}^6} f^{\text{att}}(R_{ij}) \quad (1.48)$$

$$f^{\text{att}}(R_{ij}) = \frac{1}{1 + e^{-D(R_{ij}/R_0 - 1)}} \quad (1.49)$$

où S_6 et D sont deux paramètres optimisés respectivement en accord avec la fonctionnelle utilisée et avec la fonction d'atténuation ($f^{\text{att}}(R_{ij})$). Cette dernière est nécessaire pour empêcher la divergence à courte distance. N représente le nombre d'atomes, R_0 est la somme des rayons atomiques de vdW des atomes i et j . Les coefficients C_6^{ij} sont obtenus en faisant la moyenne des valeurs atomiques. Les géométries et énergies obtenues avec la DFT-D sont en excellent accord avec les méthodes post-HF incluant la dispersion. De nombreuses publications sont parues sur ce sujet [58–63], notamment sur les interactions entre paires de bases de l'ADN, dimères polyaromatiques, . . .

Chapitre 2

Évaluation des transitions électroniques

2.1 Les méthodes post Hartree-Fock

NOUS présentons ici brièvement quelques méthodes qui sont parfois utilisées dans la littérature pour évaluer les spectres d'absorption [64].

2.1.1 Interaction de configuration

Les équations d'interaction de configuration (CI) sont basées sur les équations de HF comme point de départ. La fonction d'onde électronique CI est construite à partir d'une combinaison linéaire de fonctions d'onde (exprimées comme des déterminants de Slater) de l'état fondamental (EF) et des états « excités » (EE) obtenus en déplaçant un ou plusieurs électron(s) des orbitales occupées (i, j, k, \dots) de l'état fondamental aux orbitales virtuelles (a, b, c, \dots). Les coefficients de la combinaison linéaire de fonctions d'onde sont optimisés suivant le principe variationnel. Dans un calcul CI, chaque configuration est donc exprimée avec les orbitales moléculaires décrivant l'EF, qui sont elles-même une combinaison linéaire de fonctions de base atomiques. Le nombre d'intégrales à calculer est donc extrêmement important. Lors d'un calcul *full*-CI, la fonction d'onde comprendra toutes les permutations entre les N électrons et toutes les orbitales virtuelles. C'est le traitement le plus complet mais, aussi le plus coûteux en terme de temps de calcul. Habituellement, il n'est pas nécessaire de prendre en compte toutes les orbitales. Les orbitales de cœur peuvent être négligées (*Frozen Core Approximation*). Les transitions mono- (et di-) électroniques étant les plus fréquentes, elles suffisent souvent à déterminer le spectre électronique,

on parlera de CIS(D).

Nous allons dériver l'énergie CIS, nous verrons par la suite que le CIS est un cas particulier des méthodes dépendantes du temps. La fonction d'onde CIS se développe comme suit

$$\Psi_{\text{CIS}}(\mathbf{r}) = \Psi_0 + \sum_{ia} c_i^a \Psi_i^a(\mathbf{r}) \quad (2.1)$$

En substituant cette fonction d'onde dans l'équation indépendante du temps de Schrödinger (Eq. (0.2) et en multipliant à gauche par $\langle \Psi_j^b |$ on obtient la projection sur l'espace des déterminants mono-excités

$$\langle \Psi_j^b | H | \Psi_i^a \rangle c_i^a = E_{\text{CIS}} \sum_{ia} c_i^a \delta_{ij} \delta_{ab} \quad (2.2)$$

avec

$$\langle \Psi_j^b | H | \Psi_i^a \rangle = (E_0 + \epsilon_a - \epsilon_i) \delta_{ij} \delta_{ab} + (ia||jb) \quad (2.3)$$

Le dernier terme représente les termes biélectroniques constitués du terme de Coulomb et du terme d'échange. On obtient une équation qui détermine les énergies de mono-excitations électroniques (ω_{CIS}) avec $\omega_{\text{CIS}} = E_{\text{CIS}} - E_0$:

$$\sum_{ia} [(\epsilon_a - \epsilon_i) \delta_{ij} \delta_{ab} + (ia||jb)] c_i^a = \omega_{\text{CIS}} \sum_{ia} c_i^a \delta_{ij} \delta_{ab} \quad (2.4)$$

Cette dernière équation est souvent écrite sous forme matricielle :

$$\mathbf{A}\mathbf{X} = \omega\mathbf{X} \quad (2.5)$$

où la matrice \mathbf{A} représente le Hamiltonien dans l'espace des déterminants mono-excités, ω est la matrice diagonale définissant les énergies de mono-excitations et \mathbf{X} est la matrice des coefficients d'expansion CIS. Il suffit donc de résoudre l'équation séculaire suivante pour définir les énergies d'excitations et les vecteurs propres correspondant aux coefficients de l'expansion (Eq. 2.1), en diagonalisant la matrice \mathbf{A} :

$$(\mathbf{A} - \omega)\mathbf{X} = 0 \quad (2.6)$$

D'un point de vue pratique, la résolution de cette équation se fait avec l'algorithme de Davidson [65] qui permet d'obtenir itérativement un petit nombre de vecteurs et de valeurs propres d'états excités. En général, les énergies d'excitations CIS sont surestimées de 0.5 à 2.0 eV par rapport aux valeurs expérimentales [66–68]. Cela est dû au fait qu'on ne considère que les déterminants mono-excités dérivés de l'état fondamental HF. Le CISD, lui, permet de prendre en compte une

grande partie de la corrélation dynamique. Cependant, comme le CIS, elle n'est ni *size extensive* ni *size consistent*. Une méthode est *size-consistent* si l'énergie de deux systèmes sans interaction ($E_A + E_B$) est égale à l'énergie du complexe des deux systèmes (E_{A-B}). Le concept de *size-extensivity* illustre le fait que l'énergie d'un système doit croître linéairement avec le nombre d'atomes (électrons) qui le compose.

L'approche CIS n'obéit pas à la loi de la somme des dipôles de Thomas-Reiche-Kuhn [69–71], selon laquelle la somme des moments de transitions dipolaires doit être égale au nombre d'électrons du système. Les moments de transition dipolaire sont simplement qualitatifs.

2.1.2 Les méthodes de champ auto-cohérent multi-configurationnel

La méthode CI classique est justifiée lorsque la configuration de référence est largement prédominante (systèmes monoconfigurationnels). Mais lorsque l'énergie des OM vacantes se rapproche de celle des OM occupées, le système devient de moins en moins monoconfigurationnel. Il est donc préférable d'effectuer un calcul SCF optimisant l'énergie d'une combinaison de plusieurs déterminants, impliquant donc plusieurs configurations de référence comme c'est le cas dans l'approche du champ auto-cohérent multi-configurationnel (*Multi-Configuration Self-Consistent Field*, MCSCF). Le principe du MCSCF est basé sur le même principe que le CI [72, 73]. En effet, les coefficients des déterminants de Slater sont optimisés par le principe variationnel. Cette fois, par contre, les coefficients des OM (et donc leur forme) utilisés sont aussi construits de manière à être optimum. Cela a l'avantage de décrire plus correctement les orbitales virtuelles et donc les fonctions d'onde « excitées ». La fonction d'onde MCSCF est souvent utilisée dans des méthodes multi-références comme le MRCI (*Mutli-Reference Configuration Interaction* [74] et le CASSCF (*Complete Active Space Self-Consistent Field*) [75].

En particulier, la méthode CASSCF permet d'obtenir une excellente description des états excités. Malheureusement, cette méthode requiert un temps de calcul excessif. En effet, les n OM et les m électrons d'un système (*Complete Active Space CAS(n,m)*) décrivent l'espace actif ; toutes les combinaisons possibles seront donc considérées pour toutes les configurations. Dès que le nombre total d'électrons et toutes les OM sont considérés, le calcul devient rapidement intraitable. Des méthodes ont donc été proposées pour diminuer le nombre de configurations tels le GVB (*Generalized Valence Bond*) [76] ou le RAS (*Restricted Active Space*) [77].

Notons que la généralisation de la théorie de la perturbation MP_n à la théorie multi-configura-

tionnelle induit une fonction d'onde de type CASSCF de l'état fondamental et des premiers états excités utilisés comme déterminants de référence dans la théorie des perturbations du second ordre (CASPT2, *Complete Active Space Perturbation Theory 2*). Le CASPT2 est fortement utilisé par le groupe d'Olivucci notamment pour décrire des surfaces d'énergie potentielle et des énergies de transitions de la Rhodopsine [78–85] ou de chromophores de protéines fluorescentes [86–89]. Ils utilisent le formalisme QM/MM (Link Atom) implanté dans MOLCAS [90] par N. Ferré afin de réaliser du CASPT2//CASSCF/Amber.

2.2 Les méthodes dépendantes du temps

AFIN d'évaluer les énergies de transition verticale, l'équation dépendante du temps de Schrödinger a également été développée. Nous avons déjà vu les équations du formalisme CIS, nous allons maintenant dériver celles de l'approche TD-HF et ensuite, les équations de la TD-DFT. Cette dernière méthode est fortement employée actuellement de par sa grande efficacité tant du point de vue du temps de calcul que de sa précision. Elle est aujourd'hui devenue une référence pour le calcul de spectres d'absorption de molécules organiques telles que des colorants [45, 91–97], des interrupteurs moléculaires [98–103] que ce soit en phase gazeuse ou en phase condensée [61, 104–107]. Les méthodes QM/MM étendent son champ d'action à des systèmes protéiniques contenant des chromophores [96, 108–111]. Nous avons employé la TD-DFT pour évaluer les spectres d'absorption des chromophores du complexe squaraine-tétralactame (section 5.2) [112] et des protéines fluorescentes (section 5.3) [113, 114].

2.2.1 Hartree-Fock dépendant du temps (TD-HF)

Les équations dépendantes du temps HF (TD-HF) ont été écrites pour la première fois par Dirac dans les années 1930. Pour dériver ces équations, l'équation de Schrödinger dépendante du temps est donc le point de départ :

$$H(\mathbf{r}, t)\Psi(\mathbf{r}, t) = i\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \quad (2.7)$$

où $H(\mathbf{r}, t)$ est le Hamiltonien dépendant du temps : $T(\mathbf{r}, t) + V(\mathbf{r}, t)$. Le potentiel extérieur est une perturbation qui peut-être due à un champ électrique extérieur :

$$V(\mathbf{r}, t) = \sum_i^N v_i(\mathbf{r}, t) \quad (2.8)$$

La fonction d'onde $\Psi(\mathbf{r}, t)$ peut être écrite comme un mono-déterminant de Slater de la forme :

$$\Psi(\mathbf{r}, t) = |\psi_1(\mathbf{r}, t)\psi_2(\mathbf{r}, t) \dots \psi_N(\mathbf{r}, t)| \quad (2.9)$$

Le Fockien devient aussi dépendant du temps ($F(\mathbf{r}, t)$) et comprend le potentiel à une particule dépendant du temps ($V(\mathbf{r}, t)$) :

$$[T(\mathbf{r}) + 2J(\mathbf{r}, t) - K(\mathbf{r}, t) + V(\mathbf{r}, t)] \psi(\mathbf{r}, t) = i \frac{\partial}{\partial t} \psi(\mathbf{r}, t) \quad (2.10)$$

$$F(\mathbf{r}, t) \psi(\mathbf{r}, t) = i \frac{\partial}{\partial t} \psi(\mathbf{r}, t) \quad (2.11)$$

Les opérateurs de Coulomb et d'échange restent analogues mais avec une dépendance au temps, de même pour les fonctions atomiques $\psi(\mathbf{r}, t)$. Les équations TD-HF peuvent être exprimées dans M fonctions de base atomique $\phi_j(\mathbf{r})$:

$$\psi_p(\mathbf{r}, t) = \sum_j^M c_{pj}(t) \phi_j(\mathbf{r}) \quad (2.12)$$

L'équation de HF dépendante du temps devient matriciellement,

$$\mathbf{F} \cdot \mathbf{C} = i \frac{\partial}{\partial t} (\mathbf{S} \cdot \mathbf{C}) \quad (2.13)$$

En multipliant à droite par la matrice adjointe des coefficients \mathbf{C}^\dagger et en considérant que les fonctions de base atomique sont orthogonales entre-elles ($S_{ij} = \delta_{ij}$), on obtient :

$$\sum_q F_{pq} P_{qr} = i 2 \frac{\partial}{\partial t} P_{pr} \quad (2.14)$$

avec la matrice densité P_{pr} qui est directement liée à la densité électronique du système :

$$P_{pr} = \sum_{p,r}^M c_p(t) c_r^*(t) \quad (2.15)$$

$$\rho(\mathbf{r}, t) = \sum_{p,r}^M P_{pr} \phi_p(\mathbf{r}) \phi_r^*(\mathbf{r}) \quad (2.16)$$

En soustrayant l'adjointe de l'Eq. (2.14) de l'Eq. (2.14), l'expression de Dirac dépendante du temps est obtenue :

$$\sum_q [F_{pq} P_{qr} - P_{pq} F_{qr}] = i \frac{\partial}{\partial t} P_{pr} \quad (2.17)$$

En utilisant ce formalisme de la matrice densité, on peut montrer que les énergies d'excitation sont obtenues à partir de la réponse linéaire dépendante du temps de la fonction d'onde de

l'état fondamental indépendant du temps sur laquelle est appliquée une perturbation extérieure (dépendante du temps). Au temps $t=0$, le système est dans un état stationnaire défini par le déterminant de Slater, $\Psi_{t=0}$ qui obéit à l'équation HF indépendante du temps. Le système est supposé être dans son état fondamental, l'Eq. (2.17) devient :

$$\sum_q [F_{pq}^{(0)} P_{qr}^{(0)} - P_{pq}^{(0)} F_{qr}^{(0)}] = 0 \quad (2.18)$$

$$\sum_q P_{pq}^{(0)} P_{qr}^{(0)} = P_{pr}^{(0)} \quad \text{condition d'idempotence} \quad (2.19)$$

$F_{pq}^{(0)}$ et $P_{pq}^{(0)}$ correspondent respectivement aux éléments de la matrice de Fock et ceux de la matrice densité de l'état fondamental non-perturbé. Les éléments du Fockien indépendant du temps sont :

$$F_{pq}^{(0)} = \int \psi_p^*(\mathbf{r}) \left[-\frac{1}{2} \nabla^2 - \sum_A^{\text{noyaux}} \frac{Z_A}{\mathbf{r}_{iA}} + \sum_j^{N/2} 2J_j(\mathbf{r}_i) - K_j(\mathbf{r}_i) \right] \psi_q(\mathbf{r}) d\mathbf{r}^3 \quad (2.20)$$

Dans la base des orbitales orthonormales non-perturbées de l'état fondamental, ces matrices sont :

$$F_{pq}^{(0)} = \delta_{pq} \epsilon_p \quad (2.21)$$

avec

$$P_{ij}^{(0)} = \delta_{ij} \quad (2.22)$$

$$P_{ia}^{(0)} = P_{ai}^{(0)} = P_{ab}^{(0)} = 0 \quad (2.23)$$

Les indices $i, j, k \dots$ indiquent les orbitales occupées et les indices a, b, c, \dots les orbitales virtuelles. Si une perturbation très faible (δt) est appliquée et agit sur la fonction d'onde de l'état fondamental, les orbitales répondront à cette perturbation et les équations TD-HF sont obtenues via l'approximation de la réponse linéaire (réponse au premier ordre) des orbitales et de l'opérateur de Fock. Dans ce formalisme de la matrice densité, la fonction d'onde où, dans ce cas, la matrice densité peut être écrite comme étant la somme de l'état fondamental non-perturbé et de sa modification dépendante du temps au premier ordre,

$$P_{pq} = P_{pq}^{(0)} + P_{pq}^{(1)} \quad (2.24)$$

de même pour le Fockien :

$$F_{pq} = F_{pq}^{(0)} + F_{pq}^{(1)} \quad (2.25)$$

En substituant les Eqs (2.24) et (2.25) dans l'équation de Dirac (2.17) et en rassemblant tous les termes du premier ordre, l'équation suivante est obtenue :

$$\sum_q [F_{pq}^{(0)} P_{qr}^{(1)} - P_{pq}^{(1)} F_{qr}^{(0)} + F_{pq}^{(1)} P_{qr}^{(0)} - P_{pq}^{(0)} F_{qr}^{(1)}] = 2i \frac{\partial}{\partial t} P_{pr}^{(1)} \quad (2.26)$$

Une perturbation due à l'application d'un champ électrique dépendant du temps comporte deux termes. La première contribution correspond au champ électrique lui-même, qui peut être écrit après une transformée de Fourier comme :

$$g_{pq} = \frac{1}{2} [f_{pq} e^{-i\omega t} + f_{qp}^* e^{+i\omega t}] \quad (2.27)$$

où les éléments de la matrice f_{pq} sont les opérateurs monoélectroniques décrivant la perturbation appliquée. Les termes biélectroniques de l'équation de HF seront modifiés suite à la variation de la matrice densité tel que

$$F_{pq} = F_{pq}^{(0)} + F_{pq}^{(1)} \quad (2.28)$$

$$= F_{pq}^{(0)} + g_{pq} + \Delta F_{pq}^{(0)}. \quad (2.29)$$

$$(2.30)$$

Où, la modification du Fockien de l'état fondamental due aux variations de la matrice densité est aussi exprimée uniquement au premier ordre :

$$\Delta F_{pq}^{(0)} = \sum_{st} \frac{\partial F_{pq}^{(0)}}{\partial P_{st}} P_{st}^{(1)} \quad (2.31)$$

La seconde contribution du champ électrique intervient logiquement au niveau de la matrice densité (suite à la perturbation du Fockien), celle-ci étant exprimée au premier ordre également par :

$$P_{pq} = P_{pq}^{(0)} + P_{pq}^{(1)} \quad (2.32)$$

$$= P_{pq}^{(0)} + \frac{1}{2} [d_{pq} e^{-i\omega t} + d_{qp}^* e^{+i\omega t}] \quad (2.33)$$

où d_{pq} reprend les éléments de la perturbation des orbitales, donc de la densité. En insérant les Eqs. (2.27) - (2.32) dans l'Eq. (2.26) et en rassemblant les termes en $e^{-i\omega t}$, on obtient :

$$\sum_q \left[F_{pq}^{(0)} d_{qr} - d_{pq} F_{qr}^{(0)} + \left(f_{pq} + \sum_{st} \frac{\partial F_{pq}^{(0)}}{\partial P_{st}} d_{st} \right) P_{qr}^{(0)} - P_{pq}^{(0)} \left(f_{qr} + \sum_{st} \frac{\partial F_{qr}^{(0)}}{\partial P_{st}} d_{st} \right) \right] = \omega d_{pr} \quad (2.34)$$

Les termes en $e^{+i\omega t}$ mènent au complexe conjugué de l'équation ci-dessus. La matrice d_{pq} de l'Eq. (2.34) est scindée en quatre blocs tels que les blocs d'orbitales occupées-occupées (d_{ii}) et virtuelles-virtuelles (d_{aa}) sont nuls. Seules les contributions d'orbitales occupées-virtuelles (d_{ia}) et inversement (d_{ai}) sont non-nulles et prises en compte dans le calculs des énergies d'excitations. Le Fockien non-perturbé ainsi que les matrices densités étant diagonales, on obtient la paire d'équations suivantes avec $x_{ai} = d_{ai}$ et $y_{ai} = d_{ia}$:

$$F_{aa}^{(0)}x_{ai} - x_{ai}F_{ii}^{(0)} + \left[f_{ai} + \sum_{bj} \left(\frac{\partial F_{ai}}{\partial P_{bj}} x_{bj} + \frac{\partial F_{ai}}{\partial P_{jb}} y_{bj} \right) \right] P_{ii}^{(0)} = \omega x_{ai} \quad (2.35)$$

$$F_{ii}^{(0)}y_{ai} - y_{ai}F_{aa}^{(0)} - P_{ii}^{(0)} \left[f_{ia} + \sum_{bj} \left(\frac{\partial F_{ia}}{\partial P_{bj}} x_{bj} + \frac{\partial F_{ia}}{\partial P_{jb}} y_{bj} \right) \right] = \omega y_{ai} \quad (2.36)$$

À la limite de la fréquence-zéro ($f_{ia} = f_{ai} = 0$), en supposant que la transition électronique se passe pour une perturbation infinitésimale et utilisant le fait que dans les orbitales canoniques $F_{pp}^{(0)} = \epsilon_p$ et $P_{ii}^{(0)} = 1$, on obtient une équation aux valeurs propres non-hermitiennes écrites sous forme matricielle :

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{A}^* & \mathbf{B}^* \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \omega \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \quad (2.37)$$

avec les éléments des matrices \mathbf{A} et \mathbf{B} :

$$A_{ia,jb} = (\epsilon_a - \epsilon_i) \delta_{ij} \delta_{ab} + (ia||jb) \quad (2.38)$$

$$B_{ia,jb} = (ia||jb) \quad (2.39)$$

Une manière de réécrire l'Eq. (2.37) est la suivante :

$$\Omega \mathbf{Z} = \omega^2 \mathbf{Z} \quad (2.40)$$

avec

$$\Omega = (\mathbf{A} - \mathbf{B})^{1/2} (\mathbf{A} + \mathbf{B}) (\mathbf{A} - \mathbf{B})^{1/2} \quad (2.41)$$

$$\mathbf{Z} = (\mathbf{A} - \mathbf{B})^{1/2} (\mathbf{X} + \mathbf{Y}) \quad (2.42)$$

La différence $(\mathbf{A} - \mathbf{B})$ est un terme positif, ce qui permet de réduire les équations TD-HF non-hermitiennes en équations hermitiennes de dimension moitié moindre. D'un point de vue technique, l'Eq. (2.37) peut être résolue par la procédure de Davidson [65], utilisée en CIS. Le temps computationnel est cependant deux fois plus grand que celui en CIS.

En comparant les deux formalismes, CIS et TD-HF, on remarque clairement que les Eqs. (2.5) sont contenues dans l'Eq (2.37) et cela, lorsque la matrice $\mathbf{B} = 0$: c'est l'approximation de Tamm-Dancoff [115, 116]. La méthode TD-HF n'a cependant pas été fortement utilisée, sans doute car les énergies d'excitation obtenues sont souvent plus petites que celles obtenues en CIS mais surestiment toujours les valeurs expérimentales. Par contre, contrairement à la méthode CIS, TD-HF obéit à la loi de la somme de Thomas-Reiche-Kuhn des forces d'oscillateur [117], ce qui améliore nettement la précision des moments de transition dipolaire. La polarisabilité électrique (α_{r_i, r_j}), suite à l'application d'un champ électrique, est définie en terme de la réponse du moment dipolaire moléculaire (μ_{r_i}). Dans le cadre de calculs de transitions électroniques, seul le premier ordre de la perturbation correspond à la modification du moment dipolaire.

$$\mu_{r_i}(t) = \mu_{r_i} + \sum_{r_j} \alpha_{r_i, r_j}(\omega) \mathbf{E}(\omega) + \dots \quad (2.43)$$

Où le tenseur de la polarisabilité est souvent défini comme une moyenne :

$$\alpha(\omega) = \frac{1}{3} \sum_{r_i} \alpha_{r_i, r_i}(\omega) \quad (2.44)$$

$$= \sum_{n \neq 0} \frac{f_n}{\omega_n^2 - \omega^2} \quad (2.45)$$

où f_n est la force d'oscillateur qui est définie par l'expression suivante :

$$f_n = \frac{1}{3} (f_n^{x,x} + f_n^{y,y} + f_n^{z,z}) \quad (2.46)$$

$$= \frac{2}{3} \omega_n [|\langle \Psi_0 | x | \Psi_n \rangle|^2 + |\langle \Psi_0 | y | \Psi_n \rangle|^2 + |\langle \Psi_0 | z | \Psi_n \rangle|^2] \quad (2.47)$$

et est un nombre sans dimension lié à l'intensité du pic d'absorption (voir chapitre 5).

2.2.2 Théorie de la fonctionnelle de la densité (TD-DFT)

La théorie de la fonctionnelle de la densité dépendante du temps (TD-DFT) est une approche qui décrit les excitations électroniques. Similairement au premier théorème de Hohenberg et Kohn de la DFT, le théorème de Runge-Gross [118] établit une relation directe entre la densité électronique dépendante du temps et le potentiel dépendant du temps ($v[\rho(\mathbf{r}, t)]$); dès lors le potentiel et la fonction d'onde seront dépendantes de la densité.

$$\rho(\mathbf{r}, t) \leftrightarrow v[\rho(\mathbf{r}, t)] + C(t) \leftrightarrow \Psi[\rho(\mathbf{r}, t)] e^{-i\alpha(t)} \quad (2.48)$$

Récemment van Leeuwen a généralisé le théorème de Runge-Gross aux systèmes moléculaires [119]. De la même manière, les équations de Kohn-Sham dépendantes du temps devront être dérivées. En supposant qu'il existe un système de référence dépendant du temps sans interaction avec un potentiel extérieur $v_s(\mathbf{r}, t)$ dont la densité électronique $\rho_s(\mathbf{r}, t)$ est égale à la densité exacte du système réel avec interaction $\rho(\mathbf{r}, t)$.

$$\rho(\mathbf{r}, t) = \rho_s(\mathbf{r}, t) = \sum_i^N |\phi(\mathbf{r}, t)|^2 \quad (2.49)$$

Si le potentiel extérieur existe, on peut déduire les équations de Kohn et Sham dépendantes du temps. Les équations (1.35) et (1.36) deviennent

$$E[\rho(\mathbf{r}), t] = T_s[\rho(\mathbf{r})] + \int v_{\text{eff}}[\rho(\mathbf{r}, t)] \rho(\mathbf{r}, t) d\mathbf{r} \quad (2.50)$$

$$v_{\text{eff}}[\rho(\mathbf{r}, t)] = V_{\text{ne}}(\mathbf{r}, t) + \int \frac{\rho(\mathbf{r}', t)}{\mathbf{r} - \mathbf{r}'} d\mathbf{r}' + \frac{\partial E_{\text{xc}}[\rho]}{\partial \rho(\mathbf{r}, t)} \quad (2.51)$$

Il suffit ensuite de résoudre l'équation de Schrödinger dépendante du temps,

$$[T_s + v_{\text{eff}}] \psi_i(\mathbf{r}, t) = i \frac{\partial \psi_i(\mathbf{r}, t)}{\partial t} \quad (2.52)$$

ce qui a été explicité précédemment *via* le formalisme de la matrice densité. L'équation de la réponse linéaire TD-DFT est donc totalement similaire à celle pour la théorie TD-HF, Eq. (2.37). Cependant, les éléments des matrices \mathbf{A} et \mathbf{B} prendront en compte le terme d'échange et de corrélation (V_{xc}) :

$$A_{ia,jb} = (\epsilon_a - \epsilon_i) \delta_{ij} \delta_{ab} + (ia|jb) + (ia|V_{xc}|jb) \quad (2.53)$$

$$B_{ia,jb} = (ia|jb) + (ia|V_{xc}|jb) \quad (2.54)$$

Dans l'Eq. TD-HF (2.38), les intégrales biélectroniques correspondent uniquement à la réponse de l'échange exact non-local, ce qui en TD-DFT est remplacé par la réponse du potentiel d'échange-corrélation. L'approximation de Tamm-Dancoff (TDA) établie lors du passage du formalisme TD-HF aux équations CIS ($\mathbf{B}=0$) a aussi été étendue au formalisme TD-DFT. Les résultats obtenus en TDA/TD-DFT sont très similaires à ceux obtenus en TD-DFT, ce qui est dû à la présence de la fonctionnelle d'échange-corrélation. L'amplitude de \mathbf{Y} et les éléments de la matrice \mathbf{B} sont, en fait, une mesure de la corrélation manquante dans l'état fondamental ; celle-ci devrait donc être plus petite dans le formalisme TD-DFT qu'en TD-HF. La résolution des équations non hermitiennes TD-DFT se fait aussi par le passage à la matrice \mathbf{Z} (cf. Eq. (2.40)). D'un point

de vue du coût calculatoire, la TD-DFT est plus chère que la TD-HF. Ceci est dû à l'évaluation supplémentaire de la réponse de la fonctionnelle d'échange-corrélation.

De manière générale, nous pouvons réécrire les Eqs (2.38) et (2.53) de la manière suivante :

$$A_{ia,jb} = (\epsilon_a - \epsilon_i) \delta_{ij} \delta_{ab} + (ia|jb) - c_{\text{HF}} (ij|ab) + (1 - c_{\text{HF}}) (ia|V_{xc}|bj) \quad (2.55)$$

$$B_{ia,jb} = (ia|jb) - c_{\text{HF}} (ij|ab) + (1 - c_{\text{HF}}) (ia|V_{xc}|bj) \quad (2.56)$$

où le facteur c_{HF} est compris entre 1 et 0, avec les deux cas limites i) $c_{\text{HF}} = 1$, formalisme TD-HF et ii) $c_{\text{HF}} = 0$, celui de la TD-DFT.

L'utilisation de la TD-DFT est très populaire pour calculer les énergies de transition des états excités de valence, avec une erreur très raisonnable comprise entre 0.1 et 0.5 eV. Cette erreur est comparable aux méthodes hautement corrélées comme CASPT2 et EOM-CC (*Equation of Motion Coupled Cluster*) lorsque beaucoup d'orbitales virtuelles sont incluses dans le calcul. La TD-DFT est donc un outil puissant, alliant la précision et un temps de calcul acceptable (comparé aux méthodes fortement corrélées). La précision de la TD-DFT vient de la différence entre les énergies des orbitales (élément de la matrice \mathbf{A}). Dans le formalisme de Kohn et Sham, les énergies des orbitales virtuelles sont déterminées pour un système à N électrons et non à $(N+1)$ électrons comme c'est le cas en TD-HF. Tout comme en TD-HF, la TD-DFT respecte la loi de Thomas-Reiche-Kuhn des forces d'oscillateur.

Par contre, la TD-DFT échoue de plusieurs électron-Volt pour l'évaluation des états de Rydberg, des états de valence d'un système π étendu, des états de doubles excitations et lors de transferts de charges importants. Les deux premiers soucis sont causés par le fait qu'à longue portée, le comportement des fonctionnelles d'échange-corrélation décroît plus rapidement qu'en $1/r$, r étant la distance noyaux-électron. Concernant les transferts de charges, nous avons vu que récemment des fonctionnelles hybrides avec un comportement différent à longue et courte portée avaient été mises au point spécialement pour essayer de reproduire le comportement asymptotique de l'échange à longue distance.

Bien que l'approche TD-DFT donne des résultats très précis, le spectre d'absorption n'inclut pas encore la structure vibrationnelle du système étudié. Souvent d'ailleurs le spectre d'absorption est représenté à l'aide d'une Lorentzienne ou d'une Gaussienne construite à partir de l'énergie d'excitation de la transition et de sa force d'oscillateur [120]. Les systèmes traités étant de plus en plus grands, les contributions vibrationnelles deviennent de plus en plus importantes. Calculer

ler la structure des bandes et l'inclure dans les équations TD-DFT, introduirait évidemment un coût calculatoire énorme. Un certain nombre d'études théoriques menées par Barone [121–124] et Grimme [125] ont été développées dans ce but.

2.3 Les effets de la solvation

IL est intéressant de pouvoir comparer des propriétés moléculaires obtenues à l'aide d'outils quantiques à des données expérimentales. Très souvent ces dernières sont obtenues en phase condensée. L'effet du solvant joue un rôle important notamment avec la formation de liaisons hydrogènes, d'interactions hydrophobes ou hydrophiles, ... Il est donc nécessaire d'inclure ces effets de manière précise dans les calculs quantiques. Les méthodes de chimie théorique sont de très puissants outils quant à la compréhension des interactions soluté-solvant, la prédiction des structures, des spectres, ainsi que pour la description de la réactivité en phase liquide [126].

Il existe deux grandes manières de représenter cet environnement : explicite et implicite. La différence fondamentale entre les deux schémas est que le modèle implicite représente le solvant comme un milieu homogène tandis que le modèle explicite décrit les molécules de solvant de manière atomique. Ce dernier modèle est, physiquement, le plus réaliste parce qu'il décrit les interactions entre soluté et solvant avec plus de finesse. Cependant, il est crucial de bien définir le nombre et le positionnement des molécules de solvant sachant que le temps de calculs croît avec le nombre de molécules de solvant ajoutées. Il ne faudra prendre en compte que les molécules d'eau interagissant fortement avec le soluté [127]. Le choix de ce modèle se fera principalement par rapport au problème que l'on souhaite traiter : transfert de protons, de charges, ... Des méthodes de type QM:MM ou QM:QM'(:MM) ne doivent pas être négligées pour traiter de tels problèmes, cela allierait d'ailleurs leurs deux avantages (temps de calcul réduit par rapport à du QM complet et prise en compte de l'ensemble du système). Une autre solution est de coupler les méthodes implicites et explicites mais cette fois le nombre et le positionnement des molécules de solvant décrites explicitement sera d'autant plus important [128–130].

Représenter le solvant au moyen d'un continuum est particulièrement pertinent : comparé aux modèles qui représentent explicitement les molécules de solvant, le temps de calcul est, en effet, très significativement réduit. De plus, le modèle continuum est une approche systématique parce qu'il ne faut pas « construire » les sphères de coordination lors de chaque calcul. Nous verrons

que certains groupes vont jusqu'à utiliser un continuum isotrope pour modéliser l'environnement anisotrope d'une protéine !

Ci-dessous, nous expliquons comment le modèle de continuum de Miertus, Scrocco et Tomasi [131, 132] est construit. Nous citons néanmoins la méthode, développée au laboratoire par Jean-Louis Rivail et Daniel Rinaldi, de l'expansion multipolaire connue comme le *Nancy Continuum Model*, NCM [133–138]. Cette dernière méthode diffère du PCM essentiellement quant à la résolution analytique du terme électrostatique de l'énergie libre.

2.3.1 Principes de base des modèles de continuum

Dans un modèle de continuum, basé sur le principe du champ de réaction auto-cohérent (*Self-Consistent Reaction Field*, SCRF), on calcule l'ensemble des interactions moyennes entre le soluté et le solvant en équilibre thermique. Cette approximation n'est valable que si aucune interaction spécifique n'existe entre le soluté et le solvant (ponts hydrogène, interactions $\pi - \pi$, $\pi - cation$, ...). On parle de « *solvent reaction field* » pour un potentiel d'interaction obtenu dans un modèle où le solvant est représenté comme un continuum présentant les propriétés (constante diélectrique, coefficient d'expansion thermique, ...) du solvant réel.

Dans un premier temps, nous allons décrire les différents termes énergétiques, le bilan énergétique du modèle de continuum tenant compte de différents types d'interactions. La première étape implique la création d'une cavité dans le continuum de solvant. La formation de cette cavité coûte une certaine quantité d'énergie (positive) : ΔG_{cav} , l'énergie libre de cavitation. Cette quantité dépend de la nature du solvant ainsi que de la topologie de la cavité. En second temps, le soluté est placé dans la cavité et celui-ci interagit avec le continuum. On distingue trois types d'interactions soluté-continuum : électrostatiques, répulsives et dispersives. Les deux dernières sont calculées grâce à des relations empiriques.

Le processus d'interaction d'une molécule dans sa cavité se déroule de la façon suivante : la distribution de charge de ce soluté polarise la surface de la cavité du continuum. Des charges apparaissent donc à l'interface soluté-continuum. Ces dernières induisent un potentiel de réaction qui, à son tour, modifie la densité électronique du soluté. Ensuite, le continuum doit s'adapter et il y a une nouvelle modification de la distribution de charges à l'interface soluté-continuum ... et ainsi de suite jusqu'à l'obtention d'une convergence électrostatique entre la distribution de charges propre au soluté et celle de la surface de la cavité. Ce terme énergétique, toujours

négatif, est la contribution électrostatique (ΔG_{elec} , cf. section 2.3.1.B). Les termes de répulsion et de dispersion y sont souvent associés. Le terme de dispersion/de répulsion ($\Delta G_{\text{dis}}/\Delta G_{\text{rep}}$) implique l'interaction du soluté avec le solvant à l'interface de la cavité, et donne une contribution négative/positive à la variation d'énergie.

Au final, l'énergie totale d'interactions s'exprime en un terme électrostatique et en trois termes non-électrostatiques :

$$\Delta G = \Delta G_{\text{cav}} + \Delta G_{\text{elec}} + \Delta G_{\text{dis}} + \Delta G_{\text{rep}} \quad (2.57)$$

Ces différents termes seront respectivement explicités aux sections 2.3.1.A, 2.3.1.C. Deux approches principales se sont imposées pour calculer les différents termes du calcul SCRF : le PCM et le NCM. Dans l'approche PCM, le champ de réaction du solvant est exprimé en fonction de densités de charges situées sur la surface de la cavité. La densité de charges de surface est obtenue par une méthode numérique en résolvant l'équation de Poisson grâce, par exemple, à l'approche *Integral Equation Formalism* (IEF) [139–141]. Le calcul de ces charges permet la détermination de la contribution électrostatique (cf. section 2.3.1.B.a) à l'énergie de solvation. Le traitement de la contribution non-électrostatique, comprenant les termes de dispersion/répulsion et de cavitation, sont obtenus avec des relations empiriques (cf. section 2.3.1.C). La fonction d'onde du soluté dans le PCM est optimisée en prenant uniquement en compte les effets électrostatiques du continuum. La densité est alors déterminée et utilisée pour obtenir les termes non-électrostatiques.

Dans le formalisme du NCM, le terme électrostatique est calculé via un développement multipolaire de la distribution de charge du soluté (cf. section 2.3.1.B.b).

2.3.1.A La description de la cavité

Le choix de la cavité est une étape cruciale dans le traitement des effets de solvant à l'aide d'un modèle de continuum. La forme de la cavité doit, bien sûr, ressembler au mieux à la forme de la molécule de soluté mais les calculs seront simplifiés et plus rapides si la topologie de la cavité est simple (sphère, ellipsoïde, ...). En pratique, la cavité est construite à partir d'un ensemble de sphères centrées sur les noyaux et ayant un rayon de type van der Waals, soit la plus petite distance entre deux atomes dans une structure cristalline. Les rayons les plus couramment employés en chimie quantique sont ceux définis par Bondi [142].

On distingue principalement la surface accessible au solvant (SAS) et la surface excluant le solvant (SES). La SAS est définie par le centre d'une sphère génératrice de solvant qui « roule » sur la surface de van der Waals du soluté (voir Figure 2.1). La même sphère est utilisée pour définir la SES mais c'est la surface de contact avec le soluté qui la caractérise. Les molécules de

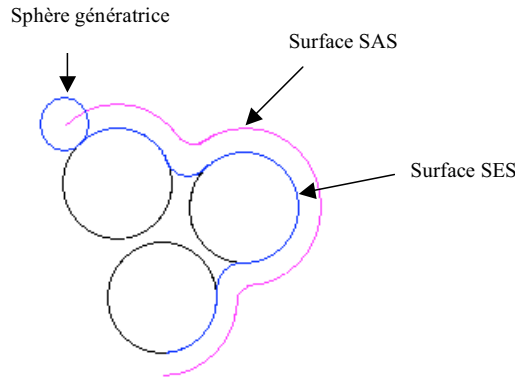


Figure 2.1: Représentation des deux différentes topologies de la cavité (SES et SAS).

soluté ont des formes irrégulières, les molécules de solvant ne savent donc pas « combler » tous les interstices du soluté. La surface de la cavité que nous utiliserons en vue de lisser la surface est la SES où de petites sphères sont introduites dans les interstices.

2.3.1.B La résolution du problème électrostatique

La distribution de charges du soluté, ρ_m , polarise le continuum diélectrique qui modifie en retour cette distribution de charges du soluté. Il s'agit donc typiquement d'un processus auto-cohérent résolu itérativement. Les modèles de continuum utilisent l'équation de Poisson de l'électrostatique classique pour solutionner le problème :

$$-\nabla \cdot [\epsilon_r(\mathbf{r}) \nabla V(\mathbf{r})] = 4\pi \rho_m(\mathbf{r}) \quad (2.58)$$

$\epsilon_r(\mathbf{r})$ est une fonction diélectrique dans le milieu et $V(\mathbf{r})$ est la somme du potentiel électrostatique (V_m) généré par la distribution de charges ρ_m et le potentiel de réaction (V_R) créé par la polarisation du milieu diélectrique :

$$V(\mathbf{r}) = V_m(\mathbf{r}) + V_R(\mathbf{r}) \quad (2.59)$$

La constante diélectrique ($\epsilon_r = \epsilon_s/\epsilon_0$) peut prendre deux valeurs :

$$\epsilon_r(\mathbf{r}) = 1, \quad \mathbf{r} \in V_{\text{int}} \quad (2.60)$$

$$\epsilon_r(\mathbf{r}) = \epsilon, \quad \mathbf{r} \in V_{\text{ext}} \quad (2.61)$$

où V_{int} (V_{ext}) est le volume à l'intérieur (à l'extérieur) de la cavité. Pour les deux régions, l'Eq. (2.58) devient :

$$-\nabla^2 V(\mathbf{r}) = 4\pi\rho_m(\mathbf{r}), \quad \mathbf{r} \in V_{\text{int}} \quad (2.62)$$

$$-\epsilon\nabla^2 V(\mathbf{r}) = 0, \quad \mathbf{r} \in V_{\text{ext}} \quad (2.63)$$

Les Eqs. (2.62) et (2.63) doivent être accompagnées de conditions frontières, à l'infini mais aussi et surtout à la surface de la cavité,

$$|V_{\text{int}} - V_{\text{ext}}| = 0 \quad (2.64)$$

afin de garantir la continuité du potentiel.

Les variantes SCRF se différencient par la manière dont sont définis la cavité et le champ de réaction. Différentes méthodes de résolution ont été proposées, nous ne citerons que les plus utilisées : l'approche de la charge de surface apparente (*Apparent Surface Charge*, ASC) et la méthode d'expansion multipolaire (*Multipole Expansion method*, MPE). La première approche étant développée par le groupe Naples (ensuite reprise par Pise). La seconde approche a été développée par D. Rinaldi et J.-L. Rivail à Nancy.

2.3.1.B.a La résolution numérique, ASC

Dans la méthode ASC, on introduit une quantité de charges de surface apparentes (σ_s) qui constitue la réponse du diélectrique à la présence de la distribution de charges à l'intérieur de la cavité. Nous utilisons le symbole \mathbf{s} pour représenter le vecteur définissant un point sur la surface de la cavité Γ . Le potentiel de réaction de l'Eq. (2.59) peut être défini dans tout l'espace :

$$V_R(\mathbf{r}) = V_\sigma(\mathbf{r}) = \int_\Gamma \frac{\sigma(\mathbf{s})}{|\mathbf{r} - \mathbf{s}|} d^2\mathbf{s} \quad (2.65)$$

Il n'y a aucune approximation dans cette dernière équation, elle est exacte lorsque $\sigma(\mathbf{s})$ est défini selon les équations de l'électrostatique. En pratique, la surface Γ est découpée en aires élémentaires (appelées *tesserae*) assez petites pour considérer constant le $\sigma(\mathbf{s})$ de chacun de ces *tesserae*.

Pour un système composé de régions à permittivité isotrope constante, la composante i du vecteur de polarisation est donnée par le gradient du potentiel $V(\mathbf{r})$:

$$P_i(\mathbf{r}) = -\frac{\epsilon_i - 1}{4\pi} \nabla V(\mathbf{r}) \quad (2.66)$$

où ϵ_i est la constante diélectrique de la région i . À la frontière des deux régions i (intérieur de la cavité) et j (extérieur de la cavité), la distribution ASC est donnée par :

$$\sigma_{ij} = -(P_j - P_i) \mathbf{n}_{ij} \quad (2.67)$$

La définition classique du PCM est dérivée de l'expression (2.66) en prenant en compte l'Eq. (2.60) et on calcule la densité de la charge à la surface comme :

$$\sigma(\mathbf{s}) = \frac{\epsilon - 1}{4\pi} \nabla V_{\text{ext}} \mathbf{n} = \frac{\epsilon - 1}{4\pi\epsilon} \frac{\partial}{\partial \mathbf{n}} (V_m + V_{\sigma_s})_{\text{int}} \quad (2.68)$$

Des charges placées dans le vide (constante diélectrique, ϵ_0), créent un champ électrique \mathbf{E}_0 au point \mathbf{r} . En présence d'un solvant, caractérisé par une constante diélectrique ϵ_s , ces charges créeront un champ électrique supplémentaire \mathbf{E} . On peut calculer la variation d'énergie libre que subit le système lorsqu'il passe du vide au solvant :

$$\Delta G_{\text{elec}} = -\frac{1}{2} \left[\iiint_{\text{espace}} \epsilon_s(\mathbf{r}) \mathbf{E}^2 d\tau - \iiint_{\text{espace}} \epsilon_0(\mathbf{r}) \mathbf{E}_0^2 d\tau \right] \quad (2.69)$$

Partant des Eqs. (2.62) et (2.63), et considérant la densité de charges nulle à l'extérieur de la cavité, on peut intégrer uniquement sur la densité de charges se trouvant à l'intérieur de la cavité. L'Eq. (2.69) devient alors :

$$\Delta G_{\text{elec}} = -\frac{1}{2} \iiint_{\text{cav}} V_R(\mathbf{r}) \rho(\mathbf{r}) \quad (2.70)$$

La variation d'énergie libre est égale à la moitié de l'énergie d'interaction des charges avec le potentiel de réaction. Il suffit donc d'introduire la valeur du potentiel de réaction dans l'Eq. (2.65). Actuellement, dans *Gaussian 03* [143], le calcul de PCM se fait par défaut par “*Integral Equation Formalism*” (IEF). Le IEF-PCM a été formulé en 1997 par Cancès et Mennucci [140, 144, 145].

2.3.1.B.b La résolution analytique, MPE

L'énergie d'interaction électrostatique entre la distribution de charge du soluté et le continuum de solvant est évaluée à l'aide d'un développement multipolaire et multicentrique de la distribution de charges. Le champ de réaction à l'équilibre (R_l^0) au point I est lié à l'expansion multipolaire d'ordre l des différents centres J :

$$R_l^0(I, J) = \sum_I f_{I'}^0(I, J) M_{I'}^0(J) \quad (2.71)$$

Dans le cas d'une cavité ellipsoïde ou sphérique, les facteurs $f_{I'}^0$ ont une expression analytique et dépendent de la constante diélectrique du solvant et de la forme de la cavité (a le rayon d'une sphère), c'est-à-dire des paramètres de l'ellipsoïde.

$$f_{I'}^0 = \frac{1}{a^{2l+1}} \frac{(l+1)(\epsilon-1)}{(l+1)\epsilon+1} \quad (2.72)$$

À partir de l'Eq. (2.71), nous obtenons alors une équation similaire à l'Eq. (2.70) :

$$\Delta G_{\text{elec}} = -\frac{1}{2} \sum_{I,J} \sum_{I'} f_{I'}^0(I, J) M_{I'}^0(J) M_{I'}^0(I) \quad (2.73)$$

La méthode des multipôles a été diversement commentée mais actuellement, elle donne des résultats similaires à ceux obtenus avec les méthodes ASC. Une étude menée par Rivail *et al.* [138] a montré que la durée des calculs impliquant le modèle de Nancy est similaire au temps de calcul de type PCM pour les optimisations de géométrie. Il en est de même pour les calculs vibrationnels. Par contre, pour le calcul de TD-DFT, le NCM se montre plus rapide que le PCM [146].

2.3.1.C Les termes non-électrostatiques

Les termes non-électrostatiques sont de trois types : cavitation, dispersion et répulsion.

2.3.1.C.a Le terme de cavitation

Conceptuellement, le transfert d'une molécule d'une phase gazeuse à une phase liquide se réalise en deux étapes. La première est la formation de la cavité dans la solution et la seconde est l'insertion de la molécule de soluté à l'intérieur de la cavité. Ces deux étapes résultent en un excès d'énergie libre. Le terme de cavitation correspond à la première étape mais n'a pas

d'équivalence expérimentale. Il est donc difficile de comparer la valeur des différents modèles qui permet de calculer ΔG_{cav} . Selon Uhlig [147], le travail pour former une cavité macroscopique est fonction de la tension de surface γ du liquide et de la surface de la cavité S_m :

$$\Delta G_{\text{cav}} = \gamma S_m \quad (2.74)$$

Avec cette relation, on peut évaluer le terme de cavitation en solution, mais elle n'est valable que lorsque les molécules de soluté et de solvant ont une taille comparable. Une autre approche a été proposée par Tuñón et *al.* pour améliorer la relation de Uhlig. [148]. Elle consiste à introduire un terme de correction dans l'Eq. (2.74) :

$$\Delta G_{\text{cav}} = \gamma S_m - k_B T \ln(1 - V_s n_s) \quad (2.75)$$

avec n_s le nombre de particules de solvant par unité de volume, V_s le volume moléculaire de solvant, k_B est la constante de Boltzmann et T la température absolue.

2.3.1.C.b Le terme de dispersion

Les forces de dispersion sont définies comme l'interaction entre des molécules présentant une polarisation instantanée. Une première approximation de ce phénomène a été réalisée par Fritz London [149]. Il propose une relation maintenant bien connue entre l'énergie de dispersion et les polarisabilités (α) ainsi que les potentiels d'ionisation (I) de deux molécules isotropiques (A et B) situées à une distance R l'une de l'autre.

$$\Delta G_{\text{disp}} = -\frac{3I_A I_B \alpha_A \alpha_B}{2R^6 (I_A + I_B)} \quad (2.76)$$

La méthode implantée dans le PCM est basée sur les formules de McWeeny [150] qui généralisent l'approche de London et utilisent la surface SAS.

2.3.1.C.c Le terme de répulsion

Dans le PCM, le terme de répulsion est évalué par l'approche d'Amovilli et Mennucci [151].

$$G_{\text{rep}} = \rho_s \int d\mathbf{r} U_{\text{ms}}^{\text{rep}}(\mathbf{r}) g_{\text{ms}}(\mathbf{r}) \quad (2.77)$$

avec m le soluté, s le solvant, \mathbf{r} est un ensemble de coordonnées qui définissent la géométrie du complexe ms , ρ_s est la densité du solvant et g_{ms} est une fonction de corrélation qui vaut 0 à

l'intérieur de la cavité et 1 à l'extérieur. En pratique, on estime le potentiel de répulsion (U^{rep}) suivant une approche de type Lennard-Jones. Comme pour le terme de dispersion, l'énergie libre de répulsion est déterminée en utilisant la surface SAS.

2.3.2 L'état de non-équilibre

Après avoir décrit les concepts fondamentaux des modèles de continuum, il est important pour la suite de pouvoir correctement décrire l'état de non équilibre présent lors des transitions électroniques. Cet état doit aussi être modélisé dans ces modèles de solvant implicites. En effet, lorsqu'un composé polaire est en présence d'un champ électromagnétique, sa polarisation est modifiée en fonction de la fréquence du champ appliqué. Quand la fréquence augmente, une situation de non-équilibre apparaît et plusieurs phénomènes interviennent du point de vue microscopique : la polarisation d'orientation (micro-ondes), de distorsion (IR) et électronique (UV). Il y a trois chutes successives de polarisation, Figure 2.2.

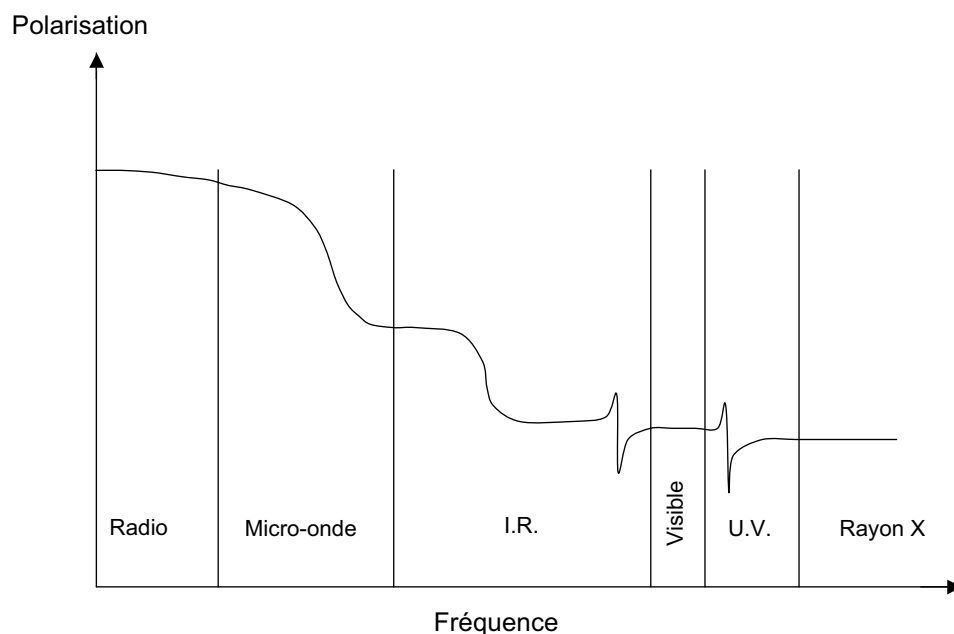


Figure 2.2: Polarisation en fonction de la fréquence

A basse fréquence, toutes les molécules ont le temps de s'aligner dans la direction du champ électrique extérieur appliqué (la polarisation est maximale) et toutes les contributions à la polarisation co-existent. Une première diminution de la polarisation se produit généralement dans les micro-ondes (10^{12} Hz). Elle est liée au fait que les molécules qui possèdent un moment dipolaire permanent ne sont plus assez rapides pour s'orienter instantanément et donc s'aligner sur les variations des oscillations du champ électrique appliqué. La contribution d'orientation ayant un temps de relaxation inférieur à 1 ps, elle ne contribue plus à la polarisation. Nous sommes dans une situation de déséquilibre où la contribution d'orientation ne participe plus à la polarisation totale, alors que les mouvements de distorsion (vibration) et électroniques parviennent toujours à s'aligner selon les oscillations du champ électrique appliqué. Au-delà d'une fréquence de 10^{14} Hz, les noyaux ne répondent plus aux oscillations du champ électrique via les vibrations parce qu'ils sont, à leur tour, trop lents ; c'est la seconde chute dans la polarisation totale (Figure 2.2). Dans le visible, la seule contribution restante est la polarisation électronique ; ce qui signifie que seul le nuage électronique est assez rapide pour continuer à suivre les variations des oscillations du champ électrique. Il est à noter qu'à très haute fréquence (10^{16} Hz) le nuage électronique s'aligne difficilement sur le champ électromagnétique.

Lorsqu'un matériau polaire est placé dans un champ électrique statique et uniforme (\mathbf{E}), sa polarisation macroscopique prend la valeur :

$$\mathbf{P} = \mathbf{P}_0 + \chi^{(1)} \mathbf{E} + \chi^{(2)} \mathbf{E}\mathbf{E} + \dots \quad (2.78)$$

Le premier terme représente la polarisation en l'absence de champ électrique, les autres termes sont caractérisés par le développement de la polarisation du matériau en série de puissances de l'amplitude du champ électrique extérieur. La susceptibilité caractérise le matériau.

Avant l'avènement des Lasers, cette équation était souvent limitée au terme en $\chi^{(2)}$. Néanmoins, cette approximation n'est valable que lorsque l'amplitude du champ électrique est faible. Macroscopiquement, la polarisation d'un matériau est donc liée à l'intensité du champ électrique extérieur :

$$\mathbf{P} = (\epsilon_s - \epsilon_0) \mathbf{E} \quad (2.79)$$

Lorsque le champ électrique est oscillant, la constante diélectrique est fonction de la fréquence, $\epsilon(\omega)$:

$$\epsilon_s(\omega) = \epsilon'_s(\omega) + i\epsilon''_s(\omega) \quad (2.80)$$

La partie réelle, $\epsilon'_s(\omega)$, est la constante diélectrique dépendante de la fréquence, décrivant la composante de la polarisation en phase avec le champ oscillant ; elle tend vers $\epsilon_s(\omega)$ lorsque $\omega \rightarrow 0$. La partie complexe ou encore facteur de perte, $\epsilon''_s(\omega)$, détermine le comportement de la polarisation ayant un déphasage de $\frac{\pi}{2}$ avec le champ et tend vers 0 quand $\omega \rightarrow 0$.

Nonobstant, les phénomènes d'orientation (dans le domaine des micro-ondes), $\epsilon'_s(\omega)$ diminuent lentement lorsque la fréquence augmente. Puis dans les régions IR, visible et UV, la diminution s'accroît. Notons aussi que $\epsilon''_s(\omega)$ augmente lorsque $\epsilon'_s(\omega)$ diminue. À haute fréquence, la structure moléculaire à l'équilibre est toujours dépendante de la constante diélectrique statique. Durant une transition électronique, les changements qui s'opèrent sont trop rapides pour que la polarisation d'orientation du solvant ait lieu. Le champ de réaction est alors divisé en deux parties : une partie électronique, qui est obtenue en remplaçant la constante diélectrique statique $\epsilon'_s(\omega)$ par la valeur à haute fréquence ϵ_∞ et une partie d'orientation, qui est la différence entre les valeurs $\epsilon'_s(\omega)$ et ϵ_∞ .

2.3.2.A Les transitions verticales

En considérant un soluté en phase condensée, dans son état fondamental, il y a un équilibre nucléaire et électronique entre soluté et solvant. Durant une transition électronique, les positions des noyaux sont les mêmes qu'à l'état initial ; on appelle cela une transition verticale (principe de Franck-Condon [152–154]). En effet, la configuration nucléaire du soluté et du solvant (polarisations d'orientation/vibration nulles) est inchangée, seuls les électrons sont assez rapides pour se relaxer dans la direction de champ appliqué. Les électrons du solvant doivent donc se réorganiser suite à la modification du nuage électronique du soluté. Le système se trouve donc dans un état de non-équilibre. Pour considérer cet état, il est nécessaire de scinder la polarisation du solvant en une contribution dynamique associée aux mouvements électroniques et en une contribution statique, liée aux mouvements nucléaires et moléculaires. Pratiquement pour considérer correctement cette réponse électronique du solvant, seule la composante dynamique de la polarisation est prise en compte et dès lors, la constante diélectrique du solvant est ϵ_∞ .

Ici, nous ne considérons que des transitions électroniques verticales, la réponse du solvant est donc divisée en deux termes :

$$\mathbf{P} = \mathbf{P}^{\text{rapide}} + \mathbf{P}^{\text{lent}} \quad (2.81)$$

Dans des processus très rapides, comme les transitions électroniques qui entraînent le changement

de l'état électronique du soluté via des absorptions de photons, la polarisation \mathbf{P} sera réduite à la réponse liée aux électrons du solvant $\mathbf{P}^{\text{rapide}}$ et \mathbf{P}^{lent} comprendra tous les autres termes liés aux différents degrés de liberté du continuum (orientation/vibration). Lors du calcul TD-DFT de non-équilibre, \mathbf{P}^{lent} est nul. Les équations de la TD-DFT en phase condensée se voient modifiées au niveau des matrices \mathbf{A} et \mathbf{B} par un terme additionnel relatif au solvant ($V_{ai,bj}$) [155] :

$$A_{ia,jb} = (\epsilon_a - \epsilon_i) \delta_{ij} \delta_{ab} + (ia|jb) + (ia|V_{xc}|jb) + V_{ai,bj}^{\text{solvant}} \quad (2.82)$$

$$B_{ia,jb} = (ia|bj) + (ia|V_{xc}|jb) + V_{ai,bj}^{\text{solvant}} \quad (2.83)$$

Les effets du solvant agissent de deux manières : (i) indirectement via la modification des orbitales et donc de leurs énergies et (ii) explicitement *via* le terme de perturbation $V_{ai,bj}^{\text{solvant}}$. Ce dernier terme peut être décrit comme l'interaction électrostatique entre la distribution de charge $\psi_a^* \psi_i$ et la contribution dynamique $\psi_b^* \psi_j$ due au potentiel de réaction du solvant.

2.3.2.A.a Le modèle PCM

Dans le modèle du PCM, $V_{ai,bj}^{\text{solvant}}$ est le potentiel du champ de réaction V_R des Eqs. (2.65) et (2.70). Nous avons vu que ce potentiel de réaction était dépendant des charges ASC ($\sigma(\mathbf{s})$) qui sont elles-même fonction de la polarisation (Eq. 2.67). Lors de la transition électronique (à autre fréquence), $\epsilon_s(\omega)$ tend vers ϵ_∞ qui est la constante diélectrique à la fréquence à laquelle la polarisation d'orientation/vibration ne peut plus suivre les mouvements des électrons ; \mathbf{P}^{lent} est nul. La polarisation devient donc :

$$\mathbf{P}(\omega) = \mathbf{P}^{\text{rapide}} = (\epsilon_s(\infty) - 1)\mathbf{E} \quad (2.84)$$

et les charges ASC sont modifiées. La contribution électrostatique ΔG_{elec} de la solvation s'en voit modifiée (cf. Eq. 2.70).

2.3.2.A.b Le modèle NCM

Dans le formalisme du NCM, la réponse électronique du solvant est ajoutée au terme électrostatique calculé à l'équilibre [156, 157]. Le modèle doit donc considérer un changement rapide des moments multipolaires de l'équilibre ($M_l^0(J)$) vers ($M_l(J)$). La forme de la cavité reste inchangée lors de cette modification de la distribution de charges. Il suffit donc de résoudre les équations de Laplace en supposant la composante statique de la constante diélectrique du solvant

comme gelée ; sa contribution dynamique constituera la réponse électronique instantannée. Le champ de réaction reste similaire à l'Eq. (2.71) en incluant les contributions statiques (polarisation lentes) et dynamiques (polarisation rapides). Ainsi, le facteur du champ de réaction $f_{ll'}^{mm'}$ est divisé en deux termes : statique ($f_{ll'}^0$) et dynamique ($f_{ll'}^\infty$). Rappelons que ce facteur dépend de la constante diélectrique du milieu ainsi que de la forme de la cavité (Eq. (2.72)).

$$R_l(I, J) = \sum_I \sum_{l'} (f_{ll'}^0(I, J) - f_{ll'}^\infty(I, J)) M_{l'}^0(J) + f_{ll'}^\infty(I, J) M_{l'}(J) \quad (2.85)$$

L'Eq. (2.73) correspondait à l'équilibre ($\Delta G_{\text{élec}}^{\text{eq}}$) et elle devient dans un état de non-équilibre :

$$\Delta G_{\text{élec}}^{\text{non-eq}} = \Delta G_{\text{élec}}^{\text{eq}} + \frac{1}{2} \sum_{ll'} [f_{ll'}^0(I, J)] M_{l'} M_{l'} - \sum_{ll'} [f_{ll'}^0(I, J) - f_{ll'}^\infty(I, J)] M_{l'}^0 M_{l'} \quad (2.86)$$

Deuxième partie

Le traitement de systèmes biologiques

Chapitre 3

La mécanique moléculaire

UNE étape importante vers le traitement complet de systèmes biologiques de quelques milliers voire de millions d'atomes est la méthode des champs de forces (CF) ou mécanique moléculaire (MM). Ces champs de forces ignorent totalement les électrons de la structure et calculent l'énergie du système comme une fonction de la position nucléaire. Ne traitant pas la réponse électronique de la structure, des propriétés telles que les excitations électroniques, ne peuvent être traitées. Par contre, ces méthodes sont très employées pour observer les changements de conformation de protéine/membrane (ex. canaux ioniques), l'importance de l'environnement sur des sites actifs d'enzymes, etc.

La MM repose sur des hypothèses de base : la première étant l'approximation de Born-Oppenheimer, sans laquelle il serait impossible de décrire l'énergie en fonction de la position des noyaux atomiques. La MM est basée sur un modèle relativement simple d'interactions au sein du système traité : l'allongement/raccourcissement des liaisons, la courbure des angles, les rotations autour d'une liaison simple, le van der Waals (vdW) et l'électrostatique. Dès lors, le principe de transférabilité est un attribut clé des CF, puisque les paramètres liés et non-liés de ces interactions sont développés et testés sur un nombre relativement faible de structures et ne prennent pas en compte toutes leurs conformations. Ces structures sont plutôt petites et très souvent isolées. Les paramètres liés sont ensuite appliqués à des systèmes nettement plus grands, qui mettent en jeu beaucoup plus d'interactions de type non-liées. C'est à ce niveau qu'il faut parfois avoir confiance dans les modèles et leur principe de transférabilité.

Il existe deux grandes catégories de CF, classiques (CFC) et polarisables (CFP). La première catégorie est actuellement la plus employée et la plus courante de par leur nombre mais aussi

par leur facilité à être utilisée. Les CFP essaient de prendre en compte la réponse électronique de l'environnement à l'aide de modèles complexes. Tant que la polarisation du système n'a pas de contribution importante, les CFC suffisent. Par contre, lorsque les effets d'inductions ne peuvent plus être négligés au long de la simulation, il est nécessaire de les prendre en compte de manière correcte. Plusieurs groupes ont montré leur importance au travers de divers exemples : les métalloprotéines [158, 159], la formation de crystal, le repliement de protéines [160, 161], ...

Nous décrivons brièvement les CF usuels de ces deux catégories après avoir discuté la forme générale de l'expression de la fonction d'énergie standard (non-polarisable).

3.1 La fonction d'énergie potentielle

LA forme de la fonction d'énergie potentielle d'un champ de forces non-polarisables comprend des termes de liaisons associés aux atomes liés de manière covalente, et des termes non-liés décrivant les interactions à longue portée (électrostatique et de van der Waals). La décomposition spécifique en chacun de ces termes dépend du champ de forces, mais la forme générale de cette fonction peut être écrite :

$$V_{\text{MM}}(\mathbf{r}_N) = \underbrace{V_{\text{liaisons}} + V_{\text{angles}} + V_{\text{dièdres}} + V_{\text{impropres}}}_{V_{\text{liés}}} + \underbrace{V_{\text{van der Waals}} + V_{\text{électrostatique}}}_{V_{\text{non-liés}}} \quad (3.1)$$

$V_{\text{MM}}(\mathbf{r}_N)$ désigne l'énergie potentielle totale dépendant de la position \mathbf{r} des N atomes présents dans le système. Les termes de liaison et d'angle (V_{liaisons} et V_{angles}) sont habituellement modélisés comme des oscillateurs harmoniques ne permettant pas la formation ou la destruction de liaisons¹.

$$V_{\text{liaisons}} = \sum_{\text{liaisons}} \frac{1}{2} k_r (\mathbf{r} - \mathbf{r}_0)^2 \quad (3.2)$$

$$V_{\text{angles}} = \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (3.3)$$

La déviation par rapport à la position d'équilibre de liaison et d'angle est exprimée respectivement par $(\mathbf{r} - \mathbf{r}_0)$ et $(\theta - \theta_0)$. La constante de force du potentiel harmonique est caractérisée par k_r et k_θ . Une description plus réaliste d'une liaison covalente soumise à un étirement important

¹Cependant, il existe quelques CF spécifiques permettant la formation et la rupture de liaisons, par exemple UNRESFF (*United-Residue Force Field*) [162] ou ReaxFF (*Reactive Force Field*) [163]

peut être reproduite par un potentiel de Morse ou anharmonique (plus coûteux), il en est de même pour le terme de courbure des angles. La forme des autres termes liés est très variable et fortement dépendante du CF utilisé mais s'écrit fréquemment :

$$V_{\text{dièdres}}(\omega) = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \quad (3.4)$$

$$V_{\text{impropres}}(\theta \text{ ou } d) = \frac{1}{2}k(\theta \text{ ou } d)^2 \quad (3.5)$$

Dans l'Eq. (3.4), les valeurs de V_n permettent de reproduire le profil de torsion, n est la multiplicité c'est-à-dire le nombre de minima présents dans le potentiel de torsion, ω est l'angle de torsion et γ est le facteur de phase qui détermine où le minimum du potentiel de torsion se situe. Les potentiels d'angles et de distances impropres (Eq. 3.5) permettent de contraindre la géométrie de molécules. Ils servent notamment à contraindre la planarité d'un groupement d'atomes (ex., groupement carbonyle, cycles aromatiques, ...).

Le terme de van der Waals (V_{vdW}) est habituellement calculé à l'aide d'un potentiel 12 – 6 de Lennard-Jones (LJ) :

$$V_{\text{vdW}} = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (3.6)$$

Le potentiel 12 – 6 de LJ contient deux paramètres : σ , la distance pour laquelle l'énergie du potentiel est nulle et ϵ , la profondeur du puit. Les paramètres σ et ϵ caractérisant les interactions entre des atomes A et B sont calculés par les relations suivantes (loi de Lorentz-Berthelot) :

$$\epsilon = (\epsilon_A \epsilon_B)^{1/2} \quad (3.7)$$

$$\sigma = \frac{1}{2}(\sigma_A + \sigma_B) \quad (3.8)$$

Le potentiel de LJ est caractérisé par une partie attractive variant en r^{-6} et une partie répulsive en r^{-12} . Le terme en r^{-6} est élevé à la même puissance que l'expression de l'énergie de dispersion. Ce potentiel 12–6 est largement utilisé pour les gros systèmes parce que le terme répulsif peut être rapidement calculé à partir du terme attractif, ce dernier peut lui-même être défini à partir du carré des distances.

Enfin, le terme électrostatique est déterminé comme une somme entre les deux charges ponctuelles associées à deux atomes (A et B) par la loi de Coulomb :

$$V_{\text{élec}} = \sum_{A>B} \frac{q_A q_B}{4\pi\epsilon_0 r_{AB}} \quad (3.9)$$

Une fois ces termes généraux définissant la fonction d'énergie potentielle décrit, la section suivante présentera les principaux champs de forces utilisés lors de ce travail.

3.2 Les champs de forces classiques

DANS cette section, nous reprenons les caractéristiques principales des CF classiques usuels. Nous avons utilisé principalement au cours de la thèse le CF Amber99 et Charmm27. Le premier a été utilisé pour traiter les protéines fluorescentes (cf. section 5.3) et la protéine Sérum Albumin Bovine (cf. section 7). Le second a été employé dans le chapitre dédié à la capture électronique sur une macromolécule (cf. chapitre 6). La grande différence entre ces deux champs de forces se situe, entre autre, au niveau de la répartition des charges des acides aminés ainsi que la manière de les calculer. Le CF Amber considère une charge globale (0, +1 ou -1) associée à chaque acide aminé (selon sa nature). Dans le CF Charmm, cette charge globale est toujours 0, +1 ou -1 mais deux contraintes sont ajoutées : la charge globale du squelette de l'acide aminé (-C(=O)-C(H)-N(H)-) est toujours nulle et la charge de la chaîne latérale de l'acide aminé est donc 0, -1 ou +1 selon l'acide aminé. La décomposition de la charge globale des acides aminés dans Charmm nous permettra de faire une analyse systématique en posant les charges des atomes de la chaîne latérale égale à 0 afin d'observer les effets électrostatiques de la chaîne latérale. Dans ce cas, le fait que la charge globale du squelette des acides aminés soit nulle a le grand avantage de ne pas créer un moment dipolaire non souhaité, ce qui serait le cas avec le CF Amber.

AMBER

a été développé par Wiener et Cornell [164]. Il contient des paramètres pour tous les acides aminés, les acides nucléiques et certaines molécules organiques (solvant). Une particularité de ce champ de forces est qu'il décrit les acides aminés comme des entités ($\text{-NH-CH}_\alpha\text{-C(O)-}$) possédant une charge globale. Nous verrons que cela se révèle très utile lorsque nous utiliserons des macromolécules où une liaison covalente lie le sous-système QM au sous-système MM. Il existe plusieurs versions de Amber : 1994, 1996, 1999, 1999sb (ADN), 2002 et 2003. Les deux dernières versions sont classées dans les CFP. GAFF (*Generalized Amber Force Field*) est la version généralisée du champ de forces Amber [165]. Il a été dérivé pour être compatible avec ce dernier. Ce CF contient un grand nombre de paramètres pour décrire des molécules organiques. GAFF peut donc être utilisé pour définir correctement

des types d'interactions liées ou non qui ne seraient pas décrites par Amber.

CHARMM

a été développé par Karplus [166–168]. Il s'agit aussi d'un des champs de forces les plus fréquemment utilisés pour traiter des systèmes biologiques tels que des protéines, des lipides et des acides nucléiques. Les paramètres de Charmm27 [169] ont été utilisés pour traiter les macromolécules contenant des ponts disulfures.

3.3 Les champs de forces polarisables

DANS ces CF particuliers, la polarisabilité est prise en compte implicitement au travers des paramètres empiriques présents dans les termes d'interactions électrostatiques et de van der Waals. Diverses approches ont été considérées ces dix dernières années :

Les moments dipolaires induits

Cette approche consiste à inclure des charges partielles et des dipôles induits sur les atomes. Le moment dipolaire (μ) induit sur une particule est proportionnel au champ électrique sur cette même particule. La constante de proportionnalité est le tenseur de la polarisabilité. Le moment dipolaire ressent le champ électrique créé par les charges ponctuelles du système et par les autres moments induits. Une implantation de cette approche a été introduite dans Charmm, nommée Charmm-PIPF (*Polarizable Intermolecular Potentials Functions*) [170]. Un autre CF connu est celui développé par Ponder et collaborateurs AMOEBA (*Atomic Multipole Optimized Energetics for Biomolecular Applications*) [171–175], où le moment dipolaire cette fois ressent le champ électrique non plus créé par des charges mais bien par un multipôle permanent ainsi que par les autres moments induits.

Les oscillateurs de Drude

Cette méthode (souvent appelée « charge sur un ressort ») modélise la polarisation en ajoutant une particule chargée (particule de Drude) sans masse, liée par un ressort au centre atomique de chaque atome polarisable . Un avantage de cette méthode est que l'interaction Coulombienne reste simple, donc son implantation peut se réaliser plus facilement. Actuellement, quelques CF incluant les oscillateurs de Drude existent TIP4P [176] et SWM4-(N)DP [177,178] pour décrire l'eau, SWM4-(N)DP a été implanté dans Charmm et ensuite les paramètres ont été étendus aux molécules organiques [109,179–181]. En 2009,

le moment dipolaire d'une membrane a d'ailleurs été reproduit par Hader et collaborateurs [182].

Les charges fluctuantes

Cette approche comprend la polarisabilité dans la fonction d'énergie en ajoutant des valeurs de charges partielles qui correspondent au champ électrique de leur environnement, c'est-à-dire en altérant la polarisabilité. Cela peut être réalisé en couplant les charges à leur environnement et en utilisant des approches de type égalisation de l'électronégativité ou égalisation de potentiel chimique. Cette approche a été implantée dans UFF par Rappé [183] et par Patel dans Charmm [184, 185].

Pour plus d'informations, nous renvoyons le lecteur à plusieurs très bonnes revues sur ce sujet [186–188].

Chapitre 4

Les méthodes hybrides QM-MM

LES méthodes hybrides alliant la mécanique quantique (QM) et la mécanique moléculaire (MM) se révèlent un intermédiaire entre les méthodes quantiques traditionnelles traitant des systèmes macromoléculaires et les méthodes de la mécanique moléculaire employées pour traiter des systèmes biologiques (protéines, membranes, ...). En effet, de nombreux systèmes sont trop grands pour être traités entièrement par la QM vu que c'est une approche très coûteuse. Une solution pourrait être de faire du QM-QM', tout en restant dans un système raisonnable (QM' dénotant à un niveau de théorie souvent nettement plus faible qu'employé en QM). Pour déterminer des propriétés moléculaires de gros systèmes biologiques, les méthodes MM échouent, ne traitant pas les électrons explicitement. N'oublions pas non plus qu'en MM, les CF usuels ne permettent pas la rupture et la formation de liaisons, ce qui empêche de décrire correctement des réactions chimiques. Le principe même des approches QM-MM est de scinder un système macromoléculaire (\mathcal{S}) en deux parties (Figure 4.1) afin d'allier les avantages de ces deux méthodes de base. Dans la suite, nous utiliserons le tiret entre les deux sous-systèmes (QM-MM) pour présenter une situation générale. La partie d'intérêt (\mathcal{I}) est traitée à l'aide de la QM et une seconde partie décrivant son environnement (\mathcal{E}) par la MM.

Par convention, si une liaison covalente lie les deux sous-systèmes, on notera la méthode QM/MM, sinon ce sera QM:MM. Deux exemples concrets d'application sont montrés ci-dessous. En QM:MM, une réaction chimique dans un solvant décrite par un CF peut-être étudiée ou l'analyse des différentes conformations d'un analogue de la proline dans un solvant. L'analogue traité possède une substitution en position β ou γ (collaboration avec A. Flores et C. Aleman) [189]. Un exemple de système où les deux parties QM et MM seront liées par une liaison covalente

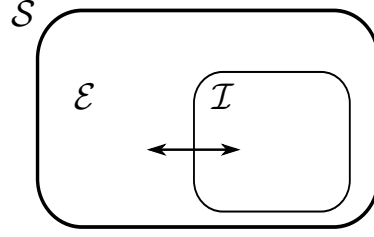


Figure 4.1: Représentation schématique des sous-systèmes en QM-MM. \mathcal{S} est le système global, \mathcal{E} décrit en MM l'environnement de la partie d'intérêt \mathcal{I} qui sera traitée au niveau QM.

sont des systèmes tels que les protéines fluorescentes, les protéines photoactives jaunes (PYP), les rhodopsines... Ci-dessous une représentation de monomère A de la rhodopsine (pdb code : 1HZX), où le rétinol (chromophore) est lié de manière covalente au reste de la protéine. La rhodopsine a aussi été maintes fois étudiée, nous renvoyons le lecteur aux publications [81–84, 89, 190–192].

Chaque atome appartiendra à un des sous-systèmes. Les deux sous-systèmes interagissant fortement entre eux, l'énergie totale du système ne peut pas être simplement la somme de l'énergie QM et de l'énergie MM. Un terme complémentaire décrivant les interactions de types électrostatiques et van der Waals entre les deux sous-systèmes doit intervenir lors du calcul de l'opérateur Hamiltonien total sous un terme de couplage ($H_{\text{QM-MM}}$) : ce dernier utilisé en QM-MM est un hamiltonien effectif H_{eff} qui agit sur la fonction d'onde (Ψ) du système. Il est donc décomposé en trois termes (Eq. (4.1))

$$H_{\text{eff}} = H_{\text{QM}} + H_{\text{MM}} + H_{\text{QM-MM}} \quad (4.1)$$

Les deux premiers termes sont respectivement propres à la partie QM et à la partie MM du système. Ils ont déjà été largement explicités dans la partie (I) et dans le chapitre (3). Pour rappel,

$$H_{\text{QM}} = H_{\text{QM}}^{\text{élec}} + H_{\text{QM}}^{\text{noyaux}} \quad (4.2)$$

$$= -\frac{\nabla_i^2}{2} - \sum_A^{\text{noyaux}} \frac{Z_A}{r_{iA}} + \sum_j^{N/2} (2J_j(r_i) - K_j(r_i)) + \sum_A^N \sum_{B>A}^N \frac{Z_A Z_B}{|R_{AB}|} \quad (4.3)$$

Quant au terme frontière de l'Eq. (4.1), il dépend fortement de la méthode QM-MM utilisée. En effet, la région frontière peut contenir des atomes additionnels (méthode *Link Atoms*) pour

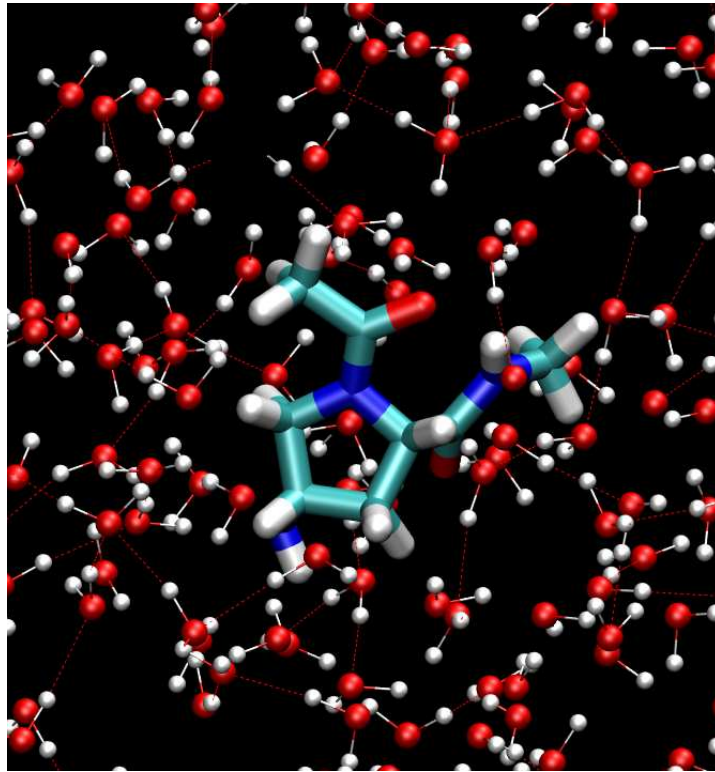


Figure 4.2: Exemple de système QM:MM. Analogue de la proline substituée en position γ par un NH_2 (Ac- γ cAmp-NHMe) dans une boîte d'eau [189].

satuer le sous-système QM où les atomes sont décrits au niveau quantique (QM) et aussi au niveau classique (MM), ces atomes sont dits hybrides ou quanto-classiques. De manière générale, on peut écrire

$$H_{\text{QM-MM}} = V_{\text{QM-MM}}^{\text{noyaux}} + V_{\text{QM-MM}}^{\text{elec}} + V_{\text{QM-MM}}^{\text{vdW}} \quad (4.4)$$

où ces termes sont respectivement les interactions noyaux QM-charges MM, électrons QM-charges MM et enfin un terme de vdW entre les deux sous-systèmes. Afin de définir les interactions de vdW, il sera nécessaire de caractériser chaque atome QM par un type d'atome du CF auquel il correspond.



Figure 4.3: Exemple de partition QM/MM

4.1 Les expressions de l'énergie QM-MM

L'ÉNERGIE totale d'un système en QM-MM sera aussi décomposée en trois termes

$$E_{\text{tot}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}} \quad (4.5)$$

avec le dernier terme défini comme

$$E_{\text{QM-MM}} = E_{\text{QM-MM}}^{\text{nuc}} + E_{\text{QM-MM}}^{\text{elec}} + E_{\text{QM-MM}}^{\text{vdW}} \quad (4.6)$$

$$E_{\text{QM-MM}}^{\text{elec}} = \langle \Psi | - \sum_{i \in \mathcal{I}} \sum_{M \in \mathcal{E}} \frac{q_M}{|\mathbf{R}_M - \mathbf{r}_i|} | \Psi \rangle \quad (4.7)$$

Nous verrons ci-après que le terme de l'Eq. 4.7 n'est calculé que dans le cas où l'*electrostatic embedding* est pris en compte. En effet, il existe plusieurs modèles pour considérer la contribution électrostatique de l'environnement sur la partie QM.

4.1.1 La contribution électrostatique

La contribution électrostatique de l'environnement peut être considérée de trois manières différentes ; elle influencera donc différemment la fonction d'onde de la partie d'intérêt QM. Ces trois approches sont explicitées dans cette partie.

4.1.1.A Le *mechanical embedding* (ME)

Dans ce premier modèle, les interactions électrostatiques QM-MM sont traitées comme des interactions électrostatiques MM-MM. Il est donc crucial que les charges MM et surtout QM soient bien définies. Le modèle de charge employé dans le champ de forces (des charges ponctuelles) sera utilisé pour décrire la partie MM et la partie QM¹. La contribution électrostatique de l'environnement ne modifie absolument pas le Fockien de la partie QM. Dès lors, la partie QM n'est pas polarisée par l'environnement électrostatique, c'est ce qu'on appelle le *mechanical embedding*. Lorsqu'on utilise cette approche, il faut donc être vigilant que la propriété étudiée ne soit pas influencée par son environnement ce qui est le cas lors de calculs d'affinité électronique ou de spectre d'absorption.

Typiquement, cette approche est celle mise au point par Morokuma dans la première version de IMOMM (*Integrated Molecular Orbital/Molecular mechanics*) [198] basée sur ONIOM (*Our N-layered Integrated Molecular Orbital and Molecular Mechanics*) [199, 200].

4.1.1.B L'*electrostatic embedding* (EE)

C'est une manière de polariser la fonction d'onde QM et donc sa densité de charges. En effet, lors du calcul de la partie QM, un terme additionnel est ajouté (Eq. (4.7)) : le terme électrostatique de la partie MM. La partie QM est donc polarisée directement par son environnement. Le Hamiltonien de cœur se voit modifié par l'introduction d'un terme :

$$V_{\text{QM-QM}}^{\text{elec}} = - \sum_{i \in \mathcal{I}}^N \sum_{M \in \mathcal{E}}^L \frac{q_M}{|\mathbf{R}_M - \mathbf{r}_i|}$$

¹Lorsque la partie d'intérêt QM n'est pas un résidu classique (présent dans le CF), ses charges devront être définies. Ces charges ne représenteront pas la véritable densité de charge de cette partie en présence de son environnement. Il est possible de définir un jeu de charges (souvent en phase gaz) en les dérivant du potentiel électrostatique de la partie QM. Ces charges peuvent être de type Merz-Kollman [193, 194], Chelp [195], ChelpG [196], ... Le logiciel R.E.D. RESP (*Restraint Electrostatic Potential*) permet de définir des charges, d'imposer des contraintes sur les charges et de prendre en compte plusieurs conformations [197].

Dans l'Eq. ci-dessus, les indices j et M sont respectivement pour les N électrons de la partie QM, les L charges ponctuelles de la partie MM. Le grand avantage de cette approche est qu'aucun jeu de charges pour la partie QM n'est à définir et que la distribution de charges de la partie QM est directement polarisée par son environnement MM. Cependant, ce modèle surpolarise parfois la fonction d'onde lorsque les charges ponctuelles de la partie MM sont très proches de la partie QM. C'est le cas lorsqu'une liaison covalente lie les deux sous-systèmes. Nous expliciterons ce cas dans les sections 4.2 et 4.3.2 .

L'EE est très souvent implanté dans les méthodes QM-MM, cependant son coût calculatoire est évidemment plus important que l'approche ME. Dans les récents développements de IMOMM [201, 202], l'EE est présent; cette approche a été implantée dans Gaussian03 dès la version C.02 [143].

4.1.1.C Le *polarizable embedding* (PE)

L'EE considère explicitement les interactions électrostatiques de la densité de charges QM, polarisée par les charges ponctuelles du CF. L'étape suivante vers une description plus exacte est d'introduire un jeu de charges MM flexibles pour décrire l'environnement. La partie QM polariserait, ainsi, à son tour la partie MM. Il existe actuellement deux grandes catégories : a) la densité de charges est polarisée par des sites polarisables via un CFP mais pas inversement et b) il y a une polarisation complète entre la partie QM et la partie MM. Grâce à ces CFP, la polarisation électronique de l'environnement serait considérée lors du traitement de spectre d'absorption/fluorescence.

Cependant, les applications de ces CFP en QM-MM sont relativement rares, dûs à leur temps de calculs excessifs et à leur faible généralisation. Nous verrons que certains groupes ont utilisé cette formule sur des systèmes photobiologiques dans la section 5.3.

4.1.2 Les contributions de van der Waals dans la région frontière

En plus des interactions électrostatiques situées dans la région frontière, les interactions entre atomes non-liés de type de van der Waals (vdW) dans cette région ne doivent pas être négligées, même si leur contribution est faible. Leur traitement est plus simple que le terme électrostatique, ce terme étant simplement ajouté à l'énergie et n'intervenant pas dans le Hamiltonien de la fonction d'onde.

Elles sont typiquement décrites par le potentiel de Lennard Jones employé dans le champ de forces choisi. Il est donc nécessaire d'assigner aux atomes QM un type d'atome.

4.2 Les différentes coupures en QM/MM

UNE attention particulière doit être portée lorsqu'une liaison covalente assure la jonction entre la partie QM et la partie MM (QM/MM). Cette coupure est, en effet, très délicate et peut induire un certain nombre d'artefacts. C'est évidemment la manière de la traiter qui les différenciera. La Figure (4.4) explicite la terminologie que nous allons utiliser tout au long de ce travail. Lorsque le sous-système \mathcal{I} est lié de manière covalente au sous-système \mathcal{E} , le premier

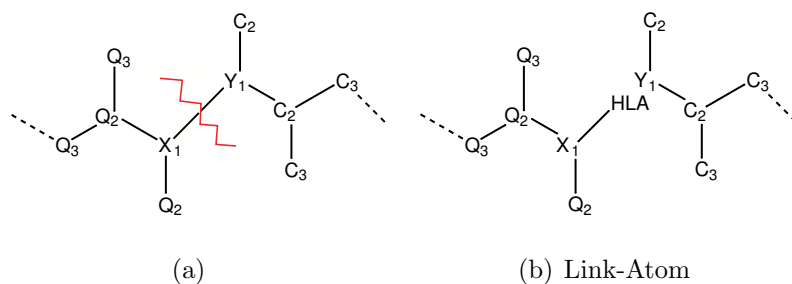


Figure 4.4: Notation pour les systèmes QM/MM. Les atomes de la partie QM sont notés Q. Le reste de l'environnement est noté C. Le premier atome de la partie QM (MM) est noté X(Y)

atome QM sera noté X_1 et le premier atome MM sera noté Y_1 . La coupure d'une liaison covalente introduit trois problèmes majeurs :

1. La liaison pendante de l'atome X_1 doit être saturée ce qui entraîne une troncature de la région QM, la liaison étant coupée de manière homolytique ou hétérolytique.
2. Comme dit précédemment dans la section 4.1.1 (page 62), dans le cas de l'EE et du PE, une surpolarisation de la densité de charges QM par les charges ponctuelles MM trop proches peut apparaître [203, 204]. Ainsi, les coupures en LA sont préférentiellement entre atomes de même nature.
3. Au niveau de la région frontière, les contributions liantes MM impliquant les atomes des deux sous-systèmes doivent être correctement définies et certaines devront être supprimées afin d'éviter un double comptage.

Dans la littérature, deux grandes approches pour traiter cette frontière ont été proposées :

i) le *Link Atom* (LA) où un atome supplémentaire (noté LA) est introduit au niveau de la liaison pendante, c'est-à-dire lié à l'atome X_1 ² et ii) l'approche des Orbitales Localisées (OL) qui place des orbitales hybrides et/ou gelées au niveau de la liaison frontière.

4.2.1 Les approches de type Link Atom

Dans cette dernière approche, la valence libre de X_1 , créée suite à la coupure QM/MM, est saturée par un atome additionnel (LA). L'atome lien est souvent un hydrogène (approche HLA) ou un atome reparamétrisé possédant une valence de libre comme dans la méthode *Connection Atom* [206] où les paramètres sont ajustés pour les méthodes semi-empiriques. Aussi, les approches *Pseudobond* [207–209] et *quantum-capping potential* [210] possèdent un atome reparamétrisé où un potentiel de cœur effectif est employé pour reproduire les propriétés de la liaison originelle.

Revenons à l'approche HLA qui est la plus employée et implantée dans de nombreux codes QM. L'introduction d'atomes lien (hydrogène) qui ne sont pas présents dans les systèmes moléculaires entraîne plusieurs complications :

- i) Chaque LA crée un artefact par l'introduction de trois degrés de liberté supplémentaires et artificiels. En effet, ce LA n'appartient normalement ni au sous-système QM, ni au sous-système MM. La définition et donc la résolution du Jacobien est donc ambiguë.
- ii) Le LA est chimiquement et électroniquement différent de la fonction qu'il remplace.
- iii) Le LA est très proche de la région MM ce qui peut amener une surpolarisation de la densité de charges QM dans le formalisme EE et PE [205]. La position de l'atome liant HLA situé sur l'axe X_i - Y_j à une distance $R(X_1 - \text{HLA})$ est souvent fixée par une règle empirique proposée par Maseras et Morokuma afin aussi de diminuer le nombre de degrés de liberté dû à la présence de cet atome [198, 199] :

$$R(X_1 - \text{HLA}) = c_{\text{HLA}} \cdot R(X_1 - Y_1) \quad (4.8)$$

avec le facteur c_{HLA} qui est souvent déterminé par le rapport des rayons covalents R_o :

$$c_{\text{HLA}} = \frac{R_o(X_1 - \text{HLA})}{R_o(X_1 - C_2)} \quad (4.9)$$

Sa valeur est souvent proche de 0.7.

²La coupure d'une liaison peptidique en LA entraîne une instabilité dans la fonction d'onde [205].

Des solutions pour remédier à cette surpolarisation sont :

- d'utiliser une base de fonctions atomiques assez petite et donc très localisée sur les atomes.
- de supprimer du Hamiltonien la charge ponctuelle de l'atome Y_1 , ainsi que celles de certains atomes MM directement liés à ce dernier atome Y_1 . Si la charge de Y_1 est annulée, on parlera du schéma Z1, si les charges de Y_1 et C_2 sont annulées, on dira Z2, ... Dans Gaussian09, c'est le schéma Z3 qui est employé par défaut dans l'approche ONIOM. Il est aussi possible d'atténuer les charge de Y_1 , C_2 et C_3 mais cela entraîne la non-neutralité de la partie MM, pouvant introduire des artefacts.
- de redistribuer (atténuer) la charge ponctuelle de l'atome Y_1 sur les atomes MM voisins.

Dans tous les cas, la charge portée par le LA est nulle.

4.2.1.A L'approche ONIOM

Dans cette section, nous allons rapidement évoquer l'approche ONIOM (OM:OM:MM– *Our N-layered Integrated Molecular Orbital and Molecular Mechanics*) qui a été développée par Morokuma et collaborateurs [199, 200] en réunissant deux méthodes : IMOMO (QM:QM') [211] et IMOMM (QM:MM) [198, 212]. Cette méthode est présente dans de nombreux logiciels dont Gaussian. Elle permet de traiter des systèmes à plusieurs couches (souvent trois). Une partie d'intérêt est traitée à un haut niveau de théorie, son environnement proche (QM') à un niveau de théorie inférieur (souvent semi-empirique) et enfin le reste de la structure en MM. Lors du traitement d'un système QM/MM, c'est le formalisme HLA qui est le plus employé. Notons que les CF dans Gaussian sont limités à UFF, Dreiding et Amber96. Une autre caractéristique de ONIOM est son calcul de l'énergie totale :

$$E_{\text{QM/MM}}(\mathcal{S}) = E_{\text{MM}}(\mathcal{S}) + E_{\text{QM}}(\mathcal{I} + \mathcal{LA}) - E_{\text{MM}}(\mathcal{I} + \mathcal{LA}) \quad (4.10)$$

De manière générale, cette méthode peut être considérée comme une méthode MM dans laquelle une région d'intérêt a été séparée et traitée avec un niveau de théorie plus élevé. Dans l'équation 4.10, aucun terme frontière n'apparaît, ce qui, d'un point de vue pratique, simplifie fortement le modèle, évitant toute modification des procédures habituelles en QM et en MM. L'étape de soustraction corrige implicitement l'artefact dû à la présence du LA. En fait, le couplage entre les deux sous-systèmes est traité entièrement au niveau MM. Cela s'avère relativement problématique pour les contributions électrostatiques car elles sont traitées comme de simples

charges ponctuelles rigides. En 2006, Vreven et Morokuma ont remédié à ce problème de l'EE qui n'était pas considéré en modifiant le calcul de l'énergie [202]. Ils ont ainsi pris le même schéma que l'Eq. (4.5) qui permet de prendre en compte l'EE. Cela a été implanté dans Gaussian03 dès la version C.02 [143].

4.2.2 Les approches des orbitales localisées

Dans ces approches, la liaison pendante entre les atomes frontières QM et MM est représentée à l'aide d'orbitales hybrides localisées (OL). Plusieurs approches ont été développées : la méthode du Local-SCF (*Local self-Consistent Field*) que nous verrons en détail dans la section suivante, l'approche des orbitales gelées (FO, *Frozen Orbitals*), des orbitales hybrides généralisées (GHO, *generalized hybrid orbitals*) et du potentiel de fragment effectif (EFP, *Effective Fragment Potentiels*).

De manière générale, les méthodes qui utilisent les orbitales localisées sont théoriquement plus réalistes que les approches *Link Atom* car elles décrivent de manière quantique la distribution de charges autour de la frontière QM/MM. La délocalisation des charges quantiques dans les OL aide à empêcher ou à limiter la surpolarisation qui est parfois présente dans les méthodes LA.

L'approche LSCF utilise des orbitales localisées pour décrire la liaison entre les atomes frontières des deux sous-systèmes. Une approche similaire, celle des orbitales gelées (FO), a été développée par Friesner [213,214] dans laquelle les interactions QM-MM dans la région frontière sont très fortement paramétrisées.

Dans l'approche GHO de Gao [215–219], un ensemble d'orbitales hybrides sp^3 est assigné à chaque atome frontière Y_i . L'orbitale hybride se trouvant au travers de la liaison frontière est nommée orbitale active et participe au cycle SCF, les trois autres sont qualifiées d'hybrides auxiliaires et sont gelées durant le cycle SCF, ces dernières ne peuvent pas se mélanger avec les autres orbitales .

La méthode EFP développée par Gordon [220,221] se différencie des autres méthodes d'orbitales localisées. Une partie, dite active du système, est traitée avec la QM *ab initio* et le reste est traité avec un potentiel de fragments effectifs (EFP) qui est lui même défini à partir de calculs *ab initio*. Le EFP est ensuite introduit dans l'opérateur Hamiltonien du calcul QM, où il décrit les effets de l'environnement sur la partie QM. Cette approche a été essentiellement employée pour décrire des solutés en solution, puis a été étendue aux systèmes biologiques. Dans ce dernier

cas, le système est scindé en trois parties : une partie QM, un région tampon dont les OM sont gelées et le reste de la structure décrite par l'EFP.

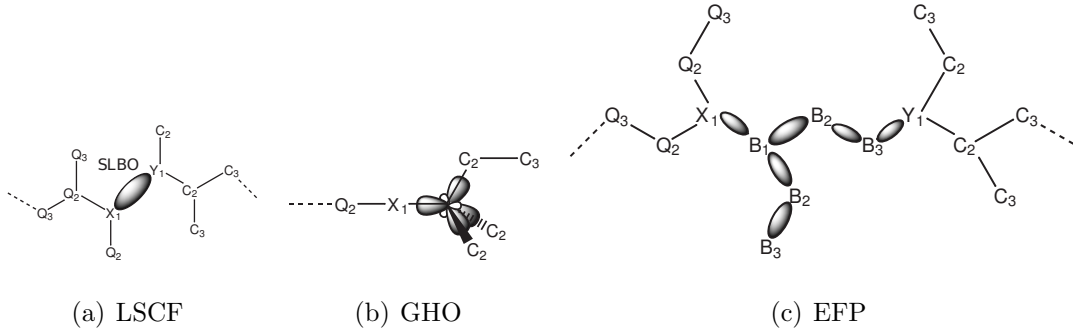


Figure 4.5: Notation pour les systèmes QM/MM. Q désigne les atomes de la partie QM et C ceux de la partie MM. Dans le schéma EFP, la zone tampon est nommée B. Le premier atome de la partie QM (MM) est noté X_1 (Y_1).

4.3 Le Local Self-Consistent Field/MM (LSCF/MM)

CETTE partie est entièrement consacrée à l'approche QM/MM développée à Nancy, le *Local Self-Consistent Field* (LSCF). Celle-ci traite à l'aide d'orbitales localisées la liaison covalente délicate à traiter dans un système où la partie QM est liée à la partie MM. Nous présenterons ces récents développements. Cette méthode fut développée par V. Théry dans laquelle la partie quantique était traitée à l'aide de méthode semi-empirique [2,3]. Elle fut ensuite étendue aux méthodes *ab initio* par X. Assfeld [4].

Dans cette approche, la liaison chimique entre le premier atome QM (X_i) et le premier atome MM (Y_i) est décrite par une orbitale localisée $|l_i\rangle$. On définit les N orbitales atomiques $|\lambda\rangle$, $|\mu\rangle$, $|\nu\rangle$ sur tous les atomes de la partie QM et sur l'atome frontière Y. Cet atome frontière est qualifié d'hybride ou d'atome quanto-classique car il possède des orbitales de fonction atomiques ainsi que tous les paramètres propres au type d'atome qu'il possède dans la partie MM. Les orbitales localisées sont décrites comme une combinaison linéaire des orbitales de fonctions atomiques :

$$|l_i\rangle = \sum_{\lambda}^N a_{\lambda i} |\lambda\rangle \quad (4.11)$$

Pour respecter les contraintes d'orthogonalité entre toutes les orbitales, il est crucial que les OM soient orthogonales aux OL. Un moyen de faire cela est de construire les OM dans une base

de fonctions atomiques à la fois orthogonales entre elles et orthogonales aux OL. La première étape de ce processus est l'orthogonalisation des OL entre elles afin d'obtenir un jeu d'orbitales localisées orthogonales (OLO) :

$$|l'_i\rangle = \sum_{\lambda}^N a'_{\lambda i} |\lambda\rangle \quad (4.12)$$

Les N orbitales atomiques du jeu de base de fonctions atomiques peuvent être ensuite orthogonalisées à ces OLO en les projetant dans le sous-espace complémentaire des OLO grâce à une fonction :

$$|\tilde{\mu}\rangle = \left[1 - \sum_i^L S_{\mu i}^2 \right]^{-1/2} \left[|\mu\rangle - \sum_i^L |l'_i\rangle \langle l'_i | \mu \rangle \right] \quad (4.13)$$

avec l'intégrale de recouvrement

$$S_{\mu i} = \sum_{\lambda}^N a'_{\lambda i} S_{\mu \lambda} \quad (4.14)$$

La fonction $|\tilde{\mu}\rangle$ est déduite à partir des fonctions originales $|\mu\rangle$ par une matrice \mathbf{M} de dimension $(N \times N)$ où les éléments (μ, ν) peuvent être écrits :

$$M_{\mu\nu} = \left[1 - \sum_i^L S_{\mu i}^2 \right]^{-1/2} \left[\delta_{\mu\nu} - \sum_{\nu}^L \sum_{\eta}^N a'_{\nu i} a'_{\eta i} S_{\eta\mu} \right] \quad (4.15)$$

Cependant, ces fonctions contiennent L dépendances linéaires dûes aux OL qui sont déjà des combinaisons linéaires des OA. Si la base de fonctions atomiques originale est linéairement indépendante, la matrice de l'intégrale de recouvrement \tilde{S} entre les OA orthogonales aux OL $|\tilde{\mu}\rangle$ a L valeurs propres nulles. Après avoir supprimé les valeurs propres et les vecteurs propres correspondants, il est alors possible d'orthogonaliser de manière canonique à l'aide de la matrice \mathbf{X} de dimension $(N \times (N - L))$:

$$X_{pq} = U_{pq} \tilde{s}_q^{-1/2} \quad (4.16)$$

où U_{pq} est le $p^{\text{ième}}$ composant du $q^{\text{ième}}$ vecteur propre de la matrice de recouvrement \tilde{S} entre les fonctions de bases atomiques $|\tilde{\mu}\rangle$ et \tilde{s} est la valeur propre correspondante [13]. Les $(N - L)$ fonctions $|\lambda'\rangle, |\mu'\rangle, |\nu'\rangle, \dots$ sont déduites de $|\tilde{\lambda}\rangle, |\tilde{\mu}\rangle, |\tilde{\nu}\rangle$ et par cette transformation elles sont linéairement indépendantes, mutuellement orthogonales et orthogonales aux OL. Les combinaisons de ces deux transformations définissent la matrice \mathbf{B} de dimension $(N \times (N - L))$:

$$\mathbf{B} = \mathbf{M} \cdot \mathbf{X} \quad (4.17)$$

La matrice \mathbf{B} joue le rôle de $\mathbf{S}^{-1/2}$ dans la procédure habituelle du SCF. Les éléments du Fockien sont définis de manière habituelle :

$$F_{\mu\nu} = H_{\mu'\nu'}^c + \sum_{\lambda\sigma} P_{\lambda\sigma}^T \left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu) \right] \quad (4.18)$$

Le premier terme est le Hamiltonien de cœur et le second représente les intégrales biélectroniques $(\mu\nu|\lambda\sigma)$. P^T est la matrice densité du système que l'on scinde en deux composantes dans le formalisme LSCF : une contribution due aux FOs (P^F) et une seconde due aux OM variationnelles (P^Q).

$$P_{\mu\nu}^T = P_{\mu\nu}^Q + P_{\mu\nu}^F \quad (4.19)$$

La matrice de Fock \mathbf{F} de dimension $(N \times N)$ est ensuite transformée dans la base $\{|\mu'\rangle\}_{1 \leq \mu \leq (N-L)}$ à l'aide de la matrice \mathbf{B} pour obtenir le nouveau Fockien \mathbf{F}' de dimension $((N-L) \times (N-L))$ ($\mathbf{F}' = \mathbf{B}^\dagger \cdot \mathbf{F} \cdot \mathbf{B}$). Ensuite, il faut diagonaliser ce nouveau Fockien \mathbf{F}' pour obtenir la matrice de $(N-L)$ vecteurs propres \mathbf{C}' et les $(N-L)$ valeurs propres (ϵ). Il faut par après re-transformer le jeu de base de fonctions atomiques initial ($\mathbf{C} = \mathbf{B} \cdot \mathbf{C}'$) pour calculer la nouvelle matrice densité correspondant aux MOs (\mathbf{P}^Q). En ajoutant la contribution des FOs (\mathbf{P}^F), on obtient alors la matrice de densité totale du système.

4.3.1 La modification des équations de Roothaan et l'algorithme

La condition de stationnarité de l'énergie par rapport aux coefficients de MOs variationnelles $\{c_{\mu p}\}_{\substack{1 \leq \mu \leq N \\ 1 \leq p \leq N-L}}$ est,

$$\frac{\delta E}{\delta c_{\mu p}} = 0 \quad (4.20)$$

et la contrainte d'orthogonalité entre toutes les orbitales (FOs et MOs) du système,

$$\sum_{\mu\nu} c_{\mu p} S_{\mu\nu} c_{\nu q} = \delta_{pq} \quad (4.21)$$

$$\sum_{\mu\nu} l_{\mu P} S_{\mu\nu} l_{\nu Q} = \delta_{PQ} \quad (4.22)$$

$$\sum_{\mu\nu} c_{\mu p} S_{\mu\nu} l_{\nu Q} = 0 \quad (4.23)$$

modifient les équations de Roothaan [5, 11] :

$$\mathbf{F} \cdot \mathbf{C} = \mathbf{S} \cdot \mathbf{C} \cdot \mathbf{E} + \mathbf{S} \cdot \mathbf{L} \cdot \mathbf{\Lambda} \quad (4.24)$$

où \mathbf{L} est la matrice des FOs et $\{\Lambda_{pQ}\}_{\substack{1 \leq p \leq (N-L) \\ 1 \leq Q \leq L}}$ sont les multiplicateurs de Lagrange du bloc des MOs-FOs :

$$\Lambda_{pQ} \equiv \varepsilon_{pQ} = \sum_{\mu\nu} c_{\mu p} F_{\mu\nu} l_{\nu Q} \quad (4.25)$$

En utilisant les relations $\mathbf{C} = \mathbf{B} \cdot \mathbf{C}'$ et $\mathbf{X}^\dagger \cdot \mathbf{S} \cdot \mathbf{X} = \mathbb{I}$, et en multipliant par la gauche par \mathbf{B}^\dagger , on obtient :

$$\mathbf{F}' \cdot \mathbf{C}' = \mathbf{C}' \cdot \mathbf{E} \quad (4.26)$$

avec $\mathbf{F}' = \mathbf{B}^\dagger \cdot \mathbf{F} \cdot \mathbf{B}$ et $\mathbf{C}' \cdot \mathbf{S} \cdot \mathbf{L} = \mathbb{O}$ (orthogonalité entre les MOs et les FOs). On aboutit à un problème hermitien qui peut être résolu en diagonalisant la matrice \mathbf{F}' . L'algorithme SCF dans Gaussian03 a été modifié de la manière suivante par X. Assfeld :

1. Orthogonaliser mutuellement les FOs.
2. Construire la matrice \mathbf{M} , orthogonalisant la base des AO aux FO.
3. Construire la matrice \mathbf{X} .
4. Transformer les fonctions obtenues en 2. en une base de fonctions orthogonales et linéairement indépendantes *via* $\mathbf{B} = \mathbf{M} \cdot \mathbf{X}$.
5. Définir la matrice de Fock \mathbf{F} dans la base originale.
6. Calculer $\mathbf{F}' = \mathbf{B}^\dagger \cdot \mathbf{F} \cdot \mathbf{B}$.
7. Diagonaliser \mathbf{F}' : $\epsilon = \mathbf{C}'^\dagger \cdot \mathbf{F}' \cdot \mathbf{C}'$ où ϵ est la matrice diagonale des valeurs propres.
8. Transformer les vecteurs propres dans la base originale : $\mathbf{C} = \mathbf{B} \cdot \mathbf{C}'$.
9. Calculer \mathbf{P}^Q .
10. Test de sortie. Si non satisfait, retour à 5.

4.3.2 Le traitement des orbitales gelées en LSCF

Dans le LSCF, il est donc possible de geler des orbitales. Celles-ci peuvent être occupées ou vides. Nous utiliserons ce formalisme pour décrire les ionisations de cœur et en les comparant à celles obtenues par la méthode de la projection asymptotique (cf. chapitre 7). Dans un système QM/MM, nous utilisons une orbitale liante strictement localisée (SLBO, *Strictly Localised Bond Orbital*) pour décrire la liaison frontière. Cette orbitale localisée sera définie entre les deux

atomes X_i et Y_i . L'atome frontière (Y_i) est dit atome hybride car il possède les bases de fonction atomique et des paramètres MM de par sa description en tant qu'atome quanto-classique. La table suivante reprend toutes les interactions qu'il faut considérer dans la région frontière :

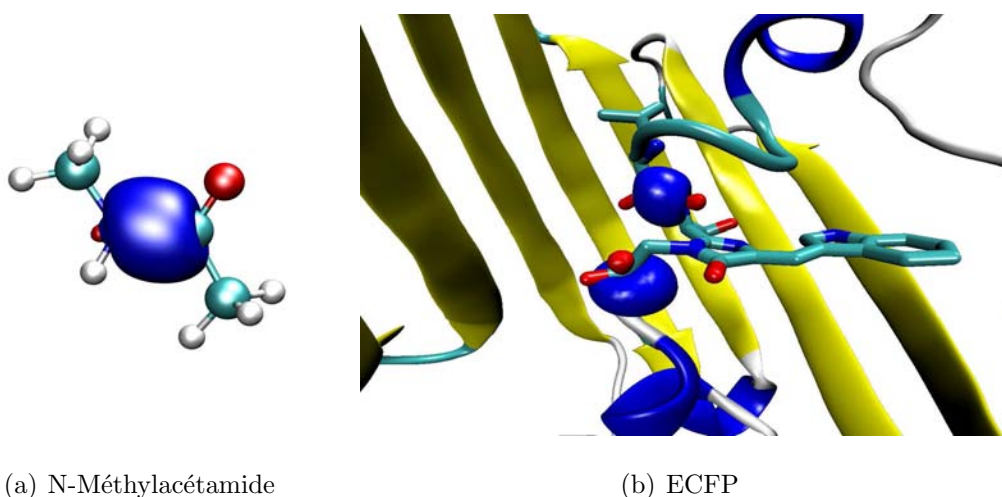
Termes liés	
Liaisons	Y-C ₁
Angles	X-Y-C ₁ , Y-C ₂ -C ₃
Torsions	Q ₂ -X-Y-C ₂ , X-Y-C ₂ -C ₃ , Y-C ₂ -C ₃ -C ₄
Termes non-liés	
vdw	Q_i -Y _k , Q_i -C _B , Y _k -C _B , Y _k -Y _l ^a
Electrostatique	Z-Z Z_{Q_i} -Z _{Y_k} , Z_{Y_k} -Z _{Y₁} ^a
	Z-q Z_{Q_i} -q _{Y_k} , Z_{Q_i} -q _{C_B} , Z_{Y_k} -q _{C_B} , Z_{Y_k} -q _{Y₁} ^a
	q-q q_{Y_k} -q _{C_B} , q_{Y_k} -q _{Y₁} ^a
	e-Z $\sum_k \sum_{\mu\nu} P_{\mu\nu}^T \left\langle \mu \left \frac{Z_{Y_k}}{r_k - r} \right \nu \right\rangle$
	e-q $\sum_k \sum_{\mu\nu} P_{\mu\nu}^T \left\langle \mu \left \frac{q_{Y_k}}{r_k - r} \right \nu \right\rangle, \sum_{A \in \text{MM}} \sum_{\mu\nu} P_{\mu\nu}^T \left\langle \mu \left \frac{q_{C_A}}{R_A - r} \right \nu \right\rangle$

Table 4.1: Description des interactions QM/MM en LSCF. Les atomes décrits en QM (MM) sont notés Q (C), Z_A et q_A correspondent à la charge nucléaire et à la charge ponctuelle de l'atome A . Les éléments de la matrice densité sont $P_{\mu\nu}$, avec μ et ν^{th} les indices des OA. ^a $k \neq l$

Afin d'assurer la neutralité du système entier, une charge de +1 est ajoutée à la charge MM de l'atome hybride parce que ce dernier contribue à la SLBO en lui donnant un électron.

Les SLBO sont déterminées par l'utilisateur sur un système modèle qui contient la liaison d'intérêt que l'on souhaite couper dans le système originel. Grâce au principe de transférabilité, les coefficients de la SLBO calculés pour le système modèle peuvent être appliqués (après rotation et normalisation) à la liaison d'intérêt dans le système réel. Il est cependant important de tester la position de la SLBO afin qu'elle minimise le nombre d'artefacts lors du calcul d'une propriété. Nous verrons que nous avons étudié cette position lors de l'étude sur la protéine fluorescente ECFP (voir section 5.3.3.A). Pour des systèmes biologiques, il serait intéressant de scinder le système en deux au niveau de la liaison peptidique. Le système modèle sera alors le N-Méthylacétamide.

La SLBO peut être localisée par différents critères regroupés en deux catégories : interne et



externe. Cette localisation est essentielle, il est donc nécessaire de tester les différents critères de localisation afin d'utiliser celui qui conviendra le mieux. Dans les critères externes, l'utilisateur doit fixer lui-même entre quels atomes la MO doit être localisée. C'est le cas des approches de Weinstein-Paucz (WP) [222, 223] et de Magnasco-Perico (MP) [224, 225]. Ces deux critères maximisent la population du recouvrement entre les deux atomes X et Y. MP maximise, en plus, la population électronique entre les deux atomes. Pipek-Mezey (PM) [226], Boys-Foster (BF) [227, 228] et Edmiston-Ruedenberg (ER) [229, 230] sont tous trois des critères de localisation internes. Le critère PM minimise la somme du carré des charges atomiques de Mulliken divisé par le nombre d'électron total, celui de BF minimise la distance au carré des deux électrons et la localisation selon ER maximise l'énergie de *self-repulsion* (inverse de la distance au carré). Le lecteur, pour plus de précisions, pourra lire le papier de Pipek-Mezey [226] qui reprend les concepts de tous les critères.

N. Ferré et collaborateurs ont ensuite montré que la SLBO et la charge nucléaire +1 attribuée à l'atome Y_i (LSCF/+1) ne suffisent pas à décrire correctement la distance inter-atomique entre X_i et Y_i [5]. En effet, les interactions électrostatiques sont sous-estimées parce que la charge nucléaire de l'atome quantique Y_i ont été remplacés par une charge nucléaire de +1. De plus, les variations du recouvrement entre les bases de fonctions atomiques en fonction de la longueur de liaison ne sont pas considérées. Pour cela, il a été nécessaire de définir un potentiel de liaison frontière (FBP, *Frontier Bond Potential*) spécifique à la liaison X_i - Y_i .

$$V_{X-Y} = (A + Br + Cr^2) e^{Dr} + \frac{E}{r_{X-Y}} \quad (4.27)$$

Le premier terme de l'Eq. (4.27) est le terme correctif pour prendre en compte les variations du recouvrement entre les OA des deux atomes en fonction de leur distance r (forme du potentiel prise d'après le recouvrement de deux orbitales 1s de Slater, $D < 0$) et le second terme est un terme correctif quant à la trop faible répulsion électrostatique entre les deux atomes. Ainsi, ils ont défini les paramètres du potentiel pour les liaisons frontières suivantes présentes dans les molécules organiques : C(sp3)-C(sp3), C(sp3)-O, O-C(sp3), C(sp3)-N, N-C(sp3), C(sp2)-N, N-C(sp2), C(sp3)-C(sp2) et C(sp2)-C(sp3) [5].

La définition de ce potentiel n'est évidemment pas triviale. Pour toutes autres liaisons frontières, il est nécessaire de redéfinir un ensemble de paramètres ce qui est laborieux. Une approche alternative pour obtenir des distances inter-atomiques correctes et pour supprimer ce FBP est de considérer explicitement les deux électrons de cœur de l'atome Y_i . Deux électrons seront considérés pour les atomes de la seconde rangée d'atomes du tableau périodique (Li \rightarrow Ne), dix électrons pour ceux de la troisième rangée (Na \rightarrow Ar), ... Le formalisme pour les atomes de la seconde rangée est appelé LSCF/+3. De cette manière, trois électrons de Y_i sont explicitement inclus dans le calcul SCF (un dans la SLBO et deux électrons de cœur de l'atome Y_i). Une charge nucléaire de +3 lui est assignée pour garantir la neutralité du système [7, 10]. Le LSCF/+3 a été testé à l'aide de deux approches : l'une plaçait les électrons de cœur dans une orbitale de cœur gelée (FCO, *Frozen Core Orbital*), voir Figure (4.6) [7], l'autre les considérait comme tous les autres électrons, c'est-à-dire dans une orbitale de cœur auto-cohérente (SCCO, *Self-consistent Core Orbital*) [10]. Les deux schémas ont des résultats comparables et ont été validés sur des critères géométriques et énergétiques sur des liaisons de nature différente : C-C, C-N, C-O [10]. Nous avons choisi dans le travail sur les protéines fluorescentes et sur l'albumine humaine l'approche SCCO.

Le SCCO a été auparavant employée sur de simples systèmes comme l'éthane, la méthylamine et le méthanol mais aussi dans une étude qui avait pour sujet les modulations de l'affinité électronique d'un acide nucléique par son environnement direct : simple et double brin de 3'-dCMPH (*2'-deoxycytidine-3'-monophosphate*) [231]. Les SLBO étaient alors placées de part et d'autre de l'acide nucléique (base, phosphate et sucre) sur la liaison C'₅-O'₅. Le choix de la position des SLBO est logique, les liaisons C'₃-O'₃ ayant une forte affinité électronique [232, 233].

Jusqu'ici, les liaisons frontières dans les systèmes traités en LSCF/+3 possédaient un carac-

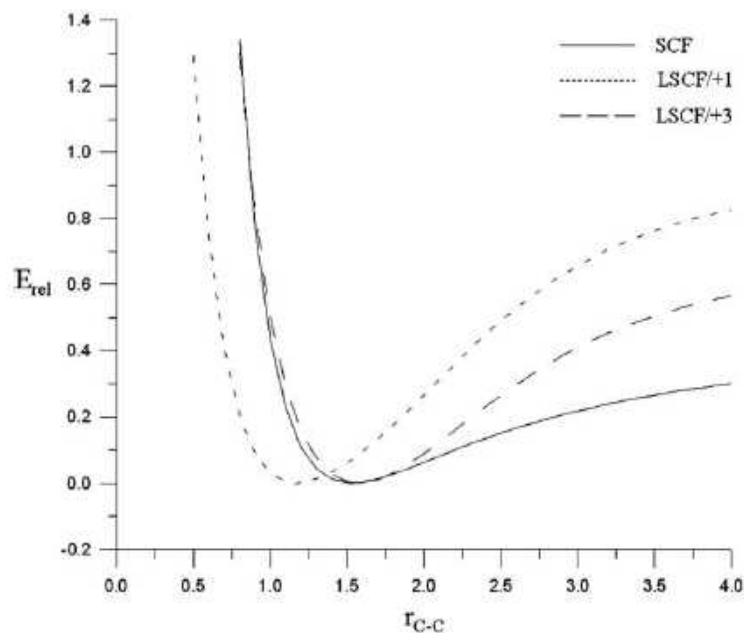


Figure 4.6: Profil de l'énergie (kcal/mole) en fonction de la distance inter-atomique (Å) entre les deux atomes de carbone du fragment $\text{CH}_3\text{-C}$ de l'éthane [7]. Les deux électrons de cœur dans LSCF/+3 sont traités par le schéma FCO.

tère de simple liaison (éthane, méthane, méthanol, ADN). P.-F. Loos et X. Assfeld ont alors considéré une liaison plus délicate à traiter et pourtant omniprésente dans les systèmes biologiques : la liaison peptidique. Celle-ci a un caractère partiel de double liaison par la présence d'un doublet libre de l'atome d'azote en conjugaison. Le nombre d'électrons entre l'atome de carbone et l'atome d'azote sera effectivement supérieur à deux. Ce nombre peut même se voir modifié en fonction de l'environnement de la liaison peptidique. La forme zwitterionique (Figure (4.8)) sera, par exemple, favorisée si des molécules d'eau [234,235] ou un acide aminé chargé se rapproche du carbonyle. Lors de l'étude d'un système biologique par la méthode LSCF/MM, deux cas peuvent se présenter. Soit l'atome de carbone de la liaison peptidique correspond à l'atome hybride Y_1 et l'atome d'azote est entièrement quantique X_1 ; soit l'inverse. Dans le premier cas, la distance inter-atomique est bien reproduite avec le schéma LSCF/+3. En effet, la paire libre de l'atome d'azote est complètement décrite quantiquement. Dans le second cas, un traitement de l'atome d'azote avec LSCF/+3 signifie que le doublet libre de l'atome d'azote est complètement négligé

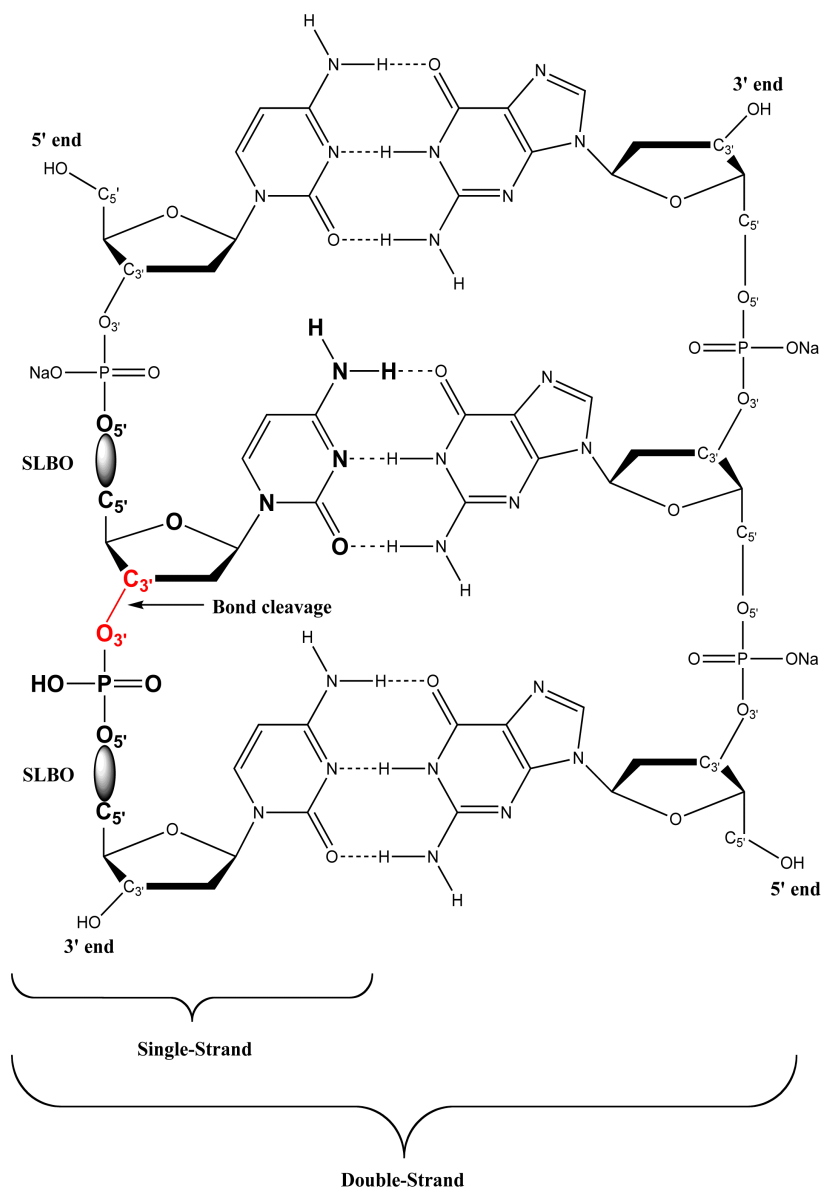


Figure 4.7: Partition QM/MM du modèle simple et double brin dCMPH. La partie QM est en trait gras. [231]

dans le calcul et la liaison C-N en est directement affectée (trop grande) [9]. P.-F. Loos a donc proposé que dans ce cas (l'atome d'azote est l'atome quanto-classique), les deux électrons de la paire libre soient explicitement considérés dans le calcul SCF. Cela revient à attribuer une charge nucléaire de +5 à cet atome hybride. Le schéma traitant ces liaisons peptidiques sera noté LSCF/+3,+5.

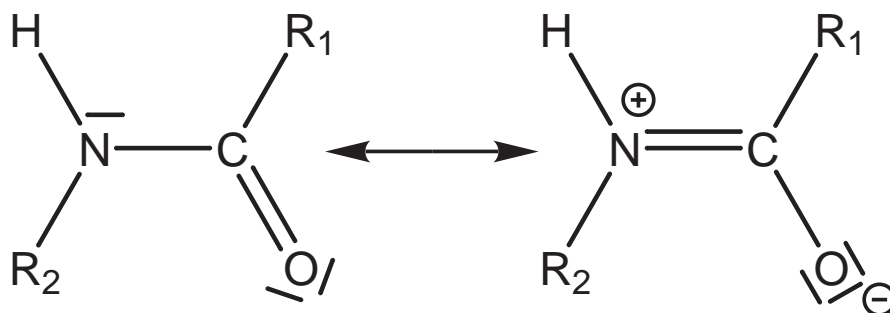


Figure 4.8: Formes de résonance de la liaison peptidique.

Le formalisme LSCF a donc été adapté pour traiter correctement des liaisons chimiques de nature et de complexité croissante. Selon la nature de ces liaisons frontières, LSCF/+3 sera utilisé (exemple ADN) ou LSCF/+3,+5 (protéine fluorescente). Un avantage majeur de la partition au niveau de la liaison peptidique (LSCF/+3, +5) est qu'il constitue une coupure naturelle des acides aminés. Ceux-ci seront donc traités de manière symétrique en QM et la charge des acides aminés adjacents à la région QM est entière dans la plupart des champs de forces. Il ne faudra donc pas le recalculer comme c'était le cas lorsque la coupure se trouvait entre le C_α - C_β ou entre $C(=O)$ et C_α (Figure (4.9)).

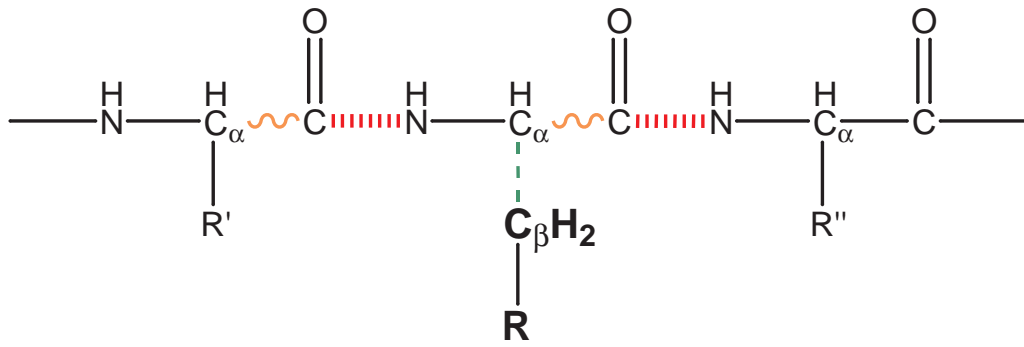


Figure 4.9: Partition schématique d'un polypeptide. Les différentes localisations de la jonction QM/MM sont représentées : le cas de la liaison C_{α} - C_{β} est représenté par une ligne discontinue verte, les liaisons C_{α} - $C(=O)$ par un trait ondulé orange, tandis que les liaisons peptidiques sont en bâtonnets rouges. Les atomes de la chaîne latérale appartenant à la partie QM sont représentés en gras.

Troisième partie

Les phénomènes électroniques appliqués aux macromolécules

Présentation

POUR rappel, cette partie constitue les applications des différentes méthodologies explicitées dans les deux parties précédentes.

Le premier chapitre est dédié aux spectres d'absorption UV/visible et contient la méthodologie que nous avons élaborée (LSCF/MM:ERS) pour prendre en compte la réponse électronique de l'environnement du chromophore. Elle est employée sur le complexe de squaraine-tétralactame ainsi que sur les protéines fluorescentes.

Le second chapitre reprend le travail réalisé sur la capture électronique d'une macromolécule contenant trois ponts disulfures.

Enfin le chapitre 7 concerne les ionisations de cœur des acides aminés et de la sérine albumine humaine. Ces ionisations ont été obtenues à l'aide de l'approche de la projection asymptotique développée par V. Glushkov pour orthogonaliser les orbitales gelées.

Chapitre 5

Absorption UV/visible

A FIN d'avoir un regard global sur cette partie dédiée à la photochimie [236–238], il est intéressant de présenter quelques notions fondamentales concernant les processus liés à l'absorption ou émission de la lumière (photon). Le diagramme de Perrin-Jablonski (Figure 5.1) rassemble les différents phénomènes électroniques liés à l'absorption d'une onde électromagnétique.

L'état fondamental singulet est noté S_0 et les états électroniques excités singulets et triplets sont notés respectivement S_1 , S_2 , ... et T_1 , T_1 , ... Des niveaux vibrationnels sont associés à chaque état. Il existe deux grandes classes de processus ; les processus radiatifs et les processus non-radiatifs. Lors des processus radiatifs, le système absorbe ou émet un photon (absorption, fluorescence, phosphorescence) tandis que les procédés non radiatifs libèrent leur énergie sous forme de chaleur dégagée lors de la relaxation vibrationnelle. Lors de procédé non-radiatif, l'énergie accumulée est, en fait, redistribuée à l'environnement (solvant, protéine). Ce procédé est nommé la conversion interne (sous entendu de l'énergie). Cette conversion interne (CI) se déroule entre des états de même multiplicité de spin selon une relaxation vibrationnelle vers le niveau le plus bas en énergie, par exemple, $S_2 \rightarrow S_1$. La conversion interne de l'état S_1 vers l'état fondamental S_0 est possible mais moins fréquente que $S_2 \rightarrow S_1$ parce qu'elle possède une plus grande différence d'énergie entre les deux états électroniques. Par conséquent, la conversion interne de $S_1 \rightarrow S_0$ peut entrer en compétition avec l'émission d'un photon (fluorescence) ou le passage/croisement intersystème (CIS) vers l'état triplet à partir duquel il est aussi possible d'émettre un photon (phosphorescence). Le processus radiatif qui fait passer une molécule de son état fondamental à un état excité singulet par l'absorption d'un photon est l'absorption.

Ce dernier phénomène est très rapide (10^{-15} s) de telle sorte que les noyaux de la molécule ne se déplacent pas lors de la transition électronique (principe de Franck-Condon [152–154]). Les processus d'émission sont tout aussi rapides (10^{-15} s) et suivent ce même principe. Les trois phénomènes radiatifs (l'absorption, la fluorescence et la phosphorescence) sont appelés transitions électroniques verticales parce que leur configuration nucléaire reste figée.

L'efficacité avec laquelle un photon caractérisé par une longueur d'onde λ , est absorbé par le chromophore est décrite par l'absorbance (A) qui suit généralement la relation linéaire de Lambert-Beer¹ :

$$A = \log \left(\frac{I_0}{I} \right) = \epsilon(\lambda)lc \quad (5.1)$$

où I_0 et I sont respectivement les intensités lumineuses des faisceaux incident et transmis. La loi de Lambert-Beer relie l'absorbance (sans dimension) à la concentration molaire (c) de la structure absorbante en (mole L^{-1}). $\epsilon(\lambda)$ est le coefficient d'absorption molaire ($L \text{ mole}^{-1} \text{ cm}^{-1}$) et l est l'épaisseur de la cellule contenant la solution (en cm). Dans ce contexte, le système d'unité utilisé est CGS. Le coefficient d'absorption molaire décrit l'efficacité avec laquelle le chromophore absorbe la lumière dans un solvant donné. C'est, en fait, la probabilité de la transition (P_{ij}) entre un état initial (i) et un état final (j).

$$P_{ij} = \left(\frac{8\pi^3 e^2}{3h^2 c} \right) G_j D_{ij} \quad (5.2)$$

où, e est la charge d'un électron (4.803×10^{-10} esu), h est la constante de Planck (6.626×10^{-27} erg s), G_j est le poids statistique de l'état final et D_{ij} est l'intensité du dipôle de transition défini comme suit :

$$D_{ij} = \left(\int \langle \Psi_i | \mu | \Psi_j \rangle d\tau \right)^2 \quad (5.3)$$

L'absorption ou l'émission de la lumière, onde électromagnétique caractérisée par une certaine fréquence ν , ne se passera que si la molécule/chromophore possède un dipole (μ) oscillant à cette même fréquence. Le coefficient d'absorption molaire peut ainsi être associé à la force d'oscillateur (f) d'une transition électronique :

$$f = I - I_0 = \left(\frac{8\pi^2 mc}{3h} \right) G_j D_{ij}(\nu) \quad (5.4)$$

$$= 4.315 \times 10^{-9} \int \epsilon(\nu) d\nu \quad (5.5)$$

¹Si, l'absorbance ne suit pas une relation linéaire avec la concentration, la formation d'agrégats ou d'autres espèces absorbantes peuvent être invoquées.

où m est la masse d'un électron (9.109×10^{-28} g), c la vitesse de la lumière (2.998×10^{10} cm s⁻¹). Nous avons vu dans le chapitre 2 comment ces transitions électroniques ainsi que la force d'oscillateur associée sont obtenues à l'aide de calculs.

Les molécules qui absorbent ou émettent des photons sont des chromophores. Ceux-ci sont fortement conjugués (alternance de simple et de double liaisons). L'absorption des ondes électromagnétiques par le chromophore est due à l'excitation des électrons d'une molécule. Selon la nature des liaisons et des atomes, une molécule absorbe à certaines longueurs d'onde et apparaît colorée, ou pas. Cette propriété permet aux électrons de se délocaliser sur tout le système grâce au recouvrement des orbitales π . L'énergie d'une excitation $\pi^* \leftarrow \pi$ sera d'autant plus faible que le nuage électronique π est plus étendu ; en conséquence, le pic d'absorption se situera d'autant plus vers les grandes longueurs d'onde (vers le rouge). Cette règle simple s'applique aux systèmes conjugués linéaires (polyène) et cycliques (molécules aromatiques). Le greffage de différents substituants susceptibles de modifier la structure électronique du chromophore aboutit généralement à un déplacement du spectre d'absorption. On appelle auxochromes, les substituants électrodonneurs tels que les groupes hydroxyle (-OH), alkoxy (-OCH₃) ou amino (-NH₂) ou les substituants électroaccepteurs comme le carbonyle (-C=O), le nitro (-NO₂), le nitroso (-N=O). Ces auxochromes sont capables d'interagir avec le chromophore en étendant(stabilisant) ou non le système conjugué. De nombreux chromophores sont présents dans notre quotidien, dans les colorants (indigo, anthraquinone, coumarine, ...), les photochromes (matériaux optiques à transmission variable tels que les diaryéthènes, les azobenzènes, ...), ...

En plus de pouvoir être affectés par des groupes auxochromes, les chromophores sont aussi très sensibles à leur environnement (solvant ou protéine). Le spectre d'absorption d'un chromophore en solution subit généralement un déplacement hypo- ou bathochromique (solvatochromisme négatif ou positif), c'est à dire un déplacement respectivement vers le bleu/petites longueurs d'onde ou vers le rouge/grandes longueurs d'onde. Dans le cas d'un déplacement bathochromique, l'état fondamental du chromophore est stabilisé par le solvant mais l'état excité l'est encore plus. D'un point de vue énergétique, l'écart d'énergie (ΔE) entre les orbitales responsables de la transition électronique diminue. Le phénomène inverse se produit pour un déplacement hypsochromique, l'état excité est moins stabilisé que l'état fondamental. Le solvatochromisme de chromophores a déjà été de nombreuses fois étudié tant expérimentalement que théoriquement et est relativement bien compris pour de nombreux composés.

Une situation plus délicate et relativement moins évidente à comprendre et encore moins à prévoir est celle du spectre d'absorption d'un chromophore se trouvant dans un milieu macromoléculaire. De la même manière, qu'en solution, ce spectre d'absorption sera considérablement modifié. Cette partie va tenter d'apporter des éléments de compréhension, d'un point de vue théorique, des effets de l'environnement macromoléculaire sur les spectres d'absorption de divers chromophores.

Les outils QM-MM permettent de prendre en compte cet environnement macromoléculaire de manière efficace. Ils seront couplés à la TD-DFT (section 2.2.2) pour évaluer les énergies de transition verticale de manière précise. Dans le chapitre QM-MM (chapitre 4), nous avons montré que le LSCF/MM incluait la contribution électrostatique de l'environnement MM ; en décrivant de manière explicite la polarisation électrostatique de la fonction d'onde (*Electrostatic Embedding*). Souvent, la prise en compte de cette contribution est suffisante pour déterminer de nombreuses propriétés comme celle de l'état fondamental. Cependant, lors d'une transition électronique, une situation de non-équilibre apparaît entre le colorant/chromophore et son environnement. Nous avons vu que lors de l'absorption d'un photon en phase gazeuse, la configuration nucléaire du chromophore restait figée (Principe de Franck-Condon [152–154]) ; cependant, le nuage électronique du chromophore est modifié suivant la fréquence du photon. Dans un solvant, à l'état fondamental, il y aura un équilibre entre noyaux et électrons du chromophore et des molécules de solvant. Lors de la transition électronique, la configuration nucléaire du chromophore et des molécules de solvant ne sera pas modifiée. Seuls les électrons du solvant seront assez rapides pour se réorganiser par rapport à la modification du nuage électronique du chromophore. Cette situation est nommée état de non-équilibre. Le nuage électronique des molécules de solvant doit alors se réarranger en fonction de la contribution électronique du chromophore. Cette réponse électronique de l'environnement (ERS, *Electronic Response of the Surroundings*) suite à la modification du nuage électronique du chromophore doit être prise en compte. Pratiquement, nous avons vu dans la section 2.3.2 que l'ERS est modélisé à l'aide d'un continuum isotrope caractérisé par la constante diélectrique extrapolée à fréquence infinie (constante diélectrique dynamique/optique, ϵ_∞).

Un chromophore dans un milieu macromoléculaire représente une situation plus complexe. Comme pour le solvant, il est nécessaire de modéliser la réponse électronique d'un environnement protéique (ERS) lors du calcul des propriétés optiques mais aussi lorsqu'un site actif

est en présence de contre-ions ou de résidus chargés. La partie suivante reprend brièvement les quelques approches utilisées actuellement pour décrire *correctement* l'environnement protéique.

5.1 La réponse électronique de l'environnement

UNE première approche est de modéliser l'entièreté de l'environnement de la protéine par un continuum caractérisé soit par la constante diélectrique de l'eau, soit par une constante diélectrique d'une valeur de quatre [61, 239]. Cette approche très *simpliste* modélise un environnement protéique totalement anisotrope par un continuum complètement isotrope ! Rappelons aussi que l'un des points faibles des modèles de continuum est que les interactions spécifiques (liaisons hydrogène) soluté-solvant/environnement ne sont pas parfaitement reproduites. Or, dans une protéine, le soluté (site actif) interagit très fortement via de nombreuses liaisons hydrogène ou interactions électrostatiques avec son environnement proche. D'autre part, la constante diélectrique utilisée pour modéliser l'intérieur des protéines peut varier entre 4 et 20 selon la polarité de l'environnement proche du site actif ou chromophore [240]. De simples calculs QM ou PCM-QM ne sont pas assez proches de la réalité de l'environnement protéique. Par contre, les outils QM-MM sont adéquats pour traiter des systèmes macromoléculaires/protéiques possédant un site d'intérêt comme un chromophore. Malheureusement, la contribution due à la polarisation électronique de l'environnement n'est pas prise en compte dans les champs de forces classiques. Alors que le développement de champ de forces classiques polarisables est en plein essor ; très peu de groupes développent des approches QM/MM_{pol} (MM_{pol} = MM avec un champ de forces classiques polarisables) qui incluraient donc la réponse électronique de l'environnement. Ces quelques méthodes sont actuellement relativement coûteuses en temps de calculs et difficilement *utilisables* par un utilisateur lambda surtout lorsqu'un résidu est non-usuel puisqu'il faudrait redéfinir toute une série de paramètres très dépendants du niveau de théorie.

Actuellement, les quelques approches QM-MM_{pol} ont été employées pour évaluer des spectres d'absorption de systèmes proches des protéines fluorescentes, les rhodopsines (cf. Figure 4.3). En 2001, Houjou développe une méthode QM-MM_{pol} où le MM_{pol} est le PMM (*Polarizable mosaic model*) [241]. Dans ce modèle, la partie MM est une mosaïque de cylindres caractérisant les liaisons avec des polarisabilités dépendantes de la nature de la liaison. Un jeu de charges induites par la densité électronique de l'état fondamental de la partie QM est défini. La matrice

de Fock prend en compte ce potentiel créé par les charges 'MM'. Ce groupe a calculé les spectres d'absorption des Rhodopsines [242], ils observent un déplacement bathochromique de +10 nm.

Dupuis et Hirao, en 2002, proposent une approche QM/MM_{pol} où la polarisation électronique de l'environnement est représentée au moyen de dipôles induits [243], ce modèle n'a cependant pas été testé sur des biomolécules.

Une autre approche, actuellement très en vogue est celle de Elstner et collaborateurs dans laquelle une couche d'atomes autour de la partie QM sont décrits à un niveau de théorie semi-empirique. C'est le principe des méthodes QM/QM'/MM ou de la SCC-DFTB (*Self-Consistent-Charge DFT tight-Binding*) [110, 244–247]. L'interaction QM/QM' dans ce modèle n'est que coulombienne, la polarisation mutuelle entre les deux sous-systèmes QM et QM' n'est donc que statique. La méthode néglige donc le transfert de charges, l'échange et la corrélation (dispersion). Un modèle polarisable semi-empirique a été mis ensuite au point, toujours par le même groupe, à l'aide du champ de forces AMOEBA qui inclut donc la polarisabilité atomique à l'aide de charges et de dipôles induits [247]. Ce modèle est relativement très efficace mais aucune indication n'a été donnée concernant le temps de calcul. Ces deux méthodes ont été utilisées en 2008 pour étudier le spectre d'absorption des dérivés de la rhodopsine (*bactério*-rhodopsine (bR) et la *pharaonis* sensory rhodopsine II (psRII)) [110, 247]. La méthode QM/QM'/MM entraîne un déplacement bathochromique de +15 nm et +19 nm par rapport à la méthode « classique » QM/MM tandis que l'approche polarisable montre un déplacement plus de deux fois plus grand (+37 nm et +30 nm).

Enfin, le groupe de Ulf Ryde associé à Olivucci a publié un papier [248], toujours sur la rhodopsine (l'isorhodopsine²), contenant un *benchmark* des champs de forces pour les méthodes QM-MM. L'effet de l'ERS est compris entre +6 et +10 nm avec le champ de forces polarisables Amber02 ou entre -4 et +13 nm en utilisant un champ de forces plus complexe contenant des polarisabilités atomiques anisotropiques et multipolaires.

5.1.1 Le couplage LSCF/MM:SCRF

Nous allons maintenant présenter la méthodologie que nous avons développée pour prendre en compte l'ERS. Cette méthode est simple et relativement efficace. Nous l'avons testée et validée sur deux types de systèmes : le complexe de squaraine-tétralactame et les protéines fluorescentes.

²Le 9-cis retinal est remplacé par le 11-cis retinal.

Ce modèle consiste à coupler le LSCF/MM avec un continuum caractérisé par une constante diélectrique extrapolée à fréquence infinie (ou constante diélectrique dynamique). Nous avons vu que la constante diélectrique peut être scindée en deux contributions : une contribution d'orientation et une contribution dynamique. En situation de non-équilibre, seule la contribution dynamique est présente. Ce continuum nous permet de ne pas négliger la réponse électronique de l'environnement, exactement comme les modèles de solvation considèrent la réponse électronique du solvant. Les contributions électrostatiques et les interactions 1-4 de l'environnement sont quant à elles modélisées par le champ de forces. Nous avons appelé ce formalisme LSCF/MM:ERS. Dans ce formalisme, l'énergie du système devient :

$$E = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM-MM}} + E_{\text{QM-SCRF}} \quad (5.6)$$

Tous ces termes ont été détaillés dans les deux chapitres relatifs à la méthodologie (les chapitres I et II). Aux trois premiers termes définissant l'énergie obtenue lors de calcul QM-MM, vient s'ajouter un terme lié au continuum caractérisé par ϵ_∞ (Eq. 2.70) qui fait intervenir le potentiel de réaction V_{R} . Dans ce cas, ϵ est égal à $\epsilon_{\text{dyn}} (= \epsilon_\infty)$. Les termes de cavitation, de dispersion et de répulsion ne sont pas présents dans ce terme car ceux-ci (excepté le terme de cavitation) sont décrits au niveau MM. L'opérateur de Fock devient, en combinant les trois Eq. (1.1), (4.7) et (2.70) :

$$\mathbf{F} = -\frac{\nabla_i^2}{2} - \sum_A \frac{\text{noyaux } Z_A}{r_{iA}} + \sum_j^{N/2} (2J_j(r_i) - K_j(r_i)) - \sum_{i \in \mathcal{I}} \sum_{M \in \mathcal{E}}^L \frac{q_M}{|\mathbf{R}_M - \mathbf{r}_i|} - \frac{1}{2} \iiint_{\text{cav}} V_{\text{R}} \quad (5.7)$$

Selon cette méthode, les charges MM polarisent la partie QM et donc la fonction d'onde. La densité électronique de la partie QM polarisée est utilisée pour définir les charges sur la surface de la cavité entourant la partie d'intérêt (QM). Il est important que les charges MM ne polarisent pas, en plus la cavité, indépendamment de la densité électronique QM qui la polarise déjà en prenant en compte les charges MM dans sa définition.

Cette approche a le grand avantage par rapport aux méthodes QM/MM_{pol} d'être très rapide. En outre, aucun paramètre ne doit être ajouté au champ de forces pour décrire la réponse électronique de l'environnement, même en cas de résidus non-usuels. Cependant, il faut garder à l'esprit que la réponse électronique est prise en compte de manière totalement isotropique dans ce modèle.

5.2 Le complexe de squaraine-tétralactame

La première application de notre méthodologie incluant la réponse électronique de l'environnement dans les spectres d'absorption (LSCF/MM:ERS) a été réalisée sur le complexe de squaraine-tétralactame (ST) et a fait l'objet d'une publication [112]. Ce projet s'inscrit dans le cadre d'une collaboration avec l'équipe de Namur (Éric A. Perpète et Denis Jacquemin) et celle de l'ENSCP (Carlo Adamo).

Le squaraine (Figure 5.3) fait partie de la catégorie des colorants qui absorbent dans le proche infra-rouge, typiquement autour de 650 nm. Ce colorant a un certain avantage, notamment sa facilité à être synthétisé ; par contre, il souffre de quelques désavantages : il est très sensible aux attaques chimiques et sous certaines conditions, il peut même s'agréger. Smith et collaborateurs ont montré en 2005 qu'il y avait un moyen de contourner ces deux limites en « encapsulant » le colorant dans un macrocycle [249–251]. Dans cette étude, la cage est le tétralactame (voir Figure 5.4). De tels complexes sont présents dans de nombreuses applications comme le FRET (*Fluorescence Resonance Energy Transfer*) et les technologies liées à l'imagerie.

L'objectif de cette étude était de modéliser rationnellement l'impact de la cage sur le colorant de squaraine ainsi que les effets géométriques et spectraux. Le chromophore a tout d'abord été considéré dans son état isolé, ensuite placé dans un continuum pour reproduire les effets de solvant [131, 252]. Les valeurs expérimentales de la longueur d'onde d'absorption maximale (λ_{\max}) sont disponibles pour le squaraine en solution (chloroforme) et complexé (solvaté par du chloroforme). Ensuite des calculs TD-DFT ont été réalisés sur le complexe entier en QM ainsi que des calculs QM:MM. Nous avons réussi ainsi à décomposer la variation du λ_{\max} lors du passage du squaraine isolé à sa forme complexée en trois effets physiques : géométrique, électrostatique et électronique.

5.2.1 Méthodologie utilisée

Pour un système de 140 atomes, il était nécessaire d'utiliser une méthode efficace et rapide pour évaluer les spectres d'absorption. Une étude préalable de Srinivas a montré que la TD-DFT était tout à fait adaptée à cette classe de colorants [253]. La DFT a été utilisée pour optimiser les géométries, la TD-DFT pour calculer les λ_{\max} . La fonctionnelle PBE0 développée par C. Adamo et collaborateurs a été employée pour ce système. Cette fonctionnelle contient 25% d'échange

Hartree-Fock (voir 1.3.1.C pour plus d'informations) et a souvent été utilisée pour évaluer les énergies de transition verticale [45, 254, 255]. La structure cristallographique a été prise comme point de départ pour l'optimisation [250], le groupe ponctuel étant C_i (cf. Figure 5.4). Sur un tel système, nous avons utilisé la base atomique 6-31G(d,p) pour définir la géométrie. Elle offre l'avantage de limiter le coût en temps de calculs. Les cinq premiers états excités ont été, quant à eux, déterminés à l'aide d'une fonction de base atomique plus diffuse, 6-31++G(d,p), permettant ainsi d'atteindre une description satisfaisante des énergies des états excités, et donc du λ_{\max} . Ces deux étapes ont été réalisées sur le squaraine isolé, sur le tétralactame isolé et sur le complexe ST en phase gazeuse. Expérimentalement, le spectre du complexe ST est obtenu dans le chloroforme. Nous n'avons pas réalisé de calculs dans le solvant dû à la taille du complexe ST. L'effet de complexation expérimentale en phase solvatée sera donc comparé théoriquement à un effet de complexation en phase gaz.

Afin d'observer les différents effets de la cage de tétralactame sur le spectre d'absorption, nous avons réalisé des calculs de type QM:MM et QM:MM:ERS sur l'ensemble de la structure optimisée en QM. Ce complexe est relativement facile à traiter par des méthodes hybrides QM-MM puisqu'il ne contient pas de liaisons covalentes entre le chromophore (partie QM) et la cage (partie MM). Concernant la partie MM, la cage de tétralactame ne fait pas partie des structures classiques présentes dans les champs de forces usuels. Deux choix s'offraient donc à nous : utiliser un champ de forces de type GAFF ou OPLS ou élaborer une méthode efficace et rapide. Nous avons choisi la dernière option. Les interactions électrostatiques sont souvent la contribution majeure parmi toutes les interactions. Nous négligeons ici les interactions de type vdW (dispersion) qui, généralement, ne modifient pas la distribution de charges ni de l'état fondamental ni de l'état excité. Un jeu de charges a donc été optimisé afin de reproduire correctement les interactions électrostatiques entre le chromophore et la cage. Nous avons choisi une procédure itérative pour déterminer les charges atomiques ponctuelles (*Atomic Point Charges*, APC) de la cage. La première étape consiste à faire un calcul de charges (en QM) sur le squaraine isolé ; les APC de Merz-Singh-Kollman ont été choisies pour reproduire le potentiel électrostatique [193, 194]. Ensuite, les APC de la cage sont, à leur tour, calculées en présence des APC du squaraine obtenues lors de la première étape. La fonction d'onde de la cage est donc polarisée par les APC du squaraine. Dans un troisième calcul, les APC du chromophore sont à nouveau recalculées en présence, cette fois des APC de la cage (obtenue lors du second calcul), et ainsi

de suite ... Les APC de la cage sont donc obtenues de manière auto-cohérente. Nous nommerons cette procédure *Self-consistent Atomic Point Charges* (SCAPC); les APC de la cage seront optimales lorsque les modifications de celles-ci entre deux cycles seront inférieures à $10^{-4}|e|$. De cette manière, nous sommes certains que le jeu de charges obtenu pour la cage reproduit bien les interactions électrostatiques entre les deux structures. Lors du calcul QM:MM_{SCAPC} pour évaluer les transitions verticales, la fonction d'onde du squaraine sera, par conséquent, polarisée par les SCAPC de la cage de tétralactame. Comme décrit ci-dessus (section 5.1.1), nous avons ensuite réalisé un calcul QM:MM pour déterminer le λ_{\max} en prenant en compte la réponse électronique du solvant à l'aide d'un continuum caractérisé par la constante diélectrique dynamique/optique (QM:MM:ERS).

5.2.2 Résultats et discussion

Nous allons comparer les paramètres géométriques obtenus de manière théorique pour le complexe de ST par rapport aux données expérimentales; et ce afin de vérifier que le couple PBE0/6-31G(d,p) n'induit pas de trop grandes divergences par rapport à la structure cristallographique. La géométrie joue, en effet, un rôle important sur le λ_{\max} [256]³.

L'évolution des longueurs de liaisons en passant du squaraine isolé au système complexé sera commentée par la suite. Nous présenterons les résultats des propriétés optiques obtenus en QM et en QM:MM:ERS. La valeur du λ_{\max} obtenue pour le complexe ST sera finalement décomposée en effets : géométriques, électrostatiques et électroniques.

5.2.2.A Analyse de la géométrie

Tout d'abord, la géométrie obtenue après optimisation en DFT est comparée à la structure obtenue expérimentalement par rayons-X [250]. La Table 5.1 reprend les paramètres géométriques du colorant, essentiellement les longueurs de liaison selon les notations reprises sur la Figure 5.3.

Nous pouvons y voir que les paramètres géométriques du complexe ST obtenus avec PBE0/6-

³Récemment, nous avons montré sur des systèmes contenant une liaison disulfure que la distance entre les deux atomes de soufre de liaison de type 2centre-3électrons (2c-3e) influence fortement le spectre d'absorption de petits peptides cystéiniques : une variation du λ_{\max} de 385 nm à 624 nm pour des distances entre atomes de soufre allant de 2.73 à 3.19 Å [256].

Distances	squaraine isolé (DFT)	complexe ST (DFT)	complexe ST (rayons-X) [250]
a	1.225	1.238	1.242
b, b'	1.470	1.459, 1.465	1.458, 1.468
c	1.408	1.408	1.409
d, d'	1.414	1.415, 1.413	1.412, 1.414
e, e'	1.376	1.375, 1.374	1.374, 1.374
f, f'	1.419	1.419, 1.419	1.417, 1.422
g	1.365	1.363	1.362
h, h'	1.446	1.447, 1.446	1.458, 1.452

Table 5.1: Comparaison entre les longueurs de liaisons en Å du colorant obtenues par diverses méthodes.

Voir Figure 5.3 pour la convention des notations utilisées.

31G(d,p) sont similaires à l'expérience. Par exemple, la distance calculée pour C=O et C-N(Me)₂ sont respectivement 1.238 et 1.363 Å, ce qui est en très bon accord avec les valeurs expérimentales de 1.242 Å et de 1.362 Å. En ce qui concerne le tétracycle central, les longueurs théoriques (1.459 Å et 1.465 Å) sont totalement comparables aux données cristallographiques (1.458 et 1.468 Å). Les paramètres de distance sont également très bons pour la liaison *c* située entre le tétracycle et les phenyls : 1.408 Å (théorie) contre 1.409 Å (exp.). Les mêmes conclusions sont observées au sein du phényl pour chacune des distances (*d*, *d'*, *e*, *e'*, *f* et *f'*).

Concernant la cage de tétralactame, la distance entre les deux anthracènes est légèrement surestimée par la théorie de 3% par rapport à la valeur expérimentale (7.0 Å contre 6.8 Å). Les principales différences entre expérience et théorie proviennent du macrocycle de tétralactame. La théorie prédit une conformation « chaise » alors que l'expérience montre une structure plus « flexible ». Néanmoins, les effets dissymétriques de la forme expérimentale sont sans doute dûs à des effets de packing dans le cristal [250]. Cette hypothèse a été confirmée par d'autres analyses de rayons-X réalisées sur des systèmes similaires (squaraine/macrocycle) qui sont dans la majorité en conformation « chaise » [249, 251]. Pour reproduire correctement l'effet du macrocycle sur le squaraine, il est important de décrire correctement les distances intramoléculaires de type liaisons hydrogènes entre le squaraine et le tétralactame. Par exemple, les distances N-H ... O=C mesurent 1.945 Å et 1.937 Å. Malheureusement ces valeurs ne peuvent être comparées à l'expérience car les atomes d'hydrogène ont été contraints durant l'étude de rayons-X. Une seule

donnée peut être directement comparée, la distance N...O relative à ces liaisons hydrogènes. Les résultats théoriques prédisent des longueurs de 2.937 Å et de 2.931 Å, celles-ci sont plus faibles et plus symétriques que les données expérimentales (3.005 Å et 2.942 Å). Cette divergence vient certainement de l'effet de packing de la structure cristallographique. Dans la référence [249], deux dérivés similaires ayant une conformation « chaise » possèdent des valeurs moyennes plus courtes pour ces ponts de 2.840 Å et 2.904 Å.

Après cette analyse des paramètres géométriques sur le complexe ST, nous pouvons affirmer que le couple PBE0/6-31G(d,p) donne une structure du complexe en très bon accord avec l'expérience et n'induit pas d'artefacts sur le spectre d'absorption.

Observons maintenant les variations géométriques entre le squaraine isolé et sa structure complexée. Celles-ci sont assez importantes au niveau du tétracycle. La longueur des liaisons C=O du colorant est allongée de 0.015 Å du fait de l'interaction avec le macrocycle. Les distances des liaisons du macrocycle sont aussi raccourcies de 0.011 et 0.005 Å par la présence de la cage, rendant ce tétracycle dissymétrique en présence du tétralactame. Les autres paramètres ne sont quasiment pas modifiés par la complexation du squaraine. Les modifications de la géométrie induites par la complexation sont donc principalement localisées sur la région possédant des liaisons hydrogènes entre le colorant et le macrocycle. Cette hypothèse a été confirmée par une optimisation du macrocycle isolé, les longueurs des liaisons N-H étaient aussi plus courtes que dans le complexe (1.009 contre 1.021 Å). La valeur de 1.009 Å est comparable à la distance N-H dans un système modèle comme le N-méthylacétamide (1.01 Å) [257].

5.2.2.B Le spectre d'absorption

Après l'analyse géométrique, des calculs TD-DFT ont été réalisés sur le chromophore isolé ou complexé. Les résultats se trouvent dans la Table 5.2. La longueur d'onde d'absorption maximale pour le squaraine isolé a été calculée à 506 nm avec une force d'oscillateur (f) égale à 1.5. Cette transition correspond à une excitation de type HOMO-LUMO caractéristique de cette classe de colorant. Cette valeur théorique est fortement sous-estimée par rapport à la valeur expérimentale (629 nm [250]), obtenue dans le chloroforme. Un calcul du colorant en solution réduit l'erreur par rapport à la valeur expérimentale avec un λ_{\max} = 539 nm. Malgré cela, l'écart reste très important (90 nm ou 0.33 eV). Or, la TD-DFT reproduit en général très bien les valeurs d'absorption avec au maximum des erreurs de 20 nm (0.05 eV)! Dans la suite de ce travail, cette déviation ne

portera pas à conséquence puisque l'objectif de cette étude est de reproduire la variation du λ_{\max} entre le composé isolé et complexé afin de comparer l'effet de complexation de +20 nm dans le chloroforme.

Système	Méthode	$\lambda_{\max}^{\text{théo}}$	$\lambda_{\max}^{\text{exp}}$	$\Delta\lambda_{\max}^{\text{théo}}$
Squarain isolé	QM	506	–	
Squaraine dans CHCl_3	QM	539	629	
Tétralactame isolé	QM	394	–	
	QM	521	649*	+15 (+ 20)
Complexe ST	QM:MM _{SCAPC} :ERS	514		+8
	QM:MM _{SCAPC}	486		-20

Table 5.2: Valeurs des longueurs d'onde d'absorption maximale (λ_{\max}) en nm du squaraine isolé, en solution et complexé. L'astérisque indique que la valeur expérimentales du complexe a été obtenue dans le chloroforme. La variation du λ_{\max} ($\Delta\lambda_{\max}$) est obtenue en prenant comme référence le squaraine isolé. La valeur entre parenthèse est le déplacement dû à la complexation obtenue expérimentalement dans le chloroforme.

Le spectre d'absorption de la cage isolée possède une longueur d'onde d'absorption maximale de 394 nm ($f=0.3$). Cette valeur est principalement due à la partie conjuguée du tétralactame, c'est-à-dire l'anthracène. D'ailleurs, ce λ_{\max} est très proche de la valeur théorique pour l'anthracène isolé (373 nm, $f=0.06$) et dans le chloroforme (378 nm, $f=0.09$). Expérimentalement, un petit pic d'absorption caractéristique de la signature du spectre vibrationnel des acènes apparaît dans cette région.

Passons maintenant au spectre du complexe ST, la transition permise en TD-DFT se situe à 521 nm. Les orbitales impliquées (HOMO-2 et LUMO) dans l'excitation sont complètement localisées sur le colorant et non sur la cage. La figure 5.5 montre la densité électronique impliquée lors de la transition. Il est intéressant de noter que la HOMO-2 possède une plus grande contribution sur l'atome d'oxygène que sur la LUMO. Cela est tout à fait cohérent avec la stabilisation relative des orbitales frontières lors de la complexation : -0.39 eV pour l'orbitale occupée et -0.27 eV pour la LUMO. Ces valeurs impliquent que l'écart entre orbitales occupées et virtuelles est plus grand pour le complexe ST que pour le chromophore isolé. En DFT, cela induit donc un

effet hypsochromique incorrect ; par contre dans le formalisme TD-DFT qui prend en compte la polarisation électronique, l'effet bathochromique expérimental (+20 nm) est correctement reproduit (+15 nm). Cette effet de réorganisation électronique est donc crucial dans ce système. Le λ_{\max} expérimental du complexe est de 649 nm.

Nous nous sommes ensuite intéressés à l'origine de ce déplacement de +20 nm en décomposant le λ_{\max} en plusieurs contributions. En réalisant un calcul sur le chromophore isolé mais avec la géométrie qu'il possède lorsqu'il est complexé, nous obtenons un λ_{\max} de 514, ce qui montre un déplacement bathochromique de +8 nm par rapport au chromophore optimisé en phase gazeuse (506 nm) et un déplacement hypsochromique de -7 nm par rapport à la structure complexée. Nous pouvons donc affirmer que la moitié de la variation du λ_{\max} entre la forme isolée et complexée du squaraine a pour origine la déformation géométrique.

Le calcul QM:MM_{SCAPC}:ERS, incluant notre nouvelle méthode pour prendre en compte la polarisation électronique de l'environnement, nous indique que nous obtenons un résultat tout à fait comparable à celui obtenu en TD-DFT sur le complexe ST (514 nm contre 521 nm). Nous pouvons ensuite décomposer cette valeur de 514 nm. En utilisant simplement le jeu de charges optimisé (QM:MM_{SCAPC}) pour modéliser la cage de tétralactame, nous obtenons une longueur d'onde de la transition de 486 nm ! Cette valeur entraîne un déplacement hypsochromique inattendu par rapport à la valeur du squaraine isolé. Cela signifie que la contribution électrostatique de la cage n'est pas responsable du déplacement bathochromique observé ⁴. Lorsque le modèle inclut la polarisation électronique de la cage (QM:MM_{SCAPC}:ERS), un déplacement bathochromique de +8 nm par rapport au chromophore isolé est obtenu ; l'effet bathochromique est donc bien reproduit. Ce cas montre clairement qu'il ne suffit pas de décrire correctement les effets électrostatiques de l'environnement, mais qu'il faut également prendre en compte la contribution électronique. Dans ce cas, un déplacement bathochromique entre les λ_{\max} QM:MM_{SCAPC} et QM:MM_{SCAPC}:ERS de +28 nm est observé. La prise en compte de l'ERS permet de contrebalancer totalement l'effet électrostatique qui induit un déplacement contraire (hypsochromique) au déplacement souhaité (bathochromique). La polarisation électrostatique de l'état fondamental, habituellement utilisée en QM-MM, n'est pas suffisante pour reproduire le spectre d'absorption de ce complexe.

⁴Un calcul du chromophore dans sa géométrie optimisée dans le complexe a été effectué en présence d'un continuum caractérisé par $\epsilon_{\infty}=1.776$. Nous obtenons une valeur de 541 nm ($f=1.73$).

5.2.3 Conclusion

La DFT nous a fourni une géométrie du complexe en bon accord avec la structure expérimentale obtenue par rayons-X. La TD-DFT reproduit correctement l'effet de la complexation du squaraine par une cage de tétralactame (+15 nm contre expérimentalement +20 nm). La moitié de ce déplacement provient des variations de la géométrie du squaraine suite à la complexation par le tétralactame. À l'aide de la méthode hybride QM:MM, un jeu de charges déterminé spécifiquement pour modéliser l'environnement ne suffit pas ; il est crucial de prendre en compte la polarisation électronique de la cage de tétralactame. L'approche QM:MM:ERS que nous proposons a donc pu être validée, reproduisant bien l'effet de la complexation bathochromique (+8 nm). Alors même qu'un calcul tout quantique est faisable, l'apport d'une analyse QM:MM est démontré puisqu'une information importante sur la contribution relative des effets géométriques et électrostatiques est clairement obtenue. Les propriétés optiques issue de notre méthodologie se sont montrées similaires au calcul tout QM avec un écart de seulement 7 nm (514 nm contre 521 nm).

5.3 Les protéines fluorescentes

UNE grande partie de cette thèse a été consacrée aux spectres d'absorption des protéines fluorescentes (FP). Les protéines fluorescentes comprennent trois grandes classes : les rhodopsines, les protéines photoactives jaunes (PYP, *Photoactive Yellow Protein*) et les protéines fluorescentes vertes (GFP, *Green Fluorescent Protein*). Nous nous sommes particulièrement intéressés aux protéines de la famille des GFP. Comme son nom l'indique, la GFP est une protéine ayant la propriété d'émettre une fluorescence dans le visible de couleur verte. C'est Shimomura en 1962 qui l'a extraite d'une méduse de la famille des *Aequorea victoria* [258]. Une caractéristique des GFP est qu'elles acquièrent leur propriété de fluorescence d'elles-mêmes. Aucune enzyme ne leur sont nécessaire, seul l'oxygène est essentiel à la formation de son site chromo-ou fluorophore. L'introduction du gène de la GFP dans une cellule permet d'observer ensuite la fluorescence verte de la protéine sous un éclairage bleu. Dans les années quatre-vingt-dix, la GFP a pu être clonée et exprimée dans d'autres organismes que les méduses *Aequorea victoria* [259, 260]. Ces protéines ne sont pas toxiques pour la cellule et deviennent très rapidement un outil de choix dans l'étude des processus biologiques des cellules ou des organismes vivants.

Les images de fluorescence obtenues ouvrent diverses possibilités pour l'usage des GFP en tant que marqueurs ou senseurs biologiques (concentration en ions, détection du pH, ...) [260–262].

De nombreuses mutations ont été effectuées, notamment par le groupe de Roger Tsien, pour améliorer le temps d'émission de la fluorescence, son éclat, sa stabilité mais aussi l'élargissement du spectre d'absorption/émission des GFP [263]. La large gamme de colorations actuellement offerte par tous les mutants de la GFP permet d'observer des processus de protéines différentes simultanément. Une des applications des GFP est le FRET (*Förster Resonance Energy Transfer*) où le contrôle des protéines fluorescentes acceptrices ou donneuses est essentiel lorsque deux espèces sont proches l'une de l'autre [264, 265]. L'importance de ces développements explique que dans le monde biologique les protéines fluorescentes se sont imposées. Certaines protéines fortement dépendantes de leur environnement peuvent changer de couleurs durant le processus observé, voire même clignoter ! L'importance de la découverte de GFP et de ses avancements a été saluée par un prix Nobel en octobre 2008 remis à Osamu Shimomura, Martin Chalfie et Roger Tsien [266–269] ⁵.

Une particularité des GFP et de ses nombreux mutants est que le chromophore fait partie intrinsèquement de la chaîne d'acides aminés. En 1979, Shimomura suggéra la structure du chromophore de la GFP, simplement par analogie avec le spectre de la luciférine qu'il avait étudié auparavant [270]. Cette structure fût confirmée en 1993 par Cody et collaborateurs [271]. Le chromophore de la GFP est composé de trois acides aminés (résidus 65 à 67) formés suite à un mécanisme en trois étapes suggéré par Tsien pour le chromophore de la GFP *sauvage* (Ser65-Tyr66-Gly67) : cyclisation, déshydratation et oxydation [265]. Ce mécanisme est explicité sur la Figure 5.8. L'étape préliminaire de repliement des trois acides aminés est toujours sujet à débat [272–278]. Toutes les variantes de la GFP sont donc modifiées par i) des mutations au sein du chromophore ou ii) des mutations dans son environnement proche ou lointain. Le résidu 66 est toujours un acide aminé qui possède un cycle conjugué de type tyrosine (Tyr), histidine (His), phénylalanine (Phe), tryptophane (Trp). Une autre mutation régulière est celle de la sérine (Ser) en position 65 par une thréonine (Thr). La mutation du résidu 65 ainsi que les mutations de résidus très proches du chromophore aident essentiellement à la formation du chromophore.

⁵Le site internet de M. Zimmer <http://www.conncoll.edu/ccacad/zimmer/GFP-ww/GFP-1.htm>.

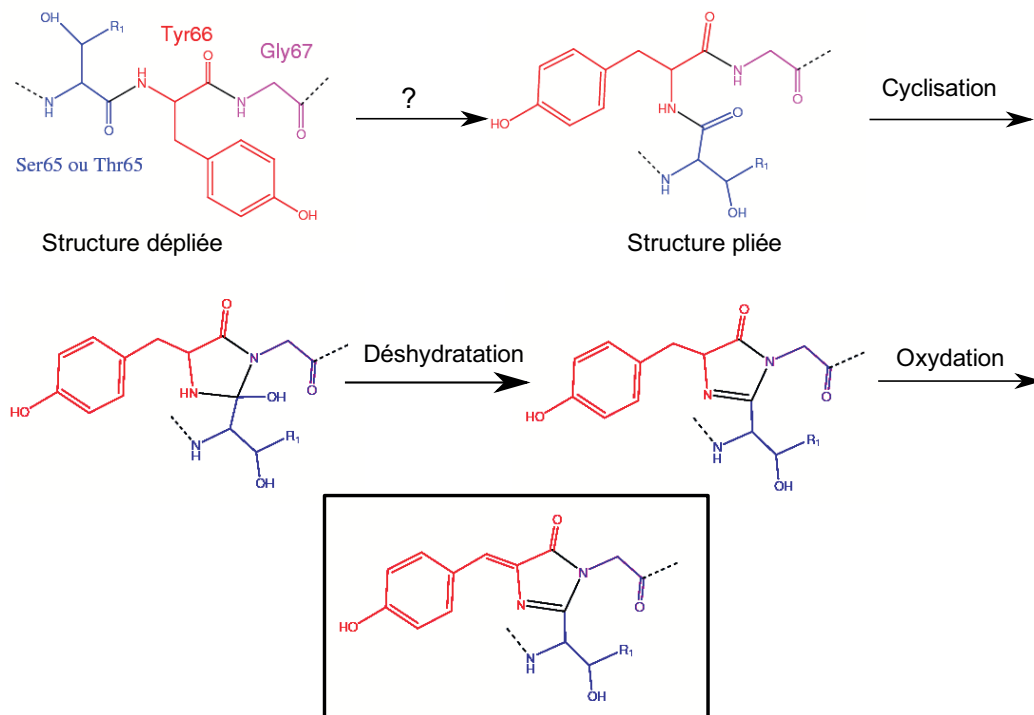


Figure 5.8: Voici le mécanisme de réaction de la formation du chromophore de la GFP proposé par Tsien [265,272]. Selon la nature du résidu 65, sérine (Ser) ou thréonine (Thr), R_1 sera un hydrogène ou un méthyle. La GFP *sauvage* possède le chromophore Ser65-Tyr66-Gly67.

Le chromophore est lié de manière covalente à une hélice α qui est, elle-même située au centre d'un tonneau de feuillets β anti-parallèles. Cette structure est donc très rigide. Une seule irrégularité est présente dans ce tonneau de feuillets β . Elle se situe entre les feuillets 7 et 8 [279]. L'écart entre ces deux feuillets est plus grand ce qui empêche la formation de liaisons hydrogènes entre les deux feuillets et permet le passage de molécules d'eau [280].

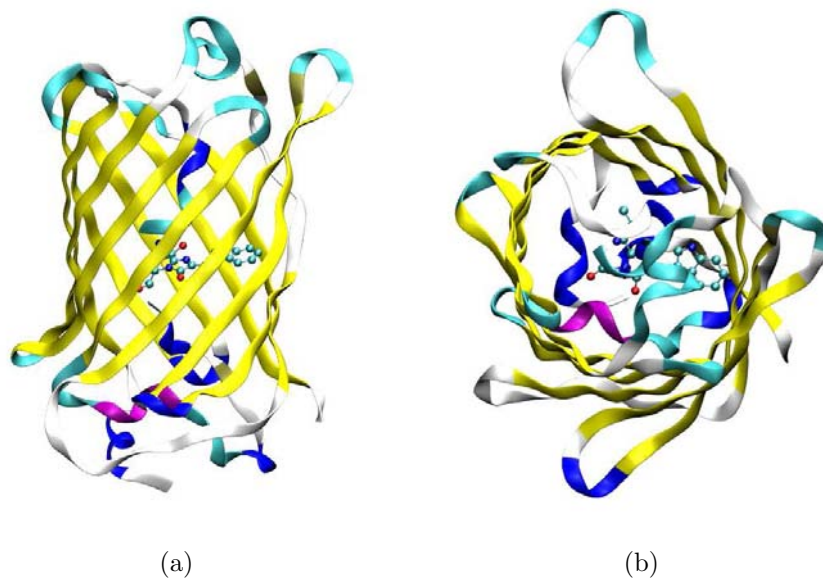


Figure 5.9: Conformation majeure de la ECFP. Le chromophore est en représentation boule et bâton, la structure secondaire de type feuillets- β est en jaune, les motifs de type hélice- α en bleu et les parties sans structure en blanc.

Les protéines de la famille des GFP sont dites photoactives. Ce sont des interrupteurs irréversibles, capables d'émettre de la lumière une fois sur un temps court mais de manière très intense. Ces structures passent d'un état non-coloré à un état coloré une seule fois. Récemment, une nouvelle classe de protéines fluorescentes a été découverte. Celles-ci sont des interrupteurs réversibles : elles peuvent passer un certain nombre de fois d'un état fluorescent à un état non-fluorescent ou inversement. Les plus connues sont la Dronpa [281, 282], l'asFP595 [283], l'EosFP [284] et l'IrisFP [285]. Les mécanismes qui rendent possible la réversibilité sont des échanges de protons et/ou des isomérisations.

Après cette présentation générale sur les GFP, nous allons décrire les propriétés spécifiques liées à la protéine fluorescente qui a fait l'objet de cette étude : la ECFP (*Enhanced Cyan Fluorescent Protein*).

5.3.1 La protéine ECFP

La CFP (*Cyan Fluorescent Protein*) est souvent nommée ECFP (*Enhanced Cyan Fluorescent Protein*) grâce à son rendement quantique de fluorescence élevé [286]. Par rapport à la GFP *sau-*

vage (wt-GFP), elle possède une mutation au sein de son chromophore Tyr66Trp et quatre autres mutations au niveau de l'environnement (Phe64Leu, Asn146Ile, Met153Thr, Val163Ala)⁶ [286]. Le chromophore de la ECFP est le Thr65-Trp66-Gly67 (Figure 5.10). La structure cristallographique de cette protéine a été définie en 2003 par Huyen Bae et collaborateurs [287]. Cette équipe souhaitait trouver une FP qui ne s'agrège pas et qui avait des propriétés de brillance plus importantes que les autres. Ils ont ainsi mis au point trois protéines qui se différencient par « simple » addition d'un groupement fonctionnel sur le Trp66 du chromophore. Nous étudierons ces mutations dans la section 5.3.5.

En 2002, grâce à une étude expérimentale de RMN du ¹⁹F, le groupe de Seifert a détecté la présence de deux conformations [288]. En obtenant la résolution de la structure des rayons X, Huyen Bae et collaborateurs ont pu quantifier le rapport entre ces deux conformations : 60% pour la conformation dite majeure et 40% pour la mineure. La Figure ci-dessous montre les principales différences entre ces deux conformations. Il s'agit de deux résidus (Tyr145 et His148) qui sont positionnés à l'intérieur ou à l'extérieur du tonneau de feuillets β selon la conformation. Ces deux acides aminés sont situés dans la crevasse formée par les feuillets β 7 et 8. Dans la conformation majeure, la Tyr145 et l'His148 se trouvent respectivement (en bleu sur la Figure 5.11) à l'intérieure et à l'extérieure du tonneau de feuillets β . La zone extérieure au tonneau de feuillet β est la partie exposée au solvant. Dans l'autre conformation (mineure), c'est l'inverse (en rouge sur la Figure 5.11). Le spectre d'absorption de la protéine ECFP présente un pic relativement large, présenté souvent comme un double pic de 434 et 452 nm [287]. Celui-ci est peu dépendant du pH. La première hypothèse proposée pour expliquer le double pic a été la présence de deux conformations chez la ECFP. Cela a été réfuté plusieurs fois notamment par la mutation des deux résidus [288] et par des études de dynamique moléculaire [280]. Les effets de la mutation du chromophore dans la conformation majeure seront étudiés.

La ECFP est très souvent utilisée pour le FRET comme donneur, l'accepteur étant la YFP (*Yellow Fluorescent Protein*). Cependant, en plus de la présence d'un double pic, la ECFP possède d'autres désavantages comme un rendement quantique bas (0.4), un coefficient d'extinction molaire faible et un petit temps de fluorescence (2.5 ns à pH basique et 1.5 ns à pH acide). En

⁶La notation conventionnelle utilisée pour les mutations est la suivante : les trois premières lettres désignent le nom de l'acide aminé muté, le nombre sa position dans la séquence de la protéine et les trois dernières lettres indiquent le nom de l'acide aminé mutant. Dans cette thèse, le code a trois lettres décrivant les acides aminés a été préféré pour plus de clarté. La structure de ces acides aminés se trouve en Annexe B.

2004, Rizzo et al. ont souhaité améliorer les propriétés de la ECFP pour le FRET. En partant de l'hypothèse de départ qu'il y a un transfert d'énergie entre les deux conformations de la ECFP, donc une perte du signal lors du FRET, la mutation de Tyr45 et/ou l'His148 par un résidu hydrophile ne produirait plus qu'une seule conformation, la conformation majeure. Ils ont donc muté trois résidus (Ser72Ala, Tyr145Ala et His148Asp) et nommé cette protéine la Cérulan (code pdb : 2Q57). Elle montre effectivement un rendement quantique 1.7 fois plus élevé et un coefficient d'extinction molaire 1.5 fois plus grand (donc plus brillante) cependant le double pic est toujours présent [289]. En 2007, la structure cristallographique de la Cérulan a été résolue et a surpris sur plusieurs points : sa structure est similaire à la conformation mineure (His148Asp vers l'intérieur) et le chromophore se trouvait en configuration *cis* [290].

Quelques études théoriques menées par des groupes français ont porté sur les FP. Le groupe d'Isabelle Demachy a publié un papier en 2005 qui contenait une étude de dynamique moléculaire et l'évaluation du λ_{\max} à l'aide de la mécanique quantique pour toute une série de *snapshot* du chromophore extrait de la dynamique. Toujours dans ce groupe, G. Vallverdu a réalisé une étude de dynamique sur le chromophore de la GFP dans son état excité [291] et ensuite a étudié le sujet délicat des effets de pH de la Céruléan lors d'une collaboration expérimentale [292]. Parallèlement, M. Lelimosin a mené une étude de dynamique moléculaire sur la Céruléan [293] en collaboration étroite avec le groupe d'expérimentateur de Grenoble qui a découvert IrisFP.

L'enjeu de cette étude sur la ECFP est de plusieurs ordres :

1. Tester et utiliser les récents développements concernant la coupure au niveau de la liaison peptidique (voir section 5.3.3.A).
2. Réaliser des calculs QM/MM sur ces protéines. Les études de spectres d'absorption réalisées sur ces protéines ont, en effet, souvent été limitées au chromophore (voir section 5.3.3.A).
3. Utiliser le modèle simple mis au point pour de prendre les effets de la réponse électronique de l'environnement de manière isotrope, le QM/MM:ERS, (voir section 5.3.4) et ainsi observer l'importance de la réponse électronique.
4. Décomposer la longueur d'onde d'absorption maximale (λ_{\max}) en différentes contributions physiques afin de déterminer l'origine physique de la valeur du λ_{\max} obtenu.
5. Observer l'influence de la mutation du chromophore de la ECFP sur le spectre d'absorption UV/visible (voir section 5.3.5).

La section suivante reprend tous les détails méthodologiques employés lors des calculs de type QM, QM/MM et dynamique moléculaire pour réaliser ces études.

5.3.2 Méthode utilisée pour les FP

Tout d'abord, le chromophore isolé de la ECFP a été considéré. Celui-ci a été saturé comme indiqué sur la Figure 5.12 par les groupements amine (-NH₂) et aldéhyde (-C(=O)H). Le niveau de théorie QM choisi pour optimiser la géométrie est B3LYP/6-311G(d,p); alors que les huit premiers états excités ont été calculés à l'aide de la TD-DFT, et plus précisément au niveau de théorie TD-B3LYP/6-311+G(2d,p). B3LYP est actuellement la fonctionnelle la plus utilisée parce qu'elle décrit correctement la géométrie des structures. Plusieurs études menées notamment par Olivucci sur ces protéines utilisent des méthodes hautement corrélées et plus chères en temps de calcul (CASPT2//CASSCF ou EOM-CC) [87,294–296]. Celles-ci utilisent par conséquent des fonctions de bases atomiques plus restreintes (souvent double- ζ). Le formalisme IEF-PCM (*Integrated Equation Formalism-PCM*) [141] a aussi été employé lorsque le chromophore est placé en solution ainsi que pour les calculs QM/MM:ERS afin de modéliser l'ERS *via* un continuum caractérisé par une constante diélectrique dynamique ($\epsilon = \epsilon_{\infty} = 1.779$).

Le chromophore étant lié de manière covalente à son environnement, nous avons testé l'effet de la coupure sur le chromophore modèle (Figure 5.12). Deux positions distinctes pour la coupure ont été testées à l'aide des approches LSCF/+1, LSCF/+3, LSCF/+3,+5 et HLA. Nous renvoyons le lecteur à la section 4.3.2 pour tous les détails concernant les différentes approches. Pour localiser les SLBO, le critère de WP a été employé sur des systèmes modèles contenant la liaison d'intérêt [222, 223]. Nous verrons dans la première partie des résultats quelles sont les deux positions de la coupure testées ainsi que la méthode que nous avons sélectionnée pour les calculs sur la protéine entière.

Concernant le traitement de la protéine fluorescente entière, les méthodes hybrides QM/MM sont parfaitement adaptées. Le chromophore absorbant de la lumière sera décrit au niveau quantique tandis que son environnement à l'aide de la MM. Le niveau de théorie QM est celui cité ci-dessus, l'environnement sera quant à lui décrit par le champ de forces classiques Amber99 [164]. Les effets de vdW pour tout le système ont été considérés étant donné que chaque atome décrit en QM a été assigné d'un type d'atome du champ de forces Amber99. La structure de départ est la structure cristallographique des deux conformations de la ECFP (code pdb 1OXD—majeure

et 1OXE–mineure). Les atomes d’hydrogène ont été ajoutés avec le module `tleap` du programme Amber9 [297]. La position de ces atomes d’hydrogène a ensuite été minimisée par le programme `minimize` de Tinker4.2 [298]. Lors des calculs QM/MM, le chromophore est traité en QM (Thr65-Trp66-Gly67), les deux résidus suivants (Leu64 et Val68) sont décrits en MM et sont relaxés lors de l’optimisation de la géométrie, le reste de la protéine étant gelé. Nous avons examiné deux approches. Dans la première, aucune molécule d’eau n’a été considérée lors des calculs QM/MM ; dans une seconde, les molécules d’eau dans un rayon de 4 Å autour du chromophore ont été prises en compte. Celles-ci ont été relaxées comme la Leu64 et Val68 lors de l’optimisation. Cette étude sur la ECFP a été menée sur les deux conformations (majeure et mineure). Les effets de la mutation (section 5.3.5) n’ont été considérés que sur la conformation majeure parce que cette conformation est celle de la plupart des protéines fluorescentes.

Les simulations de dynamique moléculaire ont été réalisées avec le logiciel NAMD [299] dans les conditions canoniques (NVT) durant 4 ns avec un pas de temps de 0.2 fs. Au préalable, nous avons minimisé les structures cristallographiques et relaxé les molécules d’eau durant 200 ps. La taille de la boîte d’eau est définie de telle sorte que la distance entre la protéine et les murs soit de 8 Å. La charge globale du système étant négative, nous avons neutralisé le système par six contre-ions Na^+ . Nous avons considéré que les systèmes étudiés se trouvaient à un pH neutre. Dès lors, les résidus aspartate et glutamate ont été déprotonés tandis que les acides aminés lysine et arginine sont protonés [300]. Les paramètres du champ de forces Amber99 ont été choisis pour l’ensemble de la protéine. Cependant, quelques paramètres ont dû être ajoutés afin de décrire correctement le chromophore. Nous avons utilisé les paramètres de la glycine et de la thréonine pour les résidus Gly65 et Thr67 du chromophore. La partie indole du Trp66 a elle aussi des paramètres similaires à ceux présents dans le tryptophane. Les paramètres spécifiques de l’imidazolinone et du pont éthylénique sont tirés de la publication de N. Reuter [301]. Les paramètres manquants ont été ajoutés en utilisant le champ de forces GAFF [165]. Le jeu de charges du chromophore a été calculé pour chaque chromophore considéré à l’aide du logiciel RESP [197]. Pour les dynamiques moléculaires, le RMSD (*Root Mean Square Deviation*) sur la protéine (sans les atomes d’hydrogène) par rapport à la structure cristallographique est inférieur à 1.5 Å. Cela indique que la structure est très rigide, comme attendu. La structure moyenne (cristallographique) est, par conséquent, un excellent point de départ pour l’optimisation de la géométrie en QM/MM.

Enfin, similairement à l'étude du complexe ST, nous avons décomposé longueur d'onde d'absorption maximale en trois contributions physiques. Pour cette étude sur les FP, cette décomposition sera systématique, la section suivante l'explique.

5.3.2.A Décomposition des propriétés optiques

La décomposition de la longueur d'onde d'absorption donne de précieuses informations sur les effets chimiques et physiques cachés derrière le spectre d'absorption des chromophores (cf. section 5.2.2.B). Nous allons ici montrer comment cette décomposition est réalisée rigoureusement. Le λ_{\max} est scindé en quatre contributions physiques :

$$\lambda_{\max} = \lambda_{\max}(\text{QM}) + \Delta\lambda_{\max}^{\text{géom}} + \Delta\lambda_{\max}^{\text{élec}} + \Delta\lambda_{\max}^{\text{ERS}}$$

Ces quatre termes sont décrits ci-dessous ⁷ :

- La dernière contribution correspond à la variation du $\Delta\lambda_{\max}^{\text{ERS}}$ dû à la réponse électronique de l'environnement ($\lambda_{\max}^{\text{ERS}}$). Celle-ci est calculée en faisant la différence entre le calcul LSCF/MM:ERS et le calcul LSCF/MM, et ce, dans la géométrie optimisée en LSCF/MM

$$\Delta\lambda_{\max}^{\text{ERS}} = \Delta\lambda_{\max}(\text{LSCF/MM : ERS}) - \lambda_{\max}(\text{LSCF/MM}) \quad (5.8)$$

- La variation du λ_{\max} dû à la présence des charges ponctuelles de l'environnement ($\Delta\lambda_{\max}^{\text{élec}}$) est calculée en soustrayant la valeur du λ_{\max} obtenue en LSCF/MM et celle obtenue lorsque les charges ponctuelles de la partie MM sont posées à zéro (LSCF/MM noEE⁸) ; toujours dans la géométrie LSCF/MM.

$$\Delta\lambda_{\max}^{\text{élec}} = \Delta\lambda_{\max}(\text{LSCF/MM}) - \lambda_{\max}(\text{LSCF/MM noEE})$$

- La modification de la longueur d'onde d'absorption maximale due aux déformations géométriques du chromophore ($\Delta\lambda_{\max}^{\text{géom}}$) est calculée en faisant la différence entre le résultat LSCF/MM noEE et celui obtenu sur le chromophore en tout QM.

$$\Delta\lambda_{\max}^{\text{géom}} = \Delta\lambda_{\max}(\text{LSCF/MM noEE}) - \lambda_{\max}(\text{QM})$$

- Le premier terme de l'Eq. (5.8) est simplement la longueur d'onde d'absorption maximale calculée en tout QM sur le chromophore isolé et saturé par des atomes d'hydrogène (voir Figure 5.10)

⁷La séparation en terme géométrique et électrostatique provenant d'un calcul LSCF/MM, la géométrie obtenue est optimisée en présence de l'*Electrostatic Embedding*

⁸noEE signifie que l'*Electrostatic Embedding* ne sera pas pris en compte.

5.3.3 Le chromophore de la ECFP isolé

Premièrement, nous avons réalisé une série de calculs quantiques sur le modèle du chromophore de la ECFP. Un grand nombre de publications sur les FP sont menées uniquement sur le chromophore. Ce chromophore étant soit en phase gaz ou mieux en solution [295, 296, 302]. Comme signalé précédemment (section 5.1), plusieurs études prétendent « reproduire » ou s'approcher de la tendance de l'environnement sur les propriétés optiques par une solution aqueuse ou par un continuum possédant une constante diélectrique entre 4 et 20 reproduisant la constante diélectrique au sein d'une protéine.

Les modèles de continuum, très souvent utilisés, sont totalement isotropes et ne reproduisent pas les interactions spécifiques entre soluté et solvant. Or, ces chromophores sont entourés par des acides aminés ou des molécules d'eau formant un réseau de liaisons hydrogènes complexes autour de lui. La Table 5.3 reprend les valeurs de λ_{\max} du chromophore de la ECFP en phase gazeuse et placé dans des continua de différentes constantes diélectriques. Par rapport au λ_{\max} obtenu en phase gaz, il y a systématiquement un déplacement vers les grandes longueurs d'onde (solvatochromisme positif) lorsque le chromophore est placé dans un continuum caractérisé par des constantes diélectriques de plus en plus grandes. Plus le chromophore se trouve en milieu polaire, plus la longueur d'onde d'absorption augmente. Le milieu stabilise les orbitales impliquées dans la transition électronique. Cette excitation électronique est de caractère $\pi^* \leftarrow \pi$ (nous verrons ces orbitales par après). Parmi ces différents milieux isotropes, nous remarquons qu'aucune valeur de λ_{\max} n'approche la valeur expérimentale du chromophore dans la protéine de 434-452 nm, bien qu'ils améliorent quelque peu le λ_{\max} en phase gazeuse (391 nm) du chromophore isolé. Le modèle du chromophore isolé ainsi que le remplacement du milieu protéique complexe par un milieu isotrope ne suffisent pas pour approcher la longueur d'onde d'absorption expérimentale du chromophore. Il est donc nécessaire de prendre en compte l'environnement de manière correcte, à l'aide de méthode hybride QM/MM. Les chromophores de la famille des GFP sont liés de manière covalente au reste de la protéine. La première étape, avant de traiter l'entièreté de la protéine en QM/MM, est de positionner correctement les coupures entre la partie QM et la partie MM de façon à ce que les coupures n'entraînent pas d'artefacts (géométriques ou sur les propriétés étudiées). La section suivante est donc dédiée aux effets de la position des coupures. Deux sites de coupure possible ont été testés sur plusieurs chromophores de la famille des GFP afin d'être certain que les coupures ne tronquaient pas les propriétés géométriques et optiques

Modèle	ϵ	λ_{\max}
Gaz	1.0	391 (0.6)
« Protéine »	4.0	404 (0.7)
	20.0	408 (0.7)
Eau	78.4	409 (0.7)
Expérience [287]		434-452

Table 5.3: Longueur d'onde d'absorption (λ_{\max}) du chromophore de la ECFP (voir Figure 5.10) en fonction du milieu considéré. Les forces d'oscillateur de la transition électroniques observées sont indiquées entre parenthèse. Le milieu est modélisé par un continuum de diverses constantes diélectriques.

du chromophore.

5.3.3.A L'effet de la coupure en LSCF/MM

Contrairement au complexe ST, le chromophore (QM) est lié de manière covalente au reste de la protéine. Il est donc nécessaire de tester la position de la coupure QM/MM ainsi que l'approche utilisée. Nous nous sommes centrés essentiellement sur le chromophore de la ECFP. Une manière d'observer si la position de la coupure n'induit pas de trop importants artefacts est d'étudier les paramètres géométriques, ainsi que la propriété étudiée, la longueur d'onde d'absorption maximale. Les propriétés de référence seront celles calculées en QM sur le chromophore modèle en phase gazeuse. Ce chromophore a été saturé par des atomes d'hydrogène (voir Figure 5.14). Les premières coupures testées se trouvent directement sur les deux liaisons voisines à l'imidazolinone (Figure 5.14 (a)). Cette coupure a été testée avec le LSCF/+1 [5] et son potentiel de liaison frontière entre les atomes N(sp3)-C(sp3) et C(sp2)-C(sp3) ainsi qu'avec le LSCF/+3 [7, 10] et avec le HLA (*Hydrogen Link Atom*). Les SLBO positionnées sur ces deux liaisons ont été obtenues sur le système modèle de l'imidazolinone incluant les deux liaisons d'intérêts (cf. Figure 5.15 (a)). Ces deux coupures impliquent une reparamétrisation des charges de la partie MM (en bleu sur la Figure 5.14 (a)). Ces charges ont été obtenues à l'aide du logiciel RESP où trois contraintes ont été imposées simultanément : a) la charge globale du système doit être nulle, b) la charge de la partie MM Glycine67 est nulle et c) la charge de la partie MM thréonine65 est nulle. Le

traitement en HLA implique en plus, une redistribution de la charge du premier atome MM sur ses voisins directs (redistribution de type Z1) ⁹.

Les paramètres géométriques et optiques sont repris dans la Table 5.4. Nous comparons ces résultats aux calculs obtenus en QM.

D'un point de vue géométrique, le LSCF/+1 raccourcit la liaison qui a été coupée, principalement celle coupant les deux atomes de même nature. Le LSCF/+3 et le HLA donnent des résultats relativement similaires concernant les paramètres géométriques. Concernant les propriétés optiques, le LSCF/+1 indique le plus grand écart par rapport à la valeur obtenue en QM (-10 nm). Le LSCF/+3 fournit une excellente valeur de 388 nm, avec un déplacement hypsochromique de seulement 3 nm par rapport 391 nm obtenu QM. Le λ_{\max} obtenu en HLA est par contre plus élevé de 4 nm (395 nm). Cette première coupure proche de l'imidazolinone présente des résultats en meilleur accord avec le calcul QM pour les approches LSCF/+3 et HLA. Cependant, la présence d'une SLBO aussi proche du site conjugué peut entraîner certains artefacts.

Nous avons positionné, ensuite, la coupure au niveau des liaisons peptidiques (Figure 5.14 (b)), seuls six atomes se trouvent dans la partie MM. La SLBO a été déterminée sur le N-méthylacétamide (Figure 5.15 (b)). Nous avons utilisé les récents développements de P.-F. Loos et X. Assfeld, le LSCF/+3,+5 [9]. Cette méthode a le grand avantage de ne pas devoir redistribuer les charges dans les protéines puisque les coupures peptidiques sont des coupures naturelles des acides aminés. Par rapport aux bons résultats du traitement en LSCF/+3 ou HLA, la géométrie du chromophore est exactement la même ainsi que la valeur de la longueur d'absorption.

L'excitation électronique impliquée dans cette transition est de type $\pi^* \leftarrow \pi$. La Figure 5.16 montre la densité électronique. Nous pouvons observer que les énergies sont totalement similaires entre le calcul QM complet et le calcul LSCF/+3,+5.

Nous avons réalisé les mêmes tests de validité de la coupure sur d'autres chromophores issus de la famille des FP. Ces chromophores sont repris dans la Figure 5.17. La propriété étudiée est la longueur d'onde d'absorption. Ces données sont reprises dans la Table 5.5. La longueur d'onde d'absorption est fortement affectée lorsque la méthode du LSCF/+1 est employée avec des écarts de 20 à 7 nm. La référence est toujours le calcul obtenu en QM. Nous pouvons y voir que, dans quasiment tous les cas le LSCF/+3 donne des valeurs de λ_{\max} relativement éloignées de la valeur de référence obtenue en QM alors que les quantités obtenues par le LSCF/+3,+5 sont en total

⁹Nous avons testé deux redistributions Z1 et Z2.

Modèles	Distances				Angles dièdres			λ_{\max}
	$C_{\text{Im}}-C$	$N_{\text{Im}}-C$	$C(O)_{\text{Gly}}-N(H)$	$N(H)_{\text{Thr}}-C(O)$	φ	τ	λ_{\max}	
QM	1.52	1.46	1.36	1.36	-6.65	-1.32	391 (0.59)	
LSCF/+1	<u>1.45</u>	<u>1.47</u>	1.33	1.33	-1.55	0.17	381 (0.48)	
LSCF/+3	<u>1.56</u>	<u>1.51</u>	1.32	1.34	-1.37	-0.21	388 (0.51)	
HLA	<u>1.51</u>	<u>1.45</u>	1.32	1.34	-1.81	-0.29	395 (0.48)	
LSCF/+3,+5	1.51	1.46	<u>1.42</u>	<u>1.36</u>	-2.47	-0.21	388 (0.61)	

Table 5.4: Distance en Å, angles dièdre en ° des frontières QM/MM pour les deux modèles. Le LSCF/+1, le LSCF/+3 et le HLA ont été testés sur le modèle (a) de la Figure 5.14, le second sous-bloc reprenant les valeurs pour le LSCF/+3,+5 et pour le HLA ont été testés sur le modèle (b) de la même Figure. La dernière colonne reprend les longueurs d'onde d'absorption maximale (λ_{\max}). Les forces d'oscillateur sont entre parenthèse. Niveau de théorie :TD-B3LYP/6-311+G(2d,p)//B3LYP/6-311G(d,p)

accord avec ces valeurs de référence à quelques nanomètres près.

Tout au long de ce travail sur les FP, nous avons donc choisi la coupure au niveau de la liaison peptidique pour des raisons pratiques mais aussi parce que la coupure en LSCF/+3 pourrait inclure des artefacts dûs aux positions des SLBO très proches du systèmes conjugués du chromophore. Nous rappelons que la méthode HLA n'est pas idéale si la coupure est située au niveau de la liaison peptidique ; cela induit une certaine instabilité dans la fonction d'onde parce que la liaison est fortement polarisée [205]. Une fois ce travail préliminaire sur un système modèle réalisé, nous allons maintenant pouvoir considérer l'ensemble de la protéine.

5.3.4 Spectre d'absorption des deux conformations de la ECFP

La Table 5.6 reprend les valeurs importantes que nous avons obtenues en LSCF/MM et en LSCF/MM:ERS. Dans un premier temps, nous allons considérer les conformations majeure et mineure de la ECFP pour un système qui ne contient aucune molécule d'eau. La première observation est que la considération de la protéine entière en LSCF/MM entraîne une augmentation de la longueur d'onde d'absorption maximale par rapport au calcul QM. Ce déplacement vers le rouge permet d'approcher étroitement la valeur expérimentale (434-452 nm). Par exemple, sur les conformations majeure et mineure de la ECFP, nous obtenons respectivement des λ_{\max} de 425 nm et de 426 nm. Ces deux valeurs sont très éloignées de la valeur obtenue sur le chromophore modèle en LSCF/MM (388 nm) mais sont seulement à 10 nm de la valeur expérimentale. Lorsqu'un continuum caractérisé par une constante diélectrique dynamique/optique est ajouté afin de décrire la réponse électronique de l'environnement (en plus des effets électrostatiques et de van der Waals), nous avons découvert que plusieurs transitions électroniques (λ_{\max}) possédaient des forces d'oscillateur non négligeables (de 0.12 à 0.18). Cela est d'autant plus étonnant que certaines de ces transitions n'apparaissent pas lors du calcul LSCF/MM. Les excitations électroniques prédites en LSCF/MM impliquent principalement des orbitales π et π^* (voir Figure 5.16).

En LSCF/MM:ERS, plusieurs autres orbitales de plus basses énergies participent à l'excitation avec des coefficients de mélange de l'ordre de 0.3. Par conséquent, les deux orbitales π et π^* sont aussi impliquées dans d'autres excitations. Le champ de réaction créé par le continuum modifie les coefficients de mélange des excitations où interviennent les orbitales π et π^* responsables de l'excitation électronique. Cela résulte en une augmentation de la force d'oscillateur.

Protéines	Chromophores	λ_{\max} (QM)				$\lambda_{\max}^{\text{exp}}$
		LSCF/+1	LSCF/+3	LSCF/+3,+5		
ECFP	Thr65-Trp66-Gly67	391	381	388	388	430-440
GdFP	Thr65-Trp66(-NH ₂)-Gly67	456	436	439	457	466
BFPF	Thr65-Phe66-Gly67	360	353	378	356	355-360
GFP(p,d)	Thr65-Tyr66-Gly67	372,415	360,409	384,476	371,431	392,414

Table 5.5: λ_{\max} en nm de chromophores extraits de la famille des AvFP (Figure 5.17) ; les deux formes de la GFP sont indiquées par p (protonée) et d (déprotonée). Les résultats QM sont obtenus en phase gazeuse. Les approches LSCF/+1, LSCF/+3 sont testées sur le modèle à gauche de la Figure 5.14 et sur le modèle LSCF/+3,5 (à droite de la figure).

Conformations	Paramètres géométriques				λ_{\max}	
	d(C _g -C-b)	d(C _b -C _i 1)	φ	τ	LSCF/MM	LSCF/MM:ERS
Majeure	1.43	1.37	-0.63	-0.66	425 (0.41)	426 (0.38)
						436 (0.18)
						447 (0.12)
Mineure	1.43	1.37	6.34	-0.97	426 (0.30)	427 (0.43)
						445 (0.22)
Majeure + eau	1.42	1.35	7.10	-2.54	405 (0.44)	419 (0.44)
						468 (0.18)
Mineure + eau	1.44	1.36	0.11	-1.06	406 (0.49)	413 (0.64)
						398 (0.08)*
Expérience [287]						434-452

Table 5.6: Les paramètres tels que les distances des deux liaisons formant le pont en Å et les angles dièdres en ° de conformations majeure et mineure de la ECFP sont repris (cf. Figure 5.12). Les valeurs des longueurs d'onde d'absorption maximale (λ_{\max}) obtenues à l'aide du LSCF/MM et du LSCF/MM:ERS se trouvent dans la table. Les valeurs entre parenthèse sont les forces d'oscillateur associées à la transition électronique. L'astérisque indique la présence d'une forte contribution de l'excitation $\pi^* \leftarrow n$ avec un coefficient de mélange de 0.64

Dans un premier temps, nous avons pensé que ce phénomène pourrait être à l'origine du double pic présent dans le spectre d'absorption expérimental de la ECFP. Nous verrons ci-après que ce n'est pas le cas. Dans la Table 5.6, pour les deux systèmes sans molécule d'eau, nous observons que les valeurs de λ_{\max} obtenues en LSCF/MM:ERS et qui possèdent les plus grandes forces d'oscillateur (426 nm et 427 nm) sont très proches des valeurs obtenues en LSCF/MM.

En plus de l'effet électronique très faible (+1 nm), le λ_{\max} est décomposé (voir section 5.3.2.A) afin de pouvoir décrire quelles sont les origines des valeurs obtenues en LSCF/MM:ERS ainsi que leurs contributions respectives. Pour rappel, les trois contributions de l'environnement sont de type géométrique, électrostatique et électronique. La Table 5.7 reprend ces contributions. Pour la conformation majeure, le passage du chromophore en phase gaz à son environnement

Conformations	$\lambda_{\max}(\text{QM})$	$\lambda_{\max}^{\text{géom}}$	$\lambda_{\max}^{\text{élec}}$	$\lambda_{\max}^{\text{ERS}}$
Majeure		+19	+15	<i>+1</i>
				+11
				+22
Mineure		+34	+1	<i>+1</i>
	391			-1
Majeure + eau		+52	-38	<i>+14</i>
				+63
Mineure + eau		+31	-16	<i>+6</i>
				-8*

Table 5.7: Cette Table reprend les contributions physiques de la longueur d'onde d'absorption en nm.

Cette décomposition du λ_{\max} est explicitée dans la section 5.3.2.A. Lorsque plusieurs valeurs sont fournies, la valeur en italique est celle pour laquelle le λ_{\max} possède la plus grand force d'oscillateur. L'astérisque indique que cette contribution est fortement due la présence de la transition $\pi^* \leftarrow n$

protéique induit une contribution géométrique très importante de +19 nm sur le spectre d'absorption. Le fait d'éteindre ou non les charges de l'environnement, entraîne un déplacement dû à la contribution électrostatique ($\Delta\lambda_{\max}^{\text{élec}}$) de + 15 nm. Ces deux premiers effets n'ont pas le même comportement pour la conformation mineure de la ECFP. En effet, pour la conformation

mineure la contribution géométrique est exacerbée (+34 nm) et la contribution électrostatique est insignifiante (+1 nm). L'*Electrostatic Embedding* a probablement un effet identique sur l'état fondamental et sur l'état excité pour cette conformation. Pour les deux conformations, la contribution géométrique induit un déplacement vers les grandes longueurs d'onde de +19 et de +34 nm. La géométrie du chromophore dans la conformation mineure est donc plus stabilisée par son environnement proche. La contribution électronique pour les deux conformations est similaire si seule la transition avec la plus importante force d'oscillateur est considérée (+1 nm).

5.3.4.A Et les molécules d'eau dans tout cela...

Les résultats obtenus ci-dessus ne prennent pas en compte les molécules d'eau se situant autour du chromophore. Or, nous avons observé que sept molécules d'eau cristallographique (W7, W11, W55, W117, W118, W162 et W163) se trouvaient dans un rayon de moins de 4.0 Å (voir Figure 5.18). Nous avons donc complexifié les calculs QM/MM, dans le souci de se rapprocher un maximum de la réalité. Ces molécules d'eau ont été prises en compte lors de l'optimisation de géométrie et lors du calcul de spectres d'absorption des deux conformations.

Les effets géométriques qu'entraîne la présence de ces molécules d'eau sur les deux conformations sont totalement différents, voire opposés. Le chromophore est moins (plus) plan en présence des molécules d'eau dans la conformation mineure (majeure). La conformation majeure devient moins planaire avec des angles dièdres de 7.10° et 2.54° tandis que la conformation mineure montre une planarité importante ($\varphi=0.11^\circ$ et $\tau=-1.06^\circ$). L'azote de l'imidazolinone joue un rôle important dans cette perturbation. Il forme une liaison hydrogène avec la molécule d'eau 117 dans la structure majeure, ce qui n'est pas le cas pour l'autre structure. La formation de liaisons hydrogènes induit une torsion importante du chromophore. Dans le cas de la contribution mineure, cinq des sept molécules d'eau interagissent uniquement entre elles, formant un amas d'eau sous le chromophore (W55, W117, W118, W162, et W163) formant ainsi un réseau de liaisons hydrogène indépendant du chromophore. Cette observation a été précédemment observée par le groupe d'Isabelle Demachy à l'aide d'une étude de dynamique moléculaire [280] : « *Another difference between major and minor conformations and concerns the five crystallographic water molecules closest to the chromophore (W55, W117, W118, W162, and W163). They form a H-bond network among each other and with neighboring residues; this network is more rigid in major conformation than in minor conformation.* ».

L'impact de l'introduction des molécules d'eau sur le spectre d'absorption est de diminuer la longueur d'onde d'absorption lorsque le chromophore est moins plan, c'est à dire dans le cas de la conformation majeure. En effet, nous observons un déplacement hypsochromique de 7 nm en LSCF/MM (passant de 426 à 419 nm). Étonnamment, cet effet hypsochromique est aussi observé dans le cas de la conformation mineure alors que le chromophore s'aplanit, allant de 427 nm à 413 nm. Le λ_{\max} de 413 nm est principalement dû à la transition $\pi^* \leftarrow \pi$ (coefficient de mélange de 0.59), dans cette contribution, la transition de type $\pi^* \leftarrow n$ est présente avec une contribution de 0.2)¹⁰. Nous avons observé les orbitales impliquées dans les transitions en LSCF/MM:ERS. Dans la conformation mineure, la contribution de la transition $\pi^* \leftarrow n$ est non négligeable. Dans cette conformation, la paire libre de l'azote n'est pas du tout impliquée dans le réseau de liaisons hydrogènes formé par la présence des molécules d'eau. Cette paire libre participe donc à une transition. La valeur de 398 nm entre parenthèse dans la Table 5.6 contient les orbitales responsables de la transition majoritaire $\pi^* \leftarrow \pi$ avec un coefficient de mélange faible de -0.12 ainsi que les orbitales de la transition $\pi^* \leftarrow n$ avec un coefficient de 0.64. Non seulement les molécules d'eau contribuent fortement à la modification de la géométrie du chromophore (+52 nm et +31 nm) mais en plus, elles jouent un rôle important dans ce cas précis sur l'implication de la paire libre de l'azote dans les transitions électroniques. La prise en compte de l'EE dans les deux cas a une contribution fortement négative (-38 nm et -16 nm). Les molécules d'eau jouent aussi un rôle déstabilisant sur la contribution électrostatique. L'ERS induit lui un déplacement légèrement plus grand de +14 et +6 nm par rapport à la contribution dans le calcul sans les molécules d'eau. La présence de la transition $\pi^* \leftarrow n$ induit par contre une effet négatif de 8 nm.

5.3.5 La mutation du chromophore

Dans l'introduction sur les protéines fluorescentes, nous avons indiqué que de nombreuses mutations de la GFP-originelle ont été réalisées afin de couvrir tout le spectre d'absorption/émission UV/visible; et ce, dans le but d'élargir le champ d'application de ces protéines. Le résidu aromatique en position 66 et l'imidazolinone formant le chromophore des protéines fluorescentes sont primordiales pour les propriétés spectroscopiques des protéines fluorescentes. Une manière

¹⁰Une autre transition apparaît aussi en LSCF/MM à une longueur d'onde de 437 nm avec les caractéristiques inverses. Nous ne l'avons pas indiquée pour plus de clarté dans les Tables.

est de générer des mutations autour du chromophore ou au sein même de ce dernier. Cela modifiera la stabilité du chromophore et donc, son λ_{\max} . Pour étudier ces effets de la mutation, deux stratégies peuvent être employées. Nous pouvons i) muter les acides aminés de l'environnement afin de modifier les propriétés optiques *via* des interactions intermoléculaires ou ii) modifier le chromophore lui-même tout en maintenant un même environnement. Cette étude traite uniquement la seconde option. Nous avons modifié le résidu 66 de la ECFP. Originellement, celui-ci est un tryptophane (Trp). Nous lui avons ajouté des groupements auxochromes comme un méthyle et une amine ou nous avons muté le Trp par d'autres acides aminés cycliques, la phenylalanine (Phe) et la tyrosine (Tyr). Ces chromophores sont représentés sur la Figure 5.19 avec la nomenclature utilisée ci-après. Pour des questions de facilité, nous avons nommé ces chromophores respectivement comme ceci : Trp66 (ECFP- originelle), Trp66-CH₃, Trp66-NH₂, Trp66Phe et Trp66Tyr. Les trois dernières notations sont les notations usuelles pour décrire les mutations ¹¹. Concernant le Trp66Tyr, nous avons traité ces deux états de protonation. Ces derniers sont fortement corrélés aux états de protonation du résidu glutamate222 (Glu222). Le Glu22 sera protoné (p) lorsque la Trp66Tyr sera déprotonée (d) ; et inversement. Toutes ces mutations seront considérées dans la conformation majeure de la ECFP parce que cette structure est la plus similaire aux autres structure de la famille des AvFP : les acides aminés His148 et Tyr145 situés respectivement à l'intérieur et à l'extérieur du tonneau de feuillet β . Pour cette étude, nous avons considéré uniquement la protéine, c'est-à-dire sans aucune molécule d'eau. Les systèmes sont donc assez simplifiés.

5.3.5.A Comparaison des résultats

La Table 5.8 reprend les données expérimentales ainsi que les λ_{\max} obtenus en QM sur le chromophore isolé et en LSCF/MM:ERS sur la protéine ECFP mutée. Notons que les données expérimentales ne correspondent pas à nos mutations mais bien à des mutations similaires ; excepté pour les valeurs de Trp66, Trp66-CH₃ et Trp66-NH₂ qui sont extraites de la publication de Huyen Bae [287]. Ceux-ci ont élaboré deux variantes de la ECFP, en ajoutant simplement un groupement fonctionnel méthyle (Trp66-CH₃) ou un groupement fonctionnel aminé (Trp66-NH₂), qui est connu sous le nom de GoldFP ou GdFP). La GoldFP est l'AvFP possédant le plus

¹¹Le premier acide aminé est le muté, le nombre indique la position dans la séquence et les trois dernières lettres correspondent au nouvel acide aminé mutant.

grand déplacement vers le rouge c'est-à-dire un déplacement de +69 nm entre le pic d'absorption et son pic d'émission. Elle possède aussi une forte stabilité et peu de tendance à s'agréger.

Lorsque le résidu 66 est muté par le résidu phenylalanine (Phe), la valeur expérimentale correspond au chromophore Ser65-Phe66-Gly67 de la GFP sauvage [265]. Le chromophore Ser65-Tyr66-Gly67 existe sous deux formes en fonction du système : protonée (p) et déprotoné (d) [265]. L'unique mutation de Ser65Thr induit un déplacement vers les grandes longueurs d'onde (+13 nm) pour la forme anionique du chromophore Xxx65-Tyr66-Gly67 [265]. Expérimentalement, les chromophores formés à partir de la Tyr66 existent dans un état protoné ou déprotoné pour la GFP sauvage, possédant respectivement un pic d'absorption à 395 nm et 476 nm [261]. La mutation Ser65Thr de la GFP sauvage entraîne la disparition d'un pic et donc, la présence d'un seul état de protonation. La forme anionique persiste avec un λ_{\max} de 489 nm. La présence de mutation dans l'environnement du chromophore induit la présence de deux pics : 366 nm sous forme protonée et 504 nm sous forme déprotonée [286]. L'état de protonation de la Tyr66 est délicat, nous avons donc considéré les deux états de protonation dans l'environnement de la ECFP.

Un graphe comparatif (Figure ??) entre les résultats obtenus en LSCF/MM:ERS et les données expérimentales montre un certain accord avec une déviation maximale entre les valeurs de -42 nm. Cet écart correspond à la mutation Trp66Tyr (d). Les autres écarts par rapport aux données expérimentales se trouvent entre -10 nm et +10 nm.

Dans la Table 5.8, nous remarquons qu'il s'agit d'une sérieuse amélioration par rapport aux calculs réalisés sur les chromophores isolés. En effet, pour les trois premiers résultats LSCF/MM:ERS de la Table 5.8, les valeurs expérimentales sont quasiment reproduites. Si l'on compare les transitions électroniques correspondantes à la force d'oscillateur la plus grande (en italique dans la Table 5.8), nous pouvons remarquer que l'addition du méthyle sur Trp66 (Trp66-CH₃) ne modifie pas le spectre d'absorption théorique et expérimental par rapport à la ECFP originelle (Trp66). Avec le LSCF/MM:ERS, nous obtenons un écart de +2 nm suite à l'addition du méthyle (426 nm à 426 nm), tandis que dans l'expérience, cette déviation est de +4 nm (434-452 nm à 438-456 nm). *A contrario*, le groupement aminé induit une forte augmentation du λ_{\max} de +29 nm (426 nm à 455 nm); ce qui peut être comparé à l'expérience (+32 nm). D'un point de vue théorique, nous observons pour Trp66 et Trp66-NH₂ plusieurs transitions électroniques actives possédant des forces d'oscillateur non négligeables (entre 0.12 et 0.18). Ces transitions

Modèle	λ_{\max}			Structure expérimentale
	QM	LSCF/MM:ERS	Expérience	
Trp66		447 (0.12)		
ECFP-originelle	391 (0.6)	436 (0.18)	434-352 [287]	Thr65-Trp66 (ECFP)
		<i>426 (0.38)</i>		
Trp66-CH3	398 (0.6)	<i>428 (0.6)</i>	438-456 [287]	Thr65-Trp66-CH ₃ (ECFP)
Trp66-NH2	456 (0.3)	<i>455 (0.43)</i>	[287]	Thr65-Trp66-NH ₂ (ECFP)
	354 (0.2)	439 (0.18)		
Trp66Phe	360 (0.6)	400 (0.31)	360 [265]	Ser65-Tyr66 (wt-GFP)
		<i>361 (0.27)</i>		
Trp66Tyr (p)	372 (0.6)	406 (0.2)	396 [303]	Ser65-Tyr66 (wt-GFP)
		<i>406 (0.48)</i>	396 [286]	Thr65-Tyr66, Met153Ala, Lys238Glu
Trp66Tyr (d)*	415 (0.9)	<i>434 (1.00)</i>	476 [303]	Ser65-Tyr66 (wt-GFP)
			504 [286]	Thr65-Tyr66, Met153Ala, Lys238Glu
			489 [286]	Thr65-Tyr66, Met153Ala, Lys238Glu

Table 5.8: Comparaison entre les valeurs de longueurs d'onde d'absorption maximale (λ_{\max}) expérimentale, en tout QM et en LSCF/MM:ERS.

Les valeurs de λ_{\max} sont en nm et les valeurs entre parenthèse sont les forces d'oscillateur. Les résultats LSCF/MM:ERS en italiques correspondent à la contribution majeure de la transition électronique $\pi^* \leftarrow \pi$. La dernière colonne reprend les structures expérimentales pour lesquelles les valeurs de λ_{\max} sont données.

ne se retrouvent pas dans le spectre d'absorption du mutant Trp66-CH₃. Habituellement, la transition permise implique exactement les mêmes orbitales π et π^* lorsque le chromophore est isolé ou dans son environnement (cf. Figure 5.16)). Les autres excitations impliquent des orbitales occupées de plus basses énergies contenant une petite contribution de l'orbitale π et ont une très petite force d'oscillateur. Comme pour les deux conformations de la ECFP, le champ de réaction du continuum, utilisé pour reproduire la contribution électronique de la partie MM, modifie les coefficients de mélange dans ces excitations. Cela induit une augmentation du coefficient de l'orbitale π résultant ainsi directement en une augmentation de la force d'oscillateur. Nous avons supposé que ces nouvelles transitions inattendues pouvaient être à l'origine du double pic présent dans la ECFP originelle. Or, lors de la mutation Trp66-CH₃, un double pic est aussi observé expérimentalement ; alors que théoriquement, nous n'observons pas ces autres transitions actives. De plus, lorsque plusieurs transitions électroniques sont mises en évidence théoriquement, aucun double pic n'apparaît expérimentalement (cas de la mutation Trp66-NH₂). Trp66Phe exhibe aussi deux transitions avec des forces d'oscillateur de 0.31 et 0.27. Il serait donc trop prétentieux de faire une relation directe entre les deux pics et les multiples transitions électroniquement actives.

Finalement, les états de (dé)protonations induisent une modification importante du spectre d'absorption, que ce soit pour la position du pic ou pour son intensité. La différence calculée (+28 nm) entre les deux formes est éloignée de la variation expérimentale de +80 nm ou de +108 nm selon les systèmes (Table 5.8). Cependant, les données expérimentales correspondent à des structures pour lesquelles l'environnement est différent de l'environnement de la ECFP. Dans ce cas, la comparaison théorie/expérience ne peut être que d'intérêt qualitatif.

5.3.5.B Les différentes contributions de la longueur d'onde d'absorption

Nous avons donc utilisé le schéma de décomposition de la longueur d'onde d'absorption décrit en section 5.3.2.A. Tous les termes de cette décomposition sont repris dans la Table 5.9. Comme attendu, les λ_{\max} calculés pour les chromophores isolés présentent une importante déviation par rapport à l'expérience (de -10 à -112 nm). Le seul composé qui « reproduit » l'expérience est la mutation Trp66Phe pour laquelle le λ_{\max} est identique (360 nm)! Les écarts sont dus principalement à l'absence d'environnement autour de ces chromophores. L'environnement joue, en effet, un rôle important sur la géométrie du chromophore et donc sur son λ_{\max} . Pour la plupart

des chromophores, le λ_{\max} a été déplacé vers les grandes longueurs d'onde lors de leur passage d'un environnement gazeux vers un environnement protéique. Une seule exception est présente, c'est le cas de la forme déprotonée de la mutation Trp66Tyr.

Dans la Table 5.10, nous pouvons observer que la géométrie des chromophores a été effectivement modifiée d'un état à l'autre. Cette Table reprend les paramètres décrivant la planarité de ce type de chromophore (les angles dièdres φ et τ , cf. chromophore (a) de la Figure 5.19). Les mutations Trp66Phe et Trp66Tyr (les deux états de protonation) exhibent une torsion plus importante entre les deux cycles de part et d'autre du pont, rompant la planarité. Si la transition électronique implique uniquement un caractère $\pi^* \leftarrow \pi$, l'écart énergétique entre ces orbitales devrait diminuer (augmenter) tant que la géométrie du chromophore est plus (moins) plan. Cela devrait donc induire une augmentation (diminution) du λ_{\max} . Ce principe simple fonctionne très bien tant qu'aucun autre effet n'existe. Il est applicable à des systèmes π très réguliers, par exemple, de type polyène. Notons que ce principe fonctionne très bien pour quatre des mutations : Trp66 (ECFP-orginelle), Trp66-CH₃, Trp66-NH₂ et la forme déprotonée de Trp66Tyr. Il est assez surprenant que ce principe marche aussi bien puisque la transition électronique implique un mélange de plusieurs excitations et que les chromophores possèdent des groupes polaires (carboxyle, amine et hydroxyle) créant des champs électriques qui peuvent interagir différemment avec l'état fondamental et l'état excité du chromophore. Donc, la déformation de la géométrie modifie le poids relatif de toutes les excitations d'une manière non triviale et il est par conséquent, normal d'observer un comportement incorrect de cette règle basée uniquement sur la planarité du chromophore. Un autre effet peut être invoqué pour la mutation Trp66Tyr dans la forme anionique. En effet, dans cette forme, il y existe un pont hydrogène entre l'hydroxyle de la thréonine en position 65 et l'atome d'azote de l'imidazolinone (Figure 5.19). Cette liaison est aussi présente suite à l'optimisation en QM sur le chromophore isolé. Elle persiste sur le système protéique Trp66Tyr (d). Comme la paire libre de l'atome d'azote est aussi légèrement impliquée dans la transition électronique, la liaison hydrogène a un effet direct sur le λ_{\max} . La liaison hydrogène stabilise la paire libre de l'azote dans l'état fondamental, ce qui augmente l'écart énergétique entre les orbitales impliquées dans la transitions électroniques. Lors de l'optimisation en QM, chaque chromophore exhibe cette liaison hydrogène, ce qui n'est pas le cas en présence de l'environnement, excepté pour le Trp66Tyr (d). Il est intuitif d'observer une plus faible λ_{\max} pour les chromophores optimisés en QM et ceux optimisés en LSCF/MM. Cela induit

des contributions géométriques toujours positives ($\Delta\lambda_{\max}^{\text{géom}} > 0$) dans la Table 5.9 excepté, donc, pour le Trp66Tyr (d). Pour ce dernier cas, la présence de la liaison hydrogène empêche la règle de planarité explicitée ci-dessus de s'appliquer.

Modèle	QM		LSCF/MM	
	φ	τ	φ	τ
ECFP-native (Trp66)	-6.65	-1.32	-2.47	0.21
Trp66-CH3	- 7.76	-1.73	3.67	-0.36
Trp66-NH2	-12.14	-1.68	-0.42	1.13
Trp66Phe	7.57	1.57	11.11	5.58
Trp66Tyr (p)	5.44	1.92	11.56	3.84
Trp66Tyr (d)*	- 2.58	-1.72	9.86	8.13

Table 5.10: Angles dièdres en degré des structures optimisées en QM et en LSCF/MM. φ et τ sont les angles dièdres représentés sur le chromophore (a) de la Figure 5.19. L'astérisque indique la présence d'une liaison hydrogène entre la Thr65 et l'imidazolinone lors du calcul LSCF/MM. Les valeurs pour la ECFP-native sont celles reprises de la Table 5.6.

La polarisation de la fonction d'onde par les charges ponctuelles possède une importante contribution sur le λ_{\max} (entre -5 nm et +29 nm). La mutation Trp66Phe a la plus petite contribution électrostatique ($\Delta\lambda_{\max}^{\text{élec}} = +6$ nm). Cet effet est certainement dû au fait que le Phe possède le plus petit cycle aromatique. Ce chromophore est le plus petit dans l'environnement de la ECFP, la distance entre le phényle de la phénylalanine et les charges ponctuelles MM (de l'environnement de la ECFP) est plus importante. La contribution Coulombique sera, par conséquent plus faible. Cet effet devrait être moindre en présence de molécules d'eau dans le modèle. La forme protonée du chromophore Trp66Tyr est plus sensible aux interactions électrostatiques. Cela semble raisonnable car souvenons-nous que dans cet état de protonation, le chromophore possède un résidu chargé (Glu222) dans son environnement proche.

À l'aide de cette décomposition, nous pouvons donc mettre en évidence l'amplitude de la contribution de l'ERS dans notre modèle. La réponse électronique de l'environnement passe de -12 nm pour la mutation Trp66Tyr (p) jusqu'à +22 nm pour le chromophore de la ECFP native. Nous avons déjà observé une variation positive de +11 et de +39 nm pour le chromophore de

la ECFP native et pour le complexe ST (+28 nm). Dans cette étude, nous observons pour la première fois un déplacement vers les petites longueurs d'onde de -6 et -12 nm pour respectivement, Trp66NH₂ et Trp66Tyr (p). Cela indique que l'état fondamental interagit de manière plus favorable avec le continuum que l'état excité.

Finalement, nous pouvons dire que pour une protéine donnée, l'environnement influence les propriétés optiques et géométriques de manière différente pour chaque chromophore distinct. Aucune propriété additive ne peut donc être tirées et une étude spécifique doit avoir lieu.

5.3.6 Conclusion générale sur l'étude des protéines fluorescentes

Dans un premier temps, nous avons testé les différentes coupures afin de scinder le système protéique en deux sous-systèmes. Sur le chromophore de la ECFP, LSCF/+3 et LSCF/+3,+5 ont montré tous deux de bons résultats quant à la géométrie et au λ_{\max} obtenus. Par contre sur quatre autres chromophores de la famille des AvFP, les résultats obtenus en LSCF/+3,5 sont nettement meilleurs que ceux obtenus en LSCF/+3 avec une erreur maximale de seulement 3 nm. La coupure au niveau des liaisons peptidiques a donc été choisie pour le traitement des protéines fluorescentes. Cette coupure facilite la définition des charges puisque celles-ci ne doivent pas être redistribuées dans la protéine parce que la liaison peptidique est une coupure naturelle des acides aminés et donc des champs de forces qui les décrivent.

Nous avons ensuite utilisé le LSCF/+3,+5 pour étudier les propriétés géométriques et optiques des deux conformation de la ECFP. En négligeant les molécules d'eau, la prise en compte de l'environnement stabilise la chromophore ce qui augmente la longueur d'onde d'absorption-maximale par rapport au chromophore en phase gazeuse. La présence des molécules d'eau avoisinantes du chromophore induit de fortes modifications, géométriques et optiques. La géométrie du chromophore de la conformation mineure est plus planaire que celle de la conformation majeure de la protéine. Cet effet est principalement dû au réseau de liaison hydrogène qui se forme sous le chromophore (mineure) ou avec le chromophore (majeure). Ce réseau de liaisons hydrogènes dans la conformation majeure implique que la transition électronique $\pi^* \leftarrow n$ est plus faible. Un mélange léger entre les transitions $\pi^* \leftarrow \pi$ et $\pi^* \leftarrow n$ est donc observé.

La dernière partie de ce chapitre sur les protéines fluorescente concerne les effets de la substitution ou mutation du chromophore, nous avons à nouveau utilisé notre modèle pour prendre en compte la réponse électronique de l'environnement. Avec ce modèle LSCF/MM:ERS, nous

obtenant une déviation de + ou -10 nm par rapport à des résultats expérimentaux possédant des mutations similaires. La décomposition de la longueur d'onde d'absorption maximale en effets géométriques, électrostatiques et électroniques a montré que ces trois contributions ont le même ordre de grandeur. L'influence de l'ERS doit être tout autant considéré que les autres contributions habituellement prise en compte (électrostatiques et géométriques) lorsque l'on traite un chromophore dans un environnement protéique.

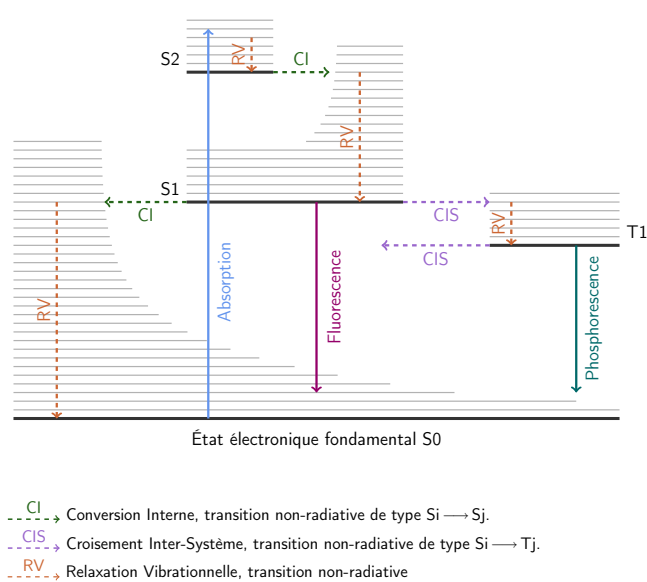


Figure 5.1: Diagramme de Perrin-Jablonski.

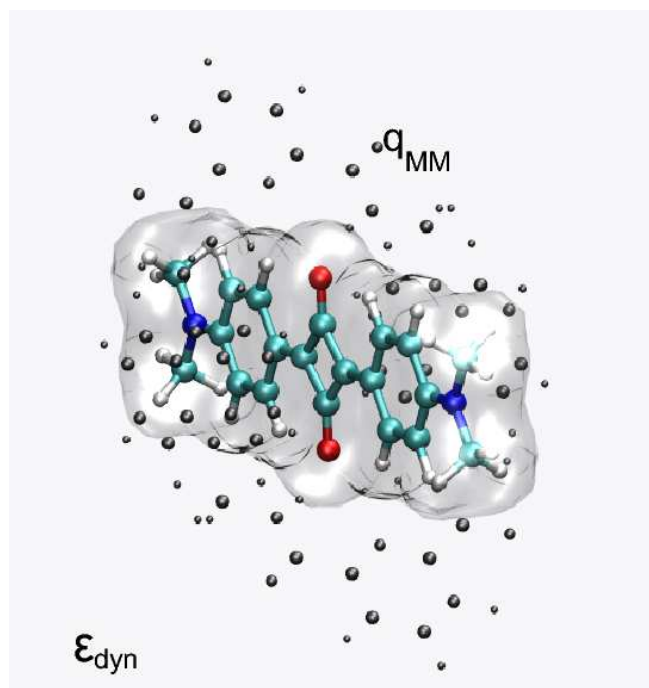


Figure 5.2: Chromophore de squaraine dans sa cavité. Les charges MM modélisant l'environnement sont représentées par des ronds noirs.

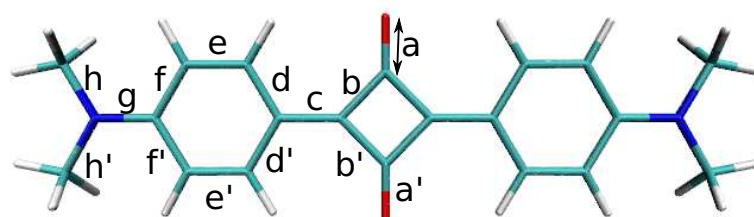


Figure 5.3: Représentation du colorant squaraine ainsi que la nomenclature employée pour définir les paramètres géométriques de liaisons.

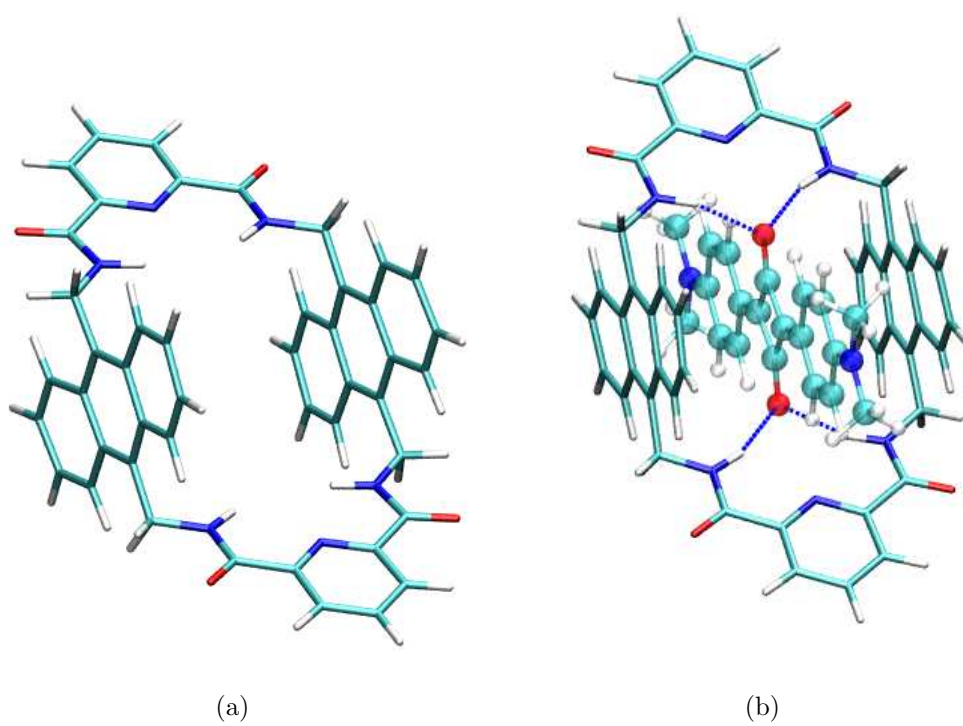


Figure 5.4: À gauche, la cage de tétralactame et à droite le complexe de squaraine-tétralactame.

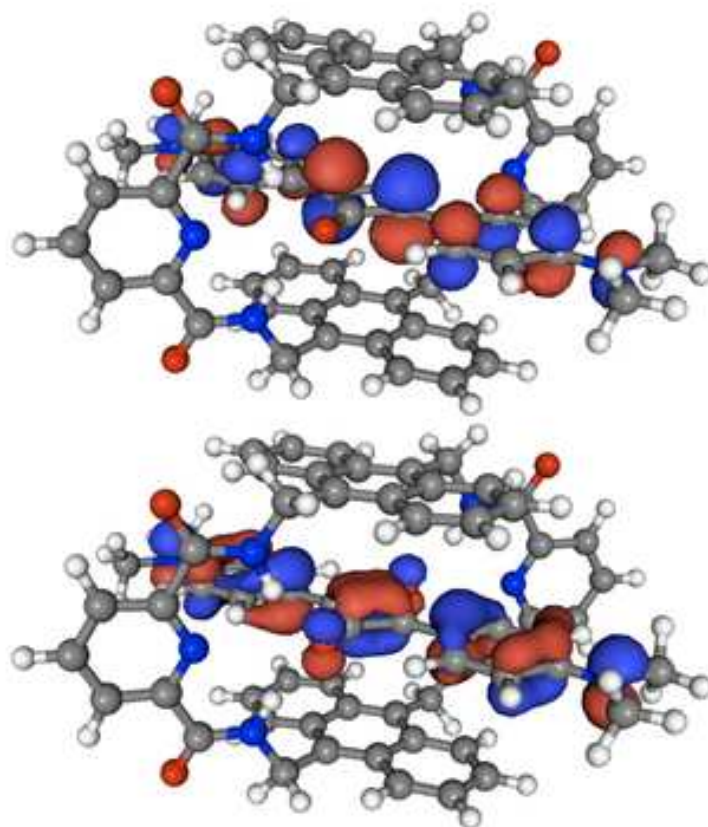


Figure 5.5: Orbitales moléculaires impliquées dans les transitions électroniques du complexe de squaraine-tétralactame, les HOMO-2 en bas et les LUMO en haut (valeur de l'isodensité = 0.035 a.u.).

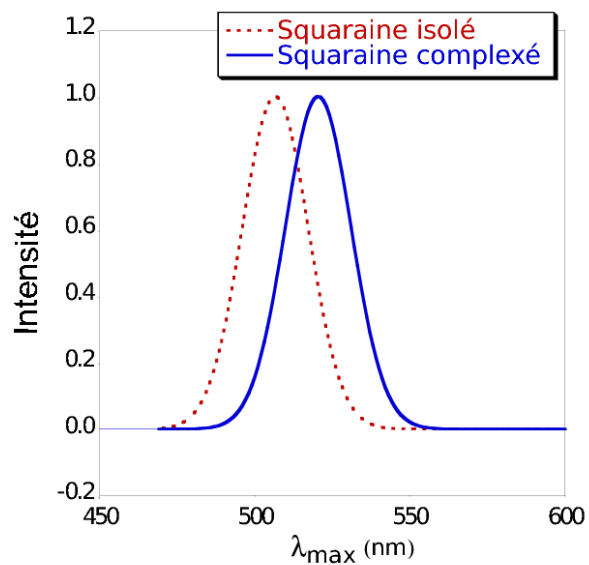


Figure 5.6: Spectre d'absorption du colorant squaraine isolé et complexé. Les pics correspondent à une gaussienne avec une largeur à mi-hauteur 25 nm. L'intensité est totalement arbitraire



Figure 5.7: Ensemble des mutants élaborés par le groupe Tsien [261].

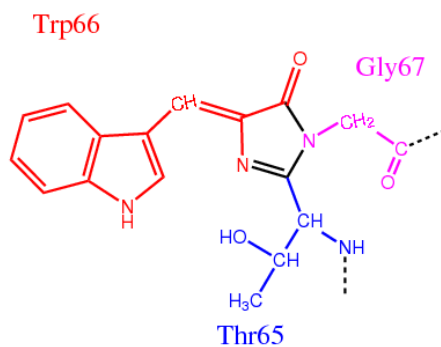


Figure 5.10: Le chromophore Thr65-Trp66-Gly67 de la ECFP.

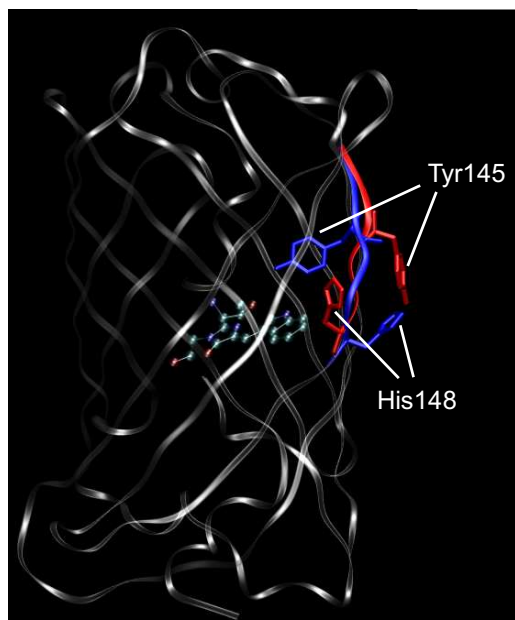


Figure 5.11: Les résidus bleus/rouges sont la Tyr145 et la His148 dans la conformation majeure/mineure.

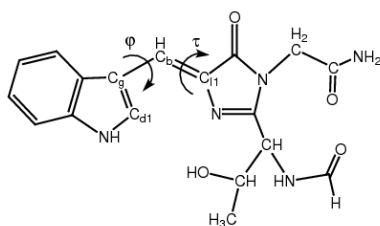


Figure 5.12: Chromophore satur  de la ECFP. La nomenclature concernant les param res g om triques observ s est not e.

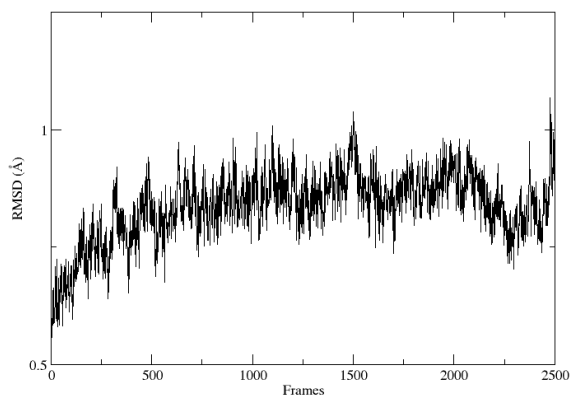


Figure 5.13: RMSD de l'ossature de la conformation majeure par rapport à sa structure cristallographique.

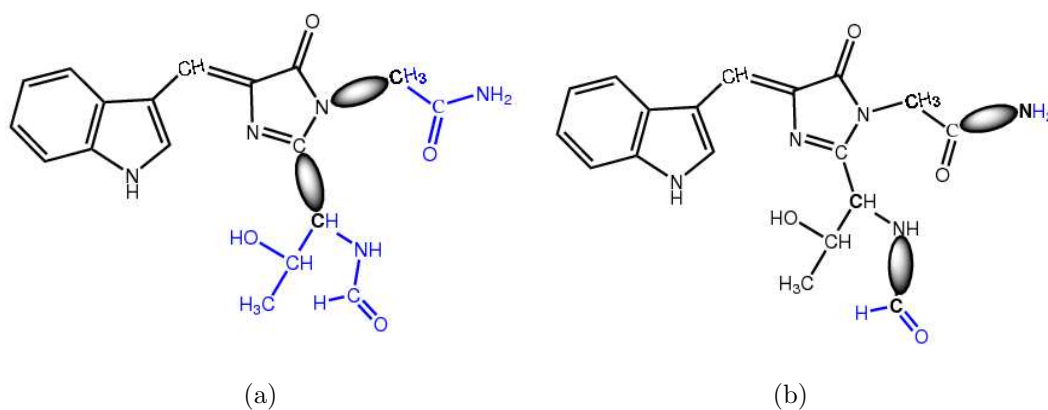


Figure 5.14: Localisation du jeu de SLBO. À gauche, sont représentées les SLBO localisées proches de l'imidazolinone. Ce modèle est utilisé pour tester le LSCF/+1 et le LSCF/+3. Le LSCF/+3,+5 est testé sur le modèle de droite où les SLBO se situent sur la liaison peptidique. Les parties en bleu sont les parties MM. L'atome frontière (YA) est représenté en gras.

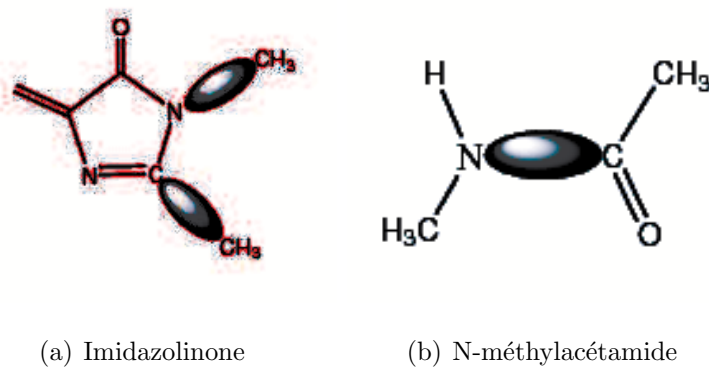


Figure 5.15: Représentation des systèmes modèles utilisés pour définir les SLBOs

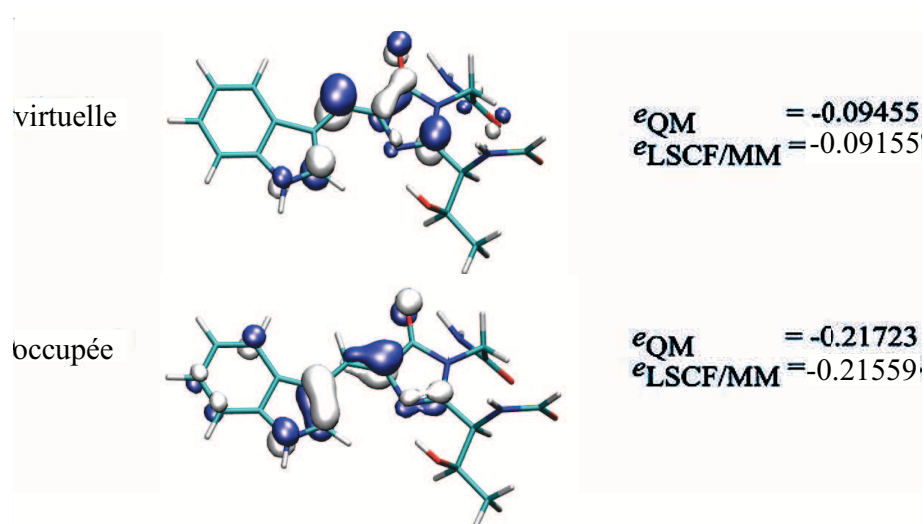


Figure 5.16: Représentation des orbitales moléculaires impliquées dans la transition électronique. Les énergies des orbitales sont données en unité atomique. Le LSCF/MM présenté est le schéma LSCF/+3,+5 sur le chromophore modèle de la ECFP.

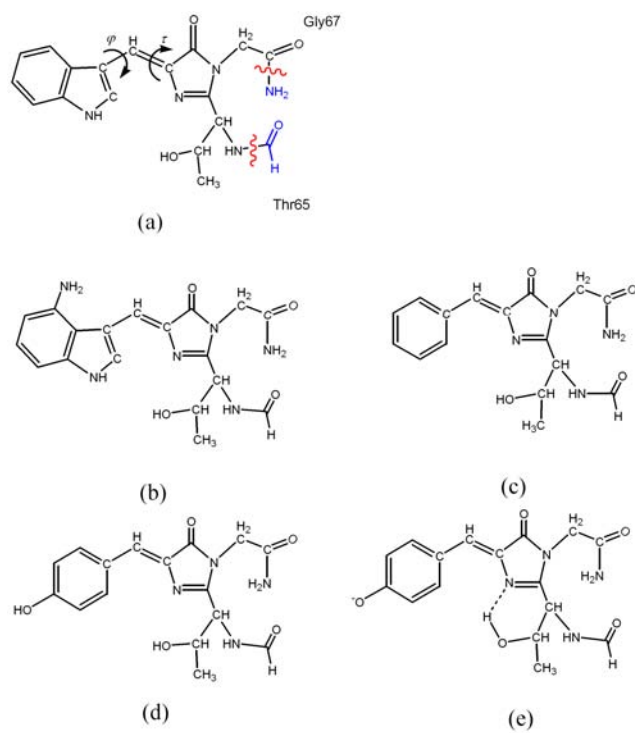


Figure 5.17: (a) est le chromophore de la ECFP (Thr65-Trp66-Gly67), (b) le chromophore de la goldFP (GdFP, Thr65-Trp66(-NH₂)-Gly67) (c) le chromophore de la BFPF (Thr65-Phe66-Gly67) et (d) et (e) sont les formes protonées et déprotonées de la GFP (Thr65-Tyr66-Gly67)

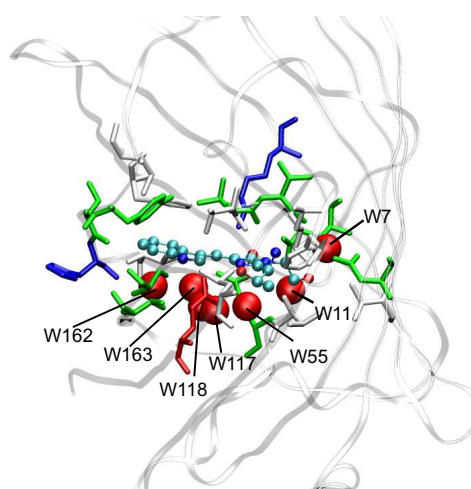


Figure 5.18: Molécules d'eau cristallographique qui ont été considérées lors des calculs QM/MM. Le chromophore se trouve dans une représentation boule et bâton, les molécules d'eau sont décrites par des boules rouges. Les résidus autour du chromophore sont représentés en bâtons. Leur couleur est attribuée en fonction du type d'acides aminés : basique bleu, acide rouge, polaire vert et non polaire blanc.

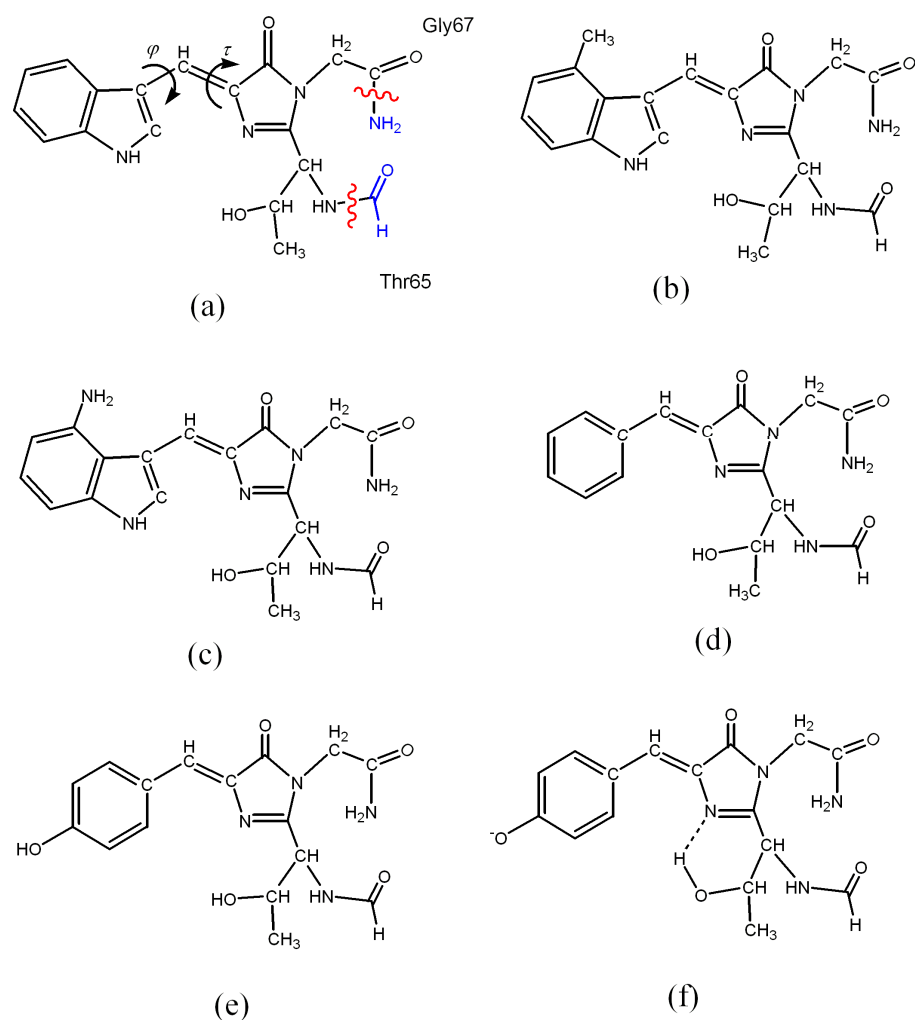


Figure 5.19: Représentation des chromophores. (a) est le chromophore de la ECFP originale (Trp66-ECFP), décrit avec la position de la frontière QM/MM. Les atomes MM sont représentés en bleu, le reste des atomes sera traité en QM. (b) Trp66-CH₃, (c) Trp66-NH₂, (d) Trp66Phe, (e) Trp66Tyr (p) et (f) Trp66Tyr (p). (p) et (d) caractérisent les formes protonée et déprotonée.

Chapitre 6

La capture électronique

CETTE partie va nous permettre de comprendre la capture électronique des ponts disulfures dans des systèmes macromoléculaires. Ce travail a été réalisé dans le cadre d'une collaboration avec le docteur Élise Dumont (ENS Lyon), co-encadrante de cette thèse, et a été le sujet de plusieurs publications [51, 256, 304, 305]. E. Dumont s'est intéressée à la liaison disulfure lors de son passage dans notre groupe [306–308]. Cette collaboration a ensuite perduré pour donner plusieurs publications et notamment une publication sur une macromolécule contenant trois ponts disulfures (un cyclotide) [304]. Cette dernière étude est l'objet de ce chapitre.

Dans les systèmes biologiques, de nombreuses réactions d'oxydo-réduction (rédox) ont lieu perpétuellement. Ces réactions d'oxydo-réductions sont caractérisées par des transferts d'électrons entre les différentes espèces (oxydantes ou réductrices). La réaction d'oxydation (réduction) entraîne une oxydation (réduction) d'une espèce réductrice. Cependant, les potentiels d'oxydo-réduction les caractérisant sont assez délicats à obtenir d'un point de vue théorique. Le transfert d'électrons dans la théorie de Marcus est un processus vertical non radiatif. Il est aussi crucial de décrire correctement l'état de non-équilibre créé par ce transfert d'électrons. Or, nous avons vu que la majorité des champs de forces ne sont paramétrisés que pour décrire des structures à l'équilibre et ne prennent pas en compte la réponse ou polarisation électronique de l'environnement. Dans le cas des réactions rédox, la contribution électronique doit être absolument bien décrite, celle-ci a, en effet, un rôle tout aussi important que la contribution orientationnelle pour déterminer un potentiel rédox [309]. Les méthodes hybrides de type Car-Parinello (CPMD) permettent de coupler l'aspect orientationnel et électronique [310]. Cependant, le niveau de théorie de la partie d'intérêt en CPMD est habituellement des fonctionelles GGA qui, malheureuse-

ment ne décrivent pas bien le transfert de charges entre molécules. Une solution actuellement employée est de considérer une demi-réaction, c'est à dire des systèmes biologiques possédant un site souvent métallique de type : $\text{Cu}^+/\text{Cu}^{2+}$, $\text{Ag}^+/\text{Ag}^{2+}$ ou encore $\text{Ru}^{2+}/\text{Ru}^{3+}$ [311,312].

Une quantité nettement plus simple à calculer pour décrire la capture électronique est l'affinité électronique. L'affinité électronique adiabatique (AEA) est la quantité d'énergie libérée suite à la capture d'un électron par un atome ou une molécule en phase gazeuse. Plus l'AEA est grande, plus la capture d'un électron libère de l'énergie et plus cette capture est favorable. Une AEA négative signifie au contraire qu'il faut fournir de l'énergie au système pour qu'il accepte un électron. La capture électronique sera donc défavorable et l'électron s'échappera tout seul rapidement. La Figure 6.1 reprend les quantités décrivant la capture électronique : l'AEA, le VEA (*Vertical Electron Affinity*) et le VDE (*Vertical Detachment Electron*).

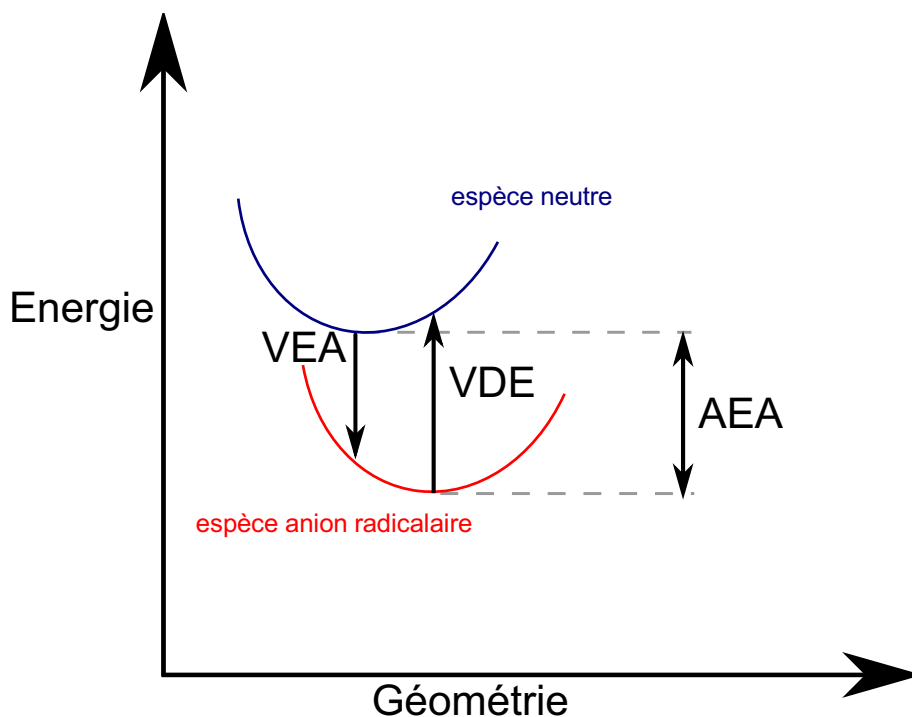


Figure 6.1: Représentation schématique des grandeurs caractérisant la capture électronique. La coupe de la surface d'énergie potentiel de l'espèce neutre (chargée négativement) est en bleu (rouge).

Dans les protéines, deux acides aminés cystéine (Cys) peuvent s'associer par formation d'une liaison covalente, appelée liaison ou pont disulfure (liaison ou pont SS). Ceux-ci, une fois formés, sont cruciaux pour la stabilisation ou le maintien de la structure tertiaire et quaternaire

de protéines. Ils sont présents dans de nombreuses structures comme les anticorps du système immunitaire, l'insuline (trois ponts SS), ... On les trouve également au sein du site actif de certaines enzymes impliquées dans des processus d'oxydo-réduction, de régularisation du potentiel rédox intra-cellulaire ou de transport d'électrons ; un exemple est la thiorédoxine (Trx). Cependant, les ponts SS formés, bien que responsables de la rigidité des protéines, sont aussi caractérisés par une énergie de dissociation relativement faible¹. Ils peuvent donc facilement être dissociés. Cette aptitude constitue un atout majeur dans de nombreuses applications. La rupture ou formation de liaisons disulfures peut notamment être utilisée pour contrôler respectivement l'ouverture ou la fermeture d'un macrocycle [314]. Elle peut aussi enclencher le dépliement d'un térapeptide [315], la formation d'une hélice α suite à une dissociation photochimique de la liaison disulfure [316]. Ces applications montrent à quel point il est important de comprendre

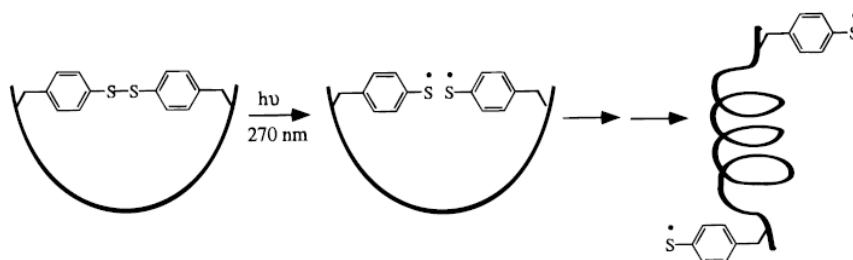


Figure 6.2: Représentation de la formation d'une hélice α (riche en alanine) suite à une photodissociation [316]

la réactivité de ces liaisons si particulières afin de pouvoir contrôler la structure/*folding* d'une protéine.

Depuis de nombreuses années, des études, tant expérimentales que théoriques ont donc essayé de comprendre les processus chimiques qu'impliquent ces liaisons lors de la capture d'un électron ou à l'absorption d'un photon. Au début des années trente, Pauling a suggéré l'existence d'anions radicalaires (RA, *Radical Anion*) possédant une liaison de type « deux-centres trois-électrons » (2c-3e) [317]. Ensuite, des études expérimentales et théoriques ont établi que les ponts disulfures pouvaient capter des électrons, formant ainsi une demi-liaison covalente² soufre-soufre allongée d'environ 0.7 Å par rapport à leur forme neutre, passant de 2.0 Å à 2.7 Å. Cette liaison 2S-3e

¹Par exemple pour le diméthyle disulfure (DMeDS), l'énergie de dissociation est d'environ 60-65 kcal.mole⁻¹ [313]

²Ordre de liaison=1

(deux atomes de soufre-trois électrons) a été ensuite qualifiée de « demi-liaison » [318, 319]. Les

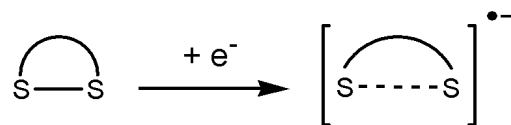


Figure 6.3: Représentation schématique de la capture électronique par un pont disulfure engagé dans un cycle. Le caractère de demi-liaison est indiqué en pointillés

RA de type 2c-3e formés ont ainsi pu être détectés avec une certaine stabilité dans des systèmes biologiques à l'aide de différentes analyses (des techniques de photolyse laser flash [320], de radiolyses à impulsion [320–322], de résonance paramagnétique électronique [323], d'électrochimie [324], d'absorption UV/visible [325, 326], ...).

Dans une protéine contenant plusieurs ponts disulfures, il est intéressant de caractériser la réactivité de chacun d'entre-eux, de déterminer les changements conformationnels suite à la capture électronique sur chacun des ponts, ... Bref, de prédire le site de l'attachement électronique et ses effets structuraux. Ces interrogations sur la formation de ces liaisons 2S-3e peuvent être quantifiées par leur affinité électronique adiabatique. L'élongation de la liaison soufre-soufre et la différence de charges entre réactifs et produits sont les deux critères de formation de liaisons à 2c-3e. Les deux contributions qui moduleront la capture électronique seront principalement géométriques (distance entre les deux atomes de soufre (d_{SS}) et angle dièdre ($\tau_{C-S-S-C}$)) et électrostatiques. Récemment, nous avons mené des études sur des systèmes modèles contenant une liaison disulfure, donnant lieu à une première compréhension des paramètres qui influencent l'affinité électronique adiabatique (AEA) :

- La tension conformationnelle ou frustration topologique favorise fortement la capture électronique [305–307]. Cet effet est géométrique, il augmente l'affinité électronique de +0.96 eV pour le motif Cys-Gly-Pro-Cys (CGPC) présent dans le site actif de la Thioredoxin (Trx) (une enzyme anti-oxydante) par rapport à la molécule linéaire DEtDS (Diéthyle Disulfure) [307].
- La structure secondaire est primordiale. En effet, nous avons montré que le dipôle créé par une hélice α de 12 résidus alanine sur le motif CGPC augmente l'AEA de +0.9 eV ($AEA_{\text{tot}} = +2.12$ eV). Ce modèle simple a permis d'approcher fortement la valeur calculée

pour la Trx de +2.20 eV [308].

- Une charge ou un résidu chargé, même situé à grande distance (10 Å), modifie fortement l'AEA d'un pont disulfure [327, 328].

Cependant, ces résultats n'ont porté que sur des systèmes mono-pont disulfure. Peu d'études théoriques ont été réalisées sur la protéine entière, se limitant souvent au site réactif. Bergès (LCT Jussieu) a considéré les Trx dans leur entièreté à l'aide de l'approche hybride ONIOM [321, 328–330]. Aucune étude théorique n'a été réalisée sur des macrocycles contenant plusieurs liaisons SS. Un travail expérimental menée sur une protéine contenant trois ponts disulfures (TcAChE, *Torpedo californica AcetylCholinesterase*, code pdb 2ACE) a été mené par Weik et collaborateur en 2000 montrant explicitement l'élongation d'une liaison soufre-soufre suite à une radiation électronique [331]. Ce travail a montré que la capture électronique était hautement spécifique [332]. Or, dans le travail de Weik sur TcAChE, plusieurs questions sur ces macrocycles à multiples ponts disulfures ont été émises :

1. Quel est l'ordre de réactivité de ces ponts SS? La figure (6.4) montre tous les cas de figure possible de la compétitivité lors de la capture d'un, puis d'un second électron, ainsi que la nomenclature utilisée pour un système contenant deux liaisons disulfures. Cette nomenclature sera utilisée par la suite pour un système contenant trois liaisons SS.
2. Où se fera l'addition du second électron? Sera-t-il capté par la liaison SS qui est déjà sous forme anion radicalaire (RA) résultant en une espèce dithiolate (DT)? Ou alors ira-t-il se positionner sur un pont SS (neutre) formant ainsi deux anions radicalaires : un dianion diradicalaire (DD)?

La seconde question concernant l'addition du second électron est intéressante parce que la formation du dithiolate entraîne une dissociation de la liaison, pouvant engendrer la déstructuration de la protéine. Ces questions sont cruciales à la compréhension du mécanisme de la capture électronique et leurs réponses ne peuvent pas être observées expérimentalement. Une étude théorique de tels systèmes est susceptible d'apporter des réponses. Nous avons considéré un macrocycle de 29 acides aminés contenant trois ponts disulfures, nommé kalata B1 (kB1).

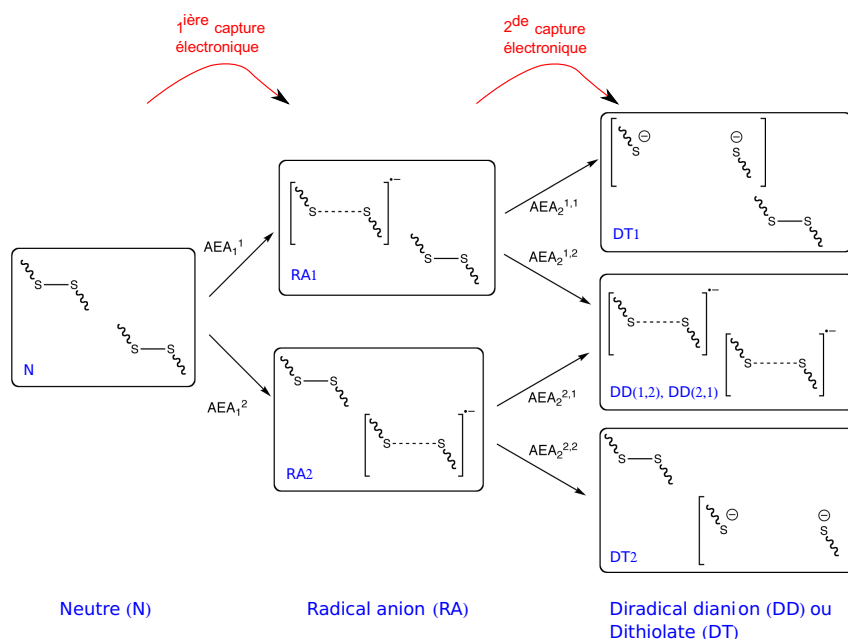


Figure 6.4: Représentation des captures électroniques d'un, puis d'un second électron, sur une molécule comprenant deux ponts disulfures. Les affinités électroniques adiabatiques (AEA) permettent de suivre l'évolution de l'addition d'électron. La capture d'un premier électron est possible sur les deux ponts, formant deux anions radicalaires (RA1 et RA2). L'addition du second électron peut former un dianion diradicalaire (DD(1,2) ou DD(2,1)) ou les dithiolates (DT1 et DT2)

6.1 Description de kalata B1 (kB1)

KALATA B1 (kB1) fait partie de la famille des cyclotides (association de deux mots *cycle* et *peptide*). Ces macrocycles contiennent entre 28 et 37 acides aminés dont six cystéines (Cys) formant deux à deux trois ponts disulfures. À l'heure actuelle, il existe deux structures tertiaires privilégiées pour ces macrocycles : une en ruban de Möbius et une en forme de « bracelet ». Ces deux conformations diffèrent par la présence ou l'absence d'un *twist* (liaison peptidique *cis* située entre le tryptophane (Trp19) et la proline (Pro20)) dans l'ossature de la protéine. kB1 est caractérisé par une structure en ruban de Möbius.

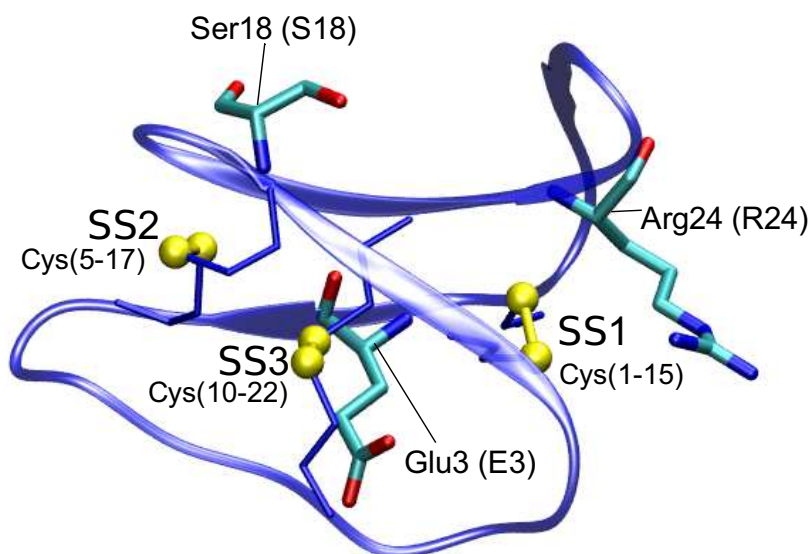


Figure 6.5: Représentation de kalata B1 (code PDB : 1NB1). Les trois ponts disulfures Cys(1-15), Cys(5-17) et Cys(10-22) sont, respectivement, notés SS1, SS2 et SS3. Les trois acides aminés jouant un rôle prépondérant sur l'AEA sont présents sur la figure.

Les caractéristiques de ces cyclotides leur confèrent une importante résistance à différents traitements chimiques, enzymatiques et thermiques qui, en principe, devraient dénaturer une protéine [333]. De tels avantages rendent ces cyclopeptides très intéressants d'un point de vue pharmacologique, notamment grâce à leur petite taille qui facilite le contrôle de la mutagénèse.

Un des objectifs de cette étude sur kB1 est donc d'observer i) sur quelle liaison SS un premier électron ira préférentiellement, ii) de quelle manière la structure secondaire sera affectée par la capture électronique, iii) quels seront les effets électrostatiques et géométriques et iv) quelle sera l'influence de la capture d'un second électron sur l'affinité électronique adiabatique (AEA) ainsi que sur sa structure tertiaire.

Pour répondre à ces questions, ce système sera traité à l'aide de l'approche HLA (*Hydrogen Link Atom*). Les liaisons disulfures étant traitées en QM, il convient néanmoins d'élaborer une stratégie pour décrire correctement l'ensemble des parties QM ainsi que pour traiter toutes les espèces formées suite à la capture d'un électron ou d'un second. Cette stratégie est résumée

dans le point suivant. Nous observerons principalement l'AEA suite à la capture d'un, puis d'un second électron, sur chacune des liaisons possibles, suivant le schéma décrit sur la Figure 6.4. L'AEA sera notamment scindée en effets géométriques et électrostatiques afin de comprendre son comportement d'un point de vue physique.

6.2 Méthodologie employée pour traiter les cyclotides

KALATA B1 sera traité à l'aide de la méthode QM/MM. Les ponts disulfures, sites de la capture électronique, sont décrits à l'aide de la mécanique quantique, précisément les fragments $-\text{CH}_2\text{-S-S-CH}_2-$.

Il est essentiel de choisir adéquatement la méthode quantique parce que celle-ci doit décrire correctement l'énergie des formes neutres et anioniques de la liaison SS permettant de calculer l'AEA. Le traitement de systèmes 2c-3e est délicat de part sa forte corrélation dynamique. Cette dernière sera décrite explicitement à l'aide du second ordre de la théorie des perturbations de Møller-Plesset [22]. Les outils de la DFT introduisent une erreur de *self-interaction* qui se traduit directement par une surestimation de la longueur de la liaison SS ainsi que de l'AEA [307]. La frontière entre la partie QM et la partie MM est définie entre les atomes C_α et C_β de chaque cystéine (Figure 6.6). La partition est décrite sur la Figure 6.6 pour la L,L-Cystine (association de deux L-cystéine avec formation d'une liaison disulfure) et est totalement similaire pour kB1. La méthode utilisée pour décrire correctement cette partition est l'approche de *Hydrogen Link Atom* (HLA) pour plusieurs raisons :

1. Pour caractériser des propriétés liées à la capture électronique, l'approche HLA est préférable à celle du LSCF. La position d'une SLBO trop proche de la liaison SS serait, en effet, délicate et introduirait des artefacts sur l'AEA. L'électron pourrait notamment aller se loger dans les orbitales atomiques vacantes de l'atom Y et non au sein de la liaison SS. Il serait donc nécessaire d'éloigner la SLBO du site de capture électronique (SS) ; mais dans ce cas, le temps computationnel deviendrait très important surtout avec des techniques de type MP2³.
2. D'un point de vue pratique, il aurait été plus simple de placer la liaison frontière au

³Actuellement, MP₂ n'est pas encore implanté pour le LSCF, cependant toutes les équations ont été dérivées [12]

niveau de la liaison peptidique comme nous l'avons fait lors du traitement des protéines fluorescentes. Cependant, le HLA a montré des instabilités de la fonction d'onde pour cette coupure [205]

La partie MM du système est décrite avec le champ de forces Charmm27. Pour minimiser la surpolarisation de la fonction d'onde et pour avoir une charge globale de la partie MM nulle, une redistribution des charges a été employée ⁴. Cette redistribution des charges a été testée sur les systèmes modèles DEtDS (diéthyle disulfure) et sur la L,L-cystine, elle n'influence pas l'AEA [308].

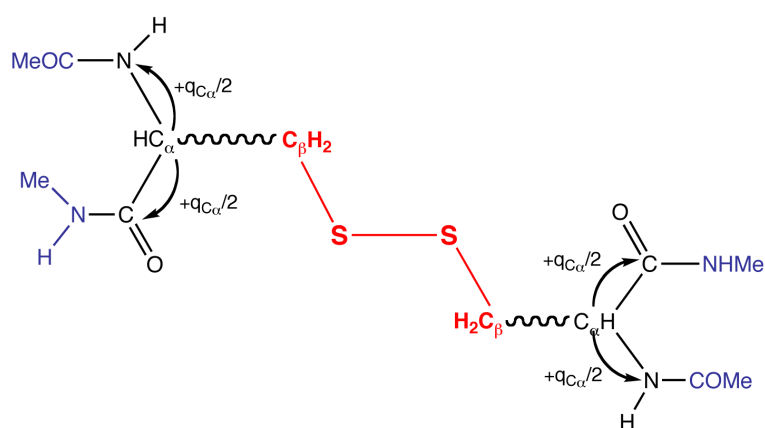


Figure 6.6: Partition QM/MM employée pour décrire l'addition électronique sur les ponts disulfures, illustrée sur le système modèle, la L,L-cystine saturée par NHMe et COMe (en bleu). Ce peptide est considéré comme référence dans cette étude. Les lignes ondulées indiquent l'endroit de la coupure, au niveau de la liaison C_{α} - C_{β} . Les atomes en gras et de couleur rouge seront traités à l'aide de la méthode MP2.

La première partie de l'étude consiste à considérer une première capture électronique sur chacune des trois liaisons. La première AEA observée (AEA_1) est définie comme étant la différence entre l'espèce neutre optimisée (N) et l'espèce anion radicalaire optimisée (RA).

$$AEA_1 = E(N) - E(RA) \quad (6.1)$$

⁴ $q_{C_{\alpha}} = 0$, la charge initiale de +0.07 a été redistribuée sur ses deux atomes voisins MM : N (-0.47e \rightarrow -0.435e) et C (+0.51 \rightarrow +0.545 e)

Les AEA sont relativement sensibles à la base de fonction atomique employée [334], parce qu'elle doit être choisie pour décrire correctement les espèces neutres et anioniques. La base choisie est mixte : 6-31+G** pour les atomes de soufre, et 6-31G* pour tout autre atome. Afin d'avoir des valeurs qui ne traduisent pas les effets de la base, nous avons considéré des valeurs d'AEA relatives (ΔAEA); la référence étant l'AEA de la L,L-Cystine saturée (cf. Figure 6.6). Cette molécule linéaire inclut en effet une liaison SS dont l'AEA ne sera pas influencée par des contraintes géométriques comme la tension de cycle.

$$\Delta\text{AEA} = \text{AEA}_{\text{kB1}} - \text{AEA}_{\text{L,L-cys}} \quad (6.2)$$

Pour observer laquelle des trois liaisons SS est la plus réactive, nous avons optimisé la structure des trois RA possibles : RA1, RA2 et RA3 respectivement si l'électron est capté par la liaison SS1, SS2 et SS3. En partant de la structure neutre de kB1, il a parfois été nécessaire d'altérer l'ordre des orbitales moléculaires de la forme neutre pour avoir le RA souhaité⁵. Ces orbitales virtuelles sont les σ^* des liaisons SS1, SS2 et SS3. Une fois ces trois RA obtenus, nous avons vérifié que la contamination de spin n'était présente que de manière très légère pour chacun des trois RA ($\langle S \rangle^2 \leq 0.77$).

La seconde partie de l'étude sur kB1 est l'addition d'un second électron sur les liaisons disulfures. Le calcul de cette seconde AEA (AEA_2) est plus délicat parce que le second électron peut i) être capturé par un anion radicalaire (RA) formant ainsi un dithiolate (DT) (trois DT différents peuvent donc se former en SS1, SS2 et SS3, ils seront notés respectivement DT1, DT2 et DT3) et ii) se fixer sur une liaison SS neutre, une autre étant au préalable sous forme RA; dans ce dernier cas, il faut examiner les trois possibilités de former ce dianion diradicalaire (DD) : DD(1,2), DD(1,3) et DD(2,3)⁶. Le schéma de ces deux additions d'électron successives est présenté sur la Figure 6.4.

Les structures des DD et DT ont été obtenues à partir de la structure optimisée de chaque RAn (n=1,2,3) et en décalant, si nécessaire les orbitales virtuelles σ^* pour que le DD ou DT souhaité soit formé. Les DD ont été considérés sous forme singulet ($\langle S \rangle^2 = 0.00$) et triplet ($\langle S \rangle^2 = 2.00$), pour les triplets, la contamination de spin n'a jamais excédé 2.03. Dans l'état singulet, nous avons obtenu $\langle S \rangle^2 \simeq 1.00$. Cependant, cette contamination de spin dans l'état

⁵Ce décalage s'effectue avec le mot clé `guess=read,alter` de Gaussian03.

⁶Les combinaisons inverses, DD(2,1), DD(3,1) et DD(3,2) ont aussi été étudiées mais les deux combinaisons sont énergétiquement et géométriquement similaires, nous ne considérerons donc que le premier set de DD(n,k)

Espèce	SS1	SS2	SS3
RA1	0.472, 0.552	0.000, 0.000	0.000, 0.000
RA2	0.000, 0.000	0.527, 0.501	0.001, 0.003
RA3	0.000, 0.000	-0.006, 0.001	0.475, 0.570

Table 6.1: Densité de spin de Mulliken des six atomes de soufre inclus dans les trois ponts disulfures des espèces anions radicalaires du cyclotide.

singulet des DD n'a pas d'influence sur l'AEA, la différence d'énergie entre l'état singulet et l'état triplet étant de moins de 1 kcal.mol⁻¹.

6.3 Première capture électronique

CETTE première partie de l'étude des macrocycles concerne la capture électronique d'un électron au sein du système, c'est-à-dire sur une des liaisons SS. Pour cela, la capture électronique a été considérée sur chacune des trois liaisons SS1, SS2 et SS3 afin d'observer sur quelle liaison irait l'électron préférentiellement. La formation des anions radicalaires RA1, RA2 et RA3 (2c-3e) permettra de décrire la compétition de cette capture entre les trois ponts SS. Les paramètres géométriques et les AEA de chaque pont disulfure se trouvent dans la Table 6.2 pour la forme neutre et pour les trois formes RA.

Tout d'abord, aucun des ponts disulfures ne s'est dissocié suite à l'addition d'un électron ; contrairement à ce que Simons et collaborateurs avaient suggéré dans leur étude théorique sur le DMeDS (diméthyle disulfure) et la L,L-cystine [335,336]. En effet, la densité de spin (Table 6.1) est répartie équitablement pour chaque RA étudié sur les deux atomes de soufre, ce qui confirme la nature de la demi-liaison.

Dans la Table 6.2, les valeurs d'AEA relatives (Δ AEA) sont de +0.52, -0.38 et -0.55 eV respectivement pour les ponts SS1, SS2 et SS3. Le pont disulfure SS1 a la plus grande affinité électronique et sera donc celui qui capturera préférentiellement un premier électron. Expérimentalement, Craik et ses collaborateurs avaient déjà constaté que ce pont SS montrait la plus grande réactivité en réduisant puis en alkylant les ponts SS de kB1 [337]. Les deux autres ponts disulfures sont quant à eux nettement moins enclins à capturer un électron. Il est donc inté-

ressant à ce niveau de pouvoir décrire les raisons de cette sélectivité : est-elle géométrique ou électrostatique ? Nous allons essayer de déduire des règles simples pour prédire l'origine de la réactivité dans un système à multiples liaisons SS.

6.3.1 Décomposition de l'AEA en terme d'effets géométriques et électrostatiques

Il est possible de dissocier les effets géométriques et électrostatiques du ΔAEA assez intuitivement. Comme expliqué dans la section dédiée au QM-MM, la fonction d'onde QM est directement polarisée par l'environnement MM (*Electrostatic Embedding*). Il est donc possible d'éteindre les charges de l'environnement MM. La contribution électrostatique absente (AEA^*), seule la contribution géométrique persiste comme terme majeur (le vdW étant très faible). Le ΔAEA est décomposé en deux contributions, dont on pourra montrer/quantifier leur importance sur le ΔAEA , par leur signe et par leur amplitude.

$$\Delta\text{AEA} = \text{AEA}_{\text{kB1}} - \text{AEA}_{\text{L,L-cys}} \quad (6.3)$$

$$= \underbrace{(\text{AEA}_{\text{kB1}} - \text{AEA}_{\text{kB1}}^* - \text{AEA}_{\text{L,L-cys}} + \text{AEA}_{\text{L,L-cys}}^*)}_{\text{élec}} + \underbrace{\text{AEA}_{\text{kB1}}^* - \text{AEA}_{\text{L,L-cys}}^*}_{\text{géom}} \quad (6.4)$$

$$= \Delta\text{AEA}^{\text{élec}} + \Delta\text{AEA}^{\text{géom}} \quad (6.5)$$

Sur des systèmes protéiques comme kB1, les ponts SS, sont dans un environnement dissymétrique caractérisé par des charges ponctuelles. Cet environnement électrostatique joue un rôle crucial vis-à-vis des propriétés moléculaires. Nous avons vu son importance lors de l'étude sur les spectres d'absorption. Lors de la capture électronique, il est tout à fait raisonnable de penser que si un site proche d'un pont disulfure est chargé négativement (glutamate, aspartate), cela serait défavorable à la capture électronique ; alors que les sites chargés positivement (lysine, arginine, et histidine protonée) favoriseraient cette capture. L'environnement *chargé* influencera certainement la réactivité/sélectivité des sites disulfures notamment les valeurs de l'affinité électronique. Il y a deux acides aminés chargés dans kB1 : le glutamate (Glu3 ou E3) dont le C_α se situe à 4.83 Å du centre de la liaison SS1 et une arginine (Arg24 ou R24) positionnée entre SS2 et SS3. La contribution de la charge négative du glutamate pourrait donc être contre-balançée par l'arginine. Une étude systématique de l'AEA a donc été réalisée en éteignant successivement les charges des chaînes latérales de tous les acides aminés à l'exception des cystéines. Les charges de

Ponts	Indices	Structure			Affinités électroniques				
		$d(\text{S-S})$	$\angle(\text{S-S-C})$	$\tau(\text{C-S-S-C})$	RMSD	AEA_1	ΔAEA_1	$\Delta\text{AEA}_1^{\text{élec}}$	$\Delta\text{AEA}_1^{\text{géom}}$
Cys(1-15)	SS1	N	2.05	103.0, 106.0	95.5	1.58	0.52	0.78	-0.26
	RA	2.75	94.1, 106.8	132.5	0.87				
Cys(5-17)	SS2	N	2.06	102.8, 101.7	76.4	0.68	-0.38	0.31	-0.69
	RA	2.77	97.5, 91.7	70.3	0.23				
Cys(10-22)	SS3	N	2.06	101.3, 102.9	106.6	0.51	-0.55	0.24	-0.79
	RA	2.83	103.6, 104.4	81.7	0.28				
L,L-cystine	—	N	2.05	102.7, 104.7	75.2	1.06			
	—	RA	2.80	95.3, 95.8	66.5	—			

Table 6.2: Les paramètres géométriques tels que les longueurs des liaisons SS (en Å), les angles de valence et angles dièdres (en °) sont repris ainsi que la première affinité électronique adiabatique (AEA_1) associée à chaque liaison SS de kB1. Les valeurs sont données en eV. N et RA réfèrent respectivement aux espèces neutres et aux espèces anions radicalaires. $\Delta\text{AEA}^{\text{élec}}$ et $\Delta\text{AEA}^{\text{géom}}$ sont obtenues à l'aide de l'Eq. 6.3. Le RMSD est défini avec la géométrie de référence de kB1 sous sa forme neutre.

l'ossature (-NH-C α H-C(=O)-) ne sont pas éteintes, évitant ainsi la création d'un dipôle artificiel. La Figure 6.7 montre l'influence des chaînes latérales des acides aminés sur le ΔAEA de chacun des trois ponts disulfures.

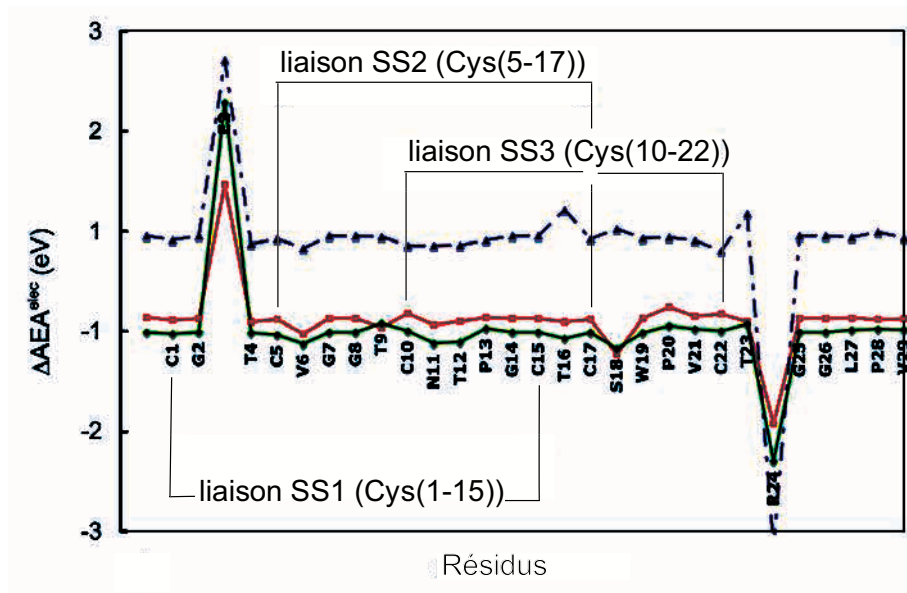


Figure 6.7: $\Delta AEA^{\text{élec}}$ pour chacune des trois liaisons disulfures de kB1, en fonction du résidu pour lequel les charges ponctuelles de la chaîne latérale ont été éteintes (mises à zéro). La ligne discontinue avec des triangles bleus reprend les valeurs pour la liaison la plus réactive (Cys(1-15) ou SS1), celles pour les ponts Cys(5-17) (SS2) et Cys(10-22) (SS3) sont respectivement indiquées par des carrés rouges et des cercles verts.

La Figure 6.7 montre clairement que

1. Comme attendu, le résidu arginine (R24) influence fortement l'AEA de SS1, SS2 et SS3 avec des incréments de, respectivement, -3.11, -1.29 et -1.05 eV sur le ΔAEA . Ces incréments sont directement liés à la distance qui sépare chacune de ces liaisons avec le C α de l'arginine, respectivement 4.83, 8.03 et 10.57 Å.
2. L'autre résidu jouant un rôle sur l'AEA des trois ponts disulfures est le résidu 3, le glutamate (E3). Étant chargé négativement, les incréments sur l'AEA sont défavorables à la capture électronique : +1.76, +1.34 et +2.30 eV. De la même manière que R24, l'intensité de ces incréments est directement en relation avec la distance de chacun des ponts disulfures, respectivement pour SS1, SS2 et SS3 : 7.11, 8.47 et 5.34 Å.

3. Étonnamment, un troisième pic de faible intensité apparaît. Il correspond à la sérine (résidu 18, S18) et favorise légèrement la capture électronique sur les ponts disulfures SS2 (-0.35 eV) et SS3 (-0.16 eV). Ce dernier résidu est bien connu pour faire partie des résidus polaires non chargés (sérine, thréonine, proline, asparagine, glutamine, cystéine).

Ces trois points convergent vers une même conclusion qui suit la loi de Coulomb : plus le résidu est proche et polaire, plus il aura un fort impact sur l'AEA. Les résidus non polaires, eux, n'influencent que très modestement l'AEA électrostatique.

Les contributions géométriques modulant l'AEA sont reprises dans la Table 6.2. La valeur est positive si l'élongation de la liaison SS induite par la capture d'un électron est associée à un relâchement de la structure comme observé pour les cycles aliphatiques [305, 306]. Elle sera, au contraire, négative si l'élongation de la liaison SS est fortement défavorisée, comme c'est le cas dans les structures secondaires de type épingle à cheveux (*hairpin*) [307]. Pour kB1, le signe de $\Delta AEA^{\text{géom}}$ est toujours négatif signifiant que systématiquement la forme anionique est défavorisée géométriquement.

D'un point de vue géométrique, SS1 se différencie des deux autres ponts SS par une contribution géométrique plus de deux fois moins importante : -0.26 eV contre -0.69 et -0.79 eV pour SS2 et SS3. On peut donc penser que la capture électronique modifie la structure en ruban de Möbius, la rendant plus apte à capturer un électron sur le pont SS1. L'élongation de la liaison SS passant de la forme N à la forme RA pour SS1 (+0.70 Å) est similaire à l'élongation de SS2 (+0.71 Å) ; par contre l'angle dièdre a été fortement affecté par la capture, pour SS1, +37.0° contre -6.1° pour SS2 et -24.9° pour SS3. Cela se traduit directement sur la valeur de RMSD, comparant la forme neutre et la forme RAn (n=1,2 ou 3), où un RMSD de 0.87 est obtenu lorsque l'électron est capturé par le pont SS1, valeur à comparer avec le RMSD de SS2 et SS3, respectivement 0.23 et 0.28. Malgré les modifications géométriques au niveau du pont SS1, la structure en ruban de Möbius est maintenue grâce à SS2 et SS3 dont les paramètres géométriques n'ont presque pas été modifiés suite à la capture électronique en SS1. Craik *et al.* ont montré expérimentalement que SS2 et SS3, à eux seuls maintiennent la structure tertiaire compacte caractéristique de la famille des cyclotides. En effet, la mutation des deux résidus cystéines formant SS1 par deux résidus alanines (Cys1Ala et Cys15Ala) n'induit aucune modification quant à la structure en ruban de Möbius. Parmi les trois ponts disulfures, SS1 est donc le pont le plus réactif, ne modifiant aucunement la structure tertiaire ; SS2 et SS3 sont quant à eux cruciaux pour maintenir la structure.

Ensuite, nous nous sommes demandé si cette conclusion restait valable lors de l'addition d'un second électron ? L'ajout de cet électron formera-t-il préférentiellement un DT en SS1 parce que ce pont est le plus réactif [337] ou alors un DD avec un deuxième pont disulfure endommagé. Quel en sera l'impact sur la structure ?

6.4 Capture d'un second électron

Nous allons maintenant discuter l'addition d'un second électron avec pour but d'observer sa localisation préférentielle ainsi que les modifications géométriques induites. La Table 6.3 reprend les paramètres géométriques et les AEA de neuf espèces : trois correspondent aux dithiolates (DT) formés sur chacune des liaisons SS ; les six autres espèces sont des dianions diradicalaires (DD) – état singulet et triplet – où le premier électron est placé sur SS n ($n=1, 2$ ou 3) et le second électron sur SS k ($k=1, 2$ ou 3 mais avec $k \neq n$). SS1 étant le plus réactif lors de la capture du premier électron, nous pouvons penser que le DT1 devrait se former préférentiellement sans modifier fortement la structure de kB1. Cela reste néanmoins à être vérifié d'un point de vue théorique.

La Figure 6.8 reprend schématiquement les résultats de l'attachement successif d'un et de deux électrons. De manière assez prévisible, nous observons clairement que la formation de DT1 est fortement favorisée, avec une énergie de 1.36 eV plus stable que le DT3 (second dianion préférentiel).

Espèces			Structures			Énergie	
2S+1	Liaison(s)		$d(\text{S-S})$	$\angle(\text{S-S-C})$	$\tau(\text{C-S-S-C})$	RMSD	ΔE
Dithiolates							
DT1	1	SS1	6.68	70.9, 120.5	146.8	1.12	2.47
DT2	1	SS2	4.70	76.2, 83.1	78.2	0.57	-0.25
DT3	1	SS3	5.77	39.6, 109.3	114.2	1.44	1.09
Dianions diradicalaires							
DD(1,2)	1	SS1	2.76	92.1, 99.3	128.8	0.83	0.88
		SS2	2.76	91.2, 96.4	68.8		
	3	SS1	2.76	92.1, 99.3	128.8	0.83	0.88
		SS2	2.76	91.2, 96.4	68.8		
DD(1,3)	1	SS1	4.97	81.7, 113.8	135.0	1.05	0.40
		SS3	2.75	98.1, 93.6	120.6		
	3	SS1	5.97	94.1, 106.0	120.2	1.51	0.40
		SS3	2.76	90.6, 90.6	125.3		
DD(2,3)	1	SS2	2.76	94.9, 92.4	62.3	0.64	-0.50
		SS3	2.71	50.3, 91.4	118.3		
	3	SS2	2.76	94.9, 92.4	62.3	0.64	-0.51
		SS3	2.72	94.2, 91.4	118.4		

Table 6.3: Paramètres géométriques et énergies relatives (en eV) des formes dianioniques de kB1 – dithiolates (DT) ou dianions diradicalaires (DD). Les énergies relatives ΔE sont calculées par rapport à l'énergie du composé neutre (N) — cf. Figure 6.8

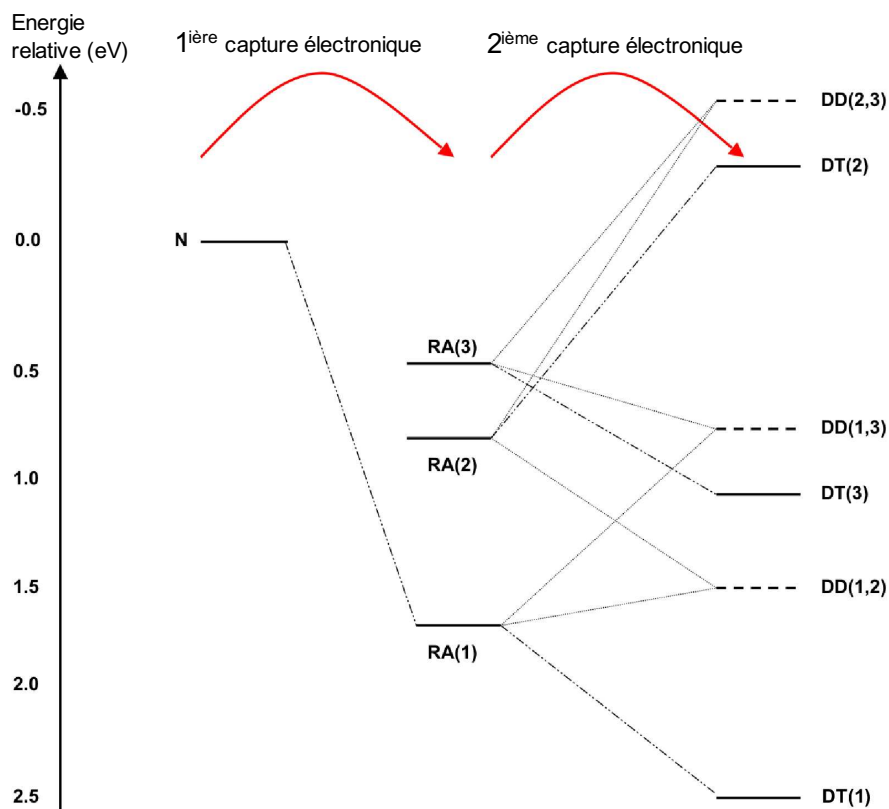


Figure 6.8: Les énergies électroniques relatives (en eV) des structures de kB1 après la capture successive d'électrons. DT (en ligne continue) sont les dithiolates, et DD (lignes discontinues) correspondent aux dianions diradicalaires avec des liaisons 2S-3e. Le système neutre (N) est pris comme référence.

Géométriquement, nous remarquons que la distance entre les deux atomes de soufre est relativement flexible, allant de 2.75 Å pour les formes RA jusqu'à 5 Å pour les formes DT. Nous tenons à préciser, à nouveau, que pour les formes DD, les densités de spin de Mulliken sont toujours réparties similairement entre les deux atomes de soufre, excepté pour le DD(1,3) (Table 6.4).

Tout d'abord, nous allons considérer les espèces dithiolates (DT), le DT1 est plus stable de 1.38 et 2.72 eV par rapport à DT3 et DT2. En observant les paramètres géométriques, l'ordre lié à l'AEA (DT1, DT2 et DT3) suit l'ordre pour la distance inter-soufre, respectivement 6.68, 5.77 et 4.70 Å. Le même ordre est aussi respecté pour les angles dièdres. En comparant la structure de DT1 avec celle de N et RA1 (cf. Figure 6.9), la chaîne latérale du résidu chargé positivement

Espèces	2S+1	SS1	SS2	SS3
DD(1,2)	1	-0.506/-0.527	0.546/ 0.483	-0.001/ 0.003
	3	0.505/ 0.528	0.546/ 0.482	-0.002/ 0.005
DD(1,3)	1	-0.002/-1.020	-0.004/ 0.000	0.435/ 0.607
	3	-0.000/ 1.021	-0.005/ 0.001	0.371/ 0.670
DD(3,2)	1	-0.000/-0.000	-0.618/-0.414	0.458/ 0.574
	3	0.000/-0.001	0.625/ 0.404	0.434/ 0.603

Table 6.4: Densité de spin de Mulliken des six atomes de soufre inclus dans les trois ponts disulfures

(arginine 24, R24) est réorientée, avec son extrémité chargée en direction de la liaison SS1 et plus particulièrement, vers le soufre de la cystéine 15 (Cys15).

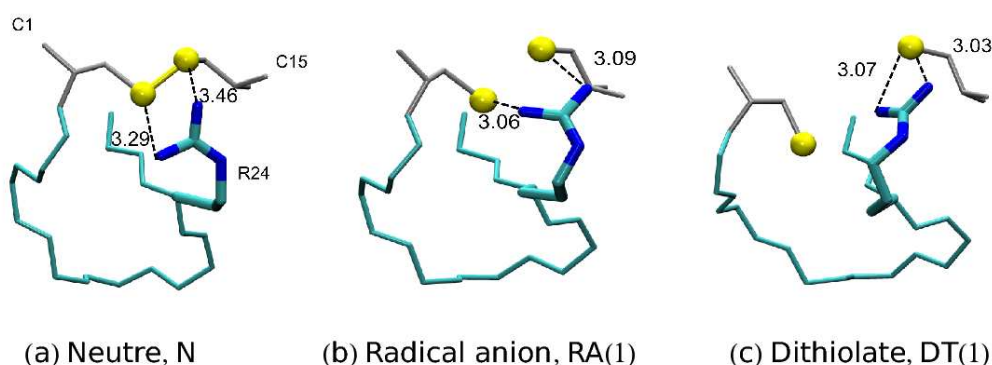


Figure 6.9: Structures optimisées de kalata B1, centrée sur la liaison SS1

Cette rotation de l'arginine s'accompagne de la formation d'un nouveau réseau de liaisons hydrogènes autour de la cystéine 15. Ces deux effets contre-balaçent la forte répulsion entre les deux atomes de soufre. Remarquablement, le DT3 est fortement stabilisé par rapport au RA3. La différence est à nouveau géométrique. Sous la forme DT3, il a y une rotation au niveau des acides aminés Pro13-Gly14 du coude β suite à la forte élongation entre les deux atomes de soufre du pont SS3 (3.71 Å) — cf. Figure 6.10. La rotation de la proline étant d'environ 80°, le RMSD de la DT3 est directement affecté par cette modification par rapport à la structure neutre. Ce mouvement entraîne une rotation de la chaîne latérale $-\text{CH}_2-\text{S}^-$ de la Cys10 de la liaison SS3.

L'atome de soufre s'oriente complètement à l'opposé de l'autre soufre de la Cys22.

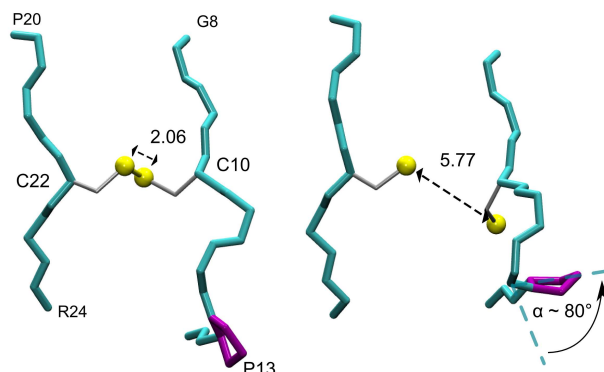


Figure 6.10: Structure optimisée de kalata B1 dans sa forme neutre, à gauche, et dithiolate (DT3), à droite, centrée sur le pont SS3. L'ossature de la protéine est décrite par les tubes bleus tandis que la proline et les ponts SS3 sont respectivement de couleur mauve et jaune. Les distances sont en Å. La rotation de la proline (P13) d'un angle de 80° est mise en évidence.

Concernant les dianions diradicalaires (DD), le plus stable est le DD(1,2). Cela pouvait être attendu suite aux résultats de la première capture électronique où les RA1 et RA2 étaient les plus réactifs (cf. Figure 6.8). Dans le cas du DD(1,3), la seconde addition électronique sur la liaison SS3 induit une dissociation de la liaison du RA1, la longueur de la liaison passant de 2.75 à 4.97 Å. L'arginine, initialement équidistante aux deux atomes de soufre du pont SS1, stabilise le thiolate (Cys1), situation relativement similaire à la Figure (6.9c). Cette évolution de la dissociation d'une liaison SS suite à un excès électronique a été étudiée expérimentalement [338] et théoriquement [336].

Il est nécessaire de garder à l'esprit que les formes DT sont énergétiquement favorables à la seconde addition électronique, même si tous les cas ont été explicités. Expérimentalement, cela n'est sans doute pas observable. Suite à une irradiation d'un cristal, de nombreux effets de « packing » empêcheraient la structure cristalline de se relaxer. Cette étude éclaire la compétitivité entre les trois liaisons SS suite à l'ajout d'un second électron.

6.5 Conclusions du travail sur kalata B1

Ce travail s'inscrit dans la suite logique de plusieurs publications réalisées sur de petits systèmes aliphatiques contenant un pont disulfure. Ces travaux montraient que l'AEA dépendait de la tension de cycle, ou frustration topologique (tétrapeptide cyclique), de l'environnement électrostatique (tétrapeptide lié à une hélice- α). Ce travail est le premier réalisé sur une macromolécule de 29 acides aminés contenant trois ponts disulfures. Il s'agissait donc de mettre en évidence les facteurs qui gouvernaient la compétition entre ces trois ponts disulfures quant à l'addition d'un, puis d'un second électron. Une étude systématique de l'AEA et plus précisément du Δ AEA a été réalisée ; notamment en décomposant cette dernière en deux contributions : électrostatique et géométrique. La contribution électrostatique nous a montré que le Δ AEA était influencé lorsque les ponts disulfures se situaient près d'acides aminés chargés (arginine et glutamate). Leur influence sur le Δ AEA est directement proportionnelle à la distance qu'ils possèdent avec les ponts SS. En effet, plus la distance est courte, plus l'AEA en sera affectée. Le résidu chargé positivement favorisent la capture électronique. SS1 est le pont le plus réactif, étant capable de capturer un électron sans grande réorganisation géométrique autour de lui. SS2 et SS3 sont, quant à eux, responsables du maintien de la structure en ruban de Möbius. La seconde capture électronique nous a appris que la liaison SS1 capturait aussi de manière plus favorable un second électron. Le glutamate étant tourné préférentiellement vers la Cys15. Le DT3 est plus réactif que le DT2 (inverse lors de l'addition d'un électron). Lors de la formation du DT3, un mouvement de rotation du coude β avoisinant est observé, favorisant la stabilisation de la Cys10.

Chapitre 7

Les ionisations de cœur

CE dernier chapitre est consacré à un troisième phénomène électronique, qui est l'ionisation de cœur. Ce travail a été réalisé dans le cadre d'une collaboration avec le professeur Vitaly N. Glushkov (Dniepropetrovsk, Ukraine) et a été soutenu dans le cadre du programme Hubert Curien *Dnipro* 19735 QK. Vitaly N. Gluchkov a, en effet, développé une méthode relativement similaire à l'approche LSCF pour orthogonaliser les orbitales localisées et les orbitales variationnelles entre elles.

Il existe deux processus responsables de l'émission d'électron(s) venant d'atomes : les processus d'excitation et de relaxation électronique. Cette partie ne concerne que l'excitation ou ionisation d'électron de cœur de molécules. L'envoi de lumière monochromatique est adéquat pour ioniser sélectivement des électrons de cœur. Il est ainsi possible d'exciter des électrons 1s, 2s ou 2p d'un élément ou d'une molécule de manière spécifique, et ce parce que les énergies ou potentiels d'ionisation sont très distincts, de l'ordre de 10 à quelques centaines électron-Volt (eV). À titre d'exemple, la Table 7.1 ci-dessous reprend les énergies d'ionisation des atomes les plus courants. Plus intéressant encore est qu'il s'avère possible expérimentalement de distinguer les effets de l'environnement d'un même élément. Par exemple, l'énergie d'ionisation du carbone dans un cycle aromatique sera différente de celle d'un carbone d'un groupement fonctionnel comme CF_3 .

Un état excité ou ionisé de cœur réfère à la situation où un électron de cœur d'un atome ou d'une molécule a été excité ou ionisé, laissant une vacance dans l'orbitale de cœur (souvent appelée « trou de cœur »). L'énergie d'ionisation de cœur d'un atome ou d'une molécule est l'énergie qu'il faut fournir à un atome neutre pour lui arracher un électron provenant du cœur. Ces

Éléments	Orbitales	E_{ion} (eV)
Carbone	1s	284
Azote	1s	410
Oxygène	1s	543
Soufre	2s	231
	2p	163

Table 7.1: Énergies d'ionisation de quelques éléments courants [339].

états excités de cœur sont étudiés par absorption de rayons X (XA, *X-ray Absorption*) qui utilise le rayonnement synchrotron. Le maximum d'absorption correspond à l'éjection d'un électron de la couche de cœur. Les spectres XA sont caractérisés par deux régions : la région étendue (EXAFS, *extended X-ray Absorption Fine Structure*) et la région NEXAFS ou XANES (*Near-Edge X-ray Absorption Fine Structure* ou *X-ray Absorption Near-Edge Structure*). La région autour (± 10 eV) du maximum d'absorption est celle du NEXAFS ou XANES, caractéristique de l'état électronique de l'élément. La zone se trouvant après le maximum d'absorption principal et s'étendant sur quelques centaines d'eV correspond à la région de l'EXAFS.

La spectroscopie NEXAFS est un outil très puissant. En effet, la région NEXAFS dépend de la géométrie, du nombre de coordination et de l'ordre de liaison. Elle est particulièrement sensible à l'environnement local de l'atome ionisé. De nombreuses applications de cette technique sont utilisées, notamment pour déterminer des structures électroniques ou l'orientation de molécules dans l'état solide ou adsorbée sur des surfaces [340]. Il est possible d'attribuer des signatures caractéristiques des spectres NEXAFS aux groupements fonctionnels et dans certains cas des liaisons spécifiques comme la liaison peptidique. Le spectre total d'une macromolécule pourra alors être considéré comme une combinaison linéaire des spectres des structures individuelles constituant la macromolécule. Cette approche est appelée *Building Block Principle* (BBP) et est le point de départ de l'interprétation de spectres expérimentaux de molécules très complexes comme des protéines. Cependant, cette approche a aussi ses limites. La délocalisation des charges électroniques aux travers des différents groupements fonctionnels entraîne une modification des orbitales moléculaires combinant les propriétés individuelles de ces groupements. Dans ces cas, les signatures des spectres NEXAFS sont fortement modifiées. Dès lors, la plupart des études

expérimentales sont réalisées sur de petits systèmes organiques [341,342] et plus récemment sur des acides aminés ou chaînes d'acides aminés [343–347]. L'étude d'ionisation de cœur des chaînes polypeptidiques a fait l'objet de quelques études expérimentales afin de pouvoir réaliser une carte chimique de leurs caractéristiques en fonction de leur environnement. Des approches théoriques peuvent ainsi venir en aide pour caractériser et comprendre les pics d'absorption des spectres expérimentaux de systèmes complexes pour lesquels le principe BBP n'est pas toujours efficace.

En chimie théorique, selon le théorème de Koopmans, l'énergie d'ionisation pour une orbitale donnée est égale à l'opposé de la valeur propre de cette orbitale dans la théorie de HF. Cette méthode très simple pour estimer le potentiel d'ionisation fonctionne assez bien tant qu'il s'agit d'une ionisation de valence [348]. Cependant, elle ne prend en compte ni la relaxation orbitalaire (redistribution de la densité électronique dans les orbitales suite à l'éjection d'un électron), ni la corrélation électronique du système. Dès qu'un électron est enlevé d'une orbitale de cœur, les effets de relaxation orbitalaire deviennent trop importants et le potentiel d'ionisation est surestimé.

Dans les années soixante-dix, Bagus a développé une méthode qui prend en compte la relaxation orbitalaire, nommée Δ SCF, pour calculer les potentiels d'ionisation de cœur [349,350]. Selon cette approche, le potentiel d'ionisation est défini comme la différence entre l'énergie HF d'un état ionisé et l'énergie correspondante de l'état fondamental d'une molécule neutre. Cette fois, les orbitales de l'état ionisé se relaxent. Ce dernier est, de cette manière, plus stabilisé que par la méthode de Koopmans. Le potentiel d'ionisation Δ SCF sera donc inférieur au potentiel d'ionisation obtenu à l'aide du théorème de Koopmans. Le calcul d'énergie de l'état ionisé est effectué en utilisant l'orbitale de cœur d'intérêt (construite dans l'état fondamental), laquelle restera simplement occupée, à la place de l'être doublement. Plusieurs méthodes de type Δ ont ensuite émergé, comme le Δ KS [351]. Cependant, avec ces approches, l'état fondamental n'est pas orthogonal aux états ionisés/excités. Plusieurs techniques ont été proposées pour assurer l'orthogonalité, cependant elles ne s'appliquent, en général, qu'à des molécules organiques et non à des systèmes macromoléculaires [352–354].

En 2002, N. Ferré et X. Assfeld utilisent une approche basée sur le LSCF pour traiter correctement les ionisations de cœur de polymères (le polyéthylène et le polytétrachloroéthylène) et d'un système biologique (la crambine) [355]. La liaison covalente entre la partie QM et la partie MM est matérialisée par une SLBO doublement occupée. Cette SLBO reste gelée lors de

l'optimisation de la fonction d'onde. Le trou de cœur est quant à lui décrit à l'aide de l'orbitale de cœur d'intérêt (CO, *Core Orbital*) vacante. Cette dernière, issue du calcul sur le neutre, sera gelée lors de l'optimisation de la fonction d'onde de l'état ionisé, assurant l'orthogonalité entre les états électroniques. Le système ionisé est donc constitué d'orbitales variationnelles et d'orbitales gelées (SLBO et CO)

Par conséquent, il suffit de réaliser deux calculs d'énergie adiabatique : un calcul de l'énergie de l'état fondamental du système neutre, $E(\Psi)$, où Ψ est le déterminant de l'état fondamental et un calcul de l'énergie de l'état électronique où l'orbitale de cœur est gelée et vacante ($E(\Psi')$). L'orbitale de cœur d'intérêt est localisée dans l'état fondamental du système par un des critères de localisation (cf. section 4.3.2).

$$E_{\text{ion}} = E(\Psi') - E(\Psi) \quad (7.1)$$

Quelques années plus tard, P.-F. Loos et X. Assfeld ont apporté une amélioration de la méthode [356]. Dans le cas des ionisations de cœur d'un atome appartenant à un système biologique décrit en QM-MM, il est donc nécessaire de définir une ou plusieurs orbitale(s) localisée(s) pour décrire l'état électronique ionisé : une CO et une SLBO. Dans l'approche LSCF, la méthode Löwdin est utilisée pour orthogonaliser les OL. Cependant, il faut signaler ici que P.-F. Loos et X. Assfeld ont mis en évidence que la situation pourrait devenir dramatique si elle était appliquée simultanément sur des orbitales localisées occupées (SLBO) et vides (CO). En effet, cette orthogonalisation des OL fonctionne très bien lorsque celles-ci sont toutes occupées, mélangeant donc des orbitales de même nature. Lors du calcul de l'énergie de l'état électronique ionisé, des OL occupées (SLBO) sont en présence de CO vacante (trou de cœur). Dans ce cas, l'orthogonalisation de Löwdin a pour effet de peupler partiellement les orbitales vides ! Cette situation sera plus importante dans le cas où une SLBO est très proche d'une CO et que le recouvrement entre ces orbitales est grand. P.-F. Loos a donc proposé d'orthogonaliser les OL vides avec la méthode de Löwdin et ensuite, d'orthogonaliser les OL occupées à l'aide de la méthode de Gram-Schmidt [13]. Cette orthogonalisation par bloc ne sera pas utilisée par la suite parce que les différences entre énergie d'ionisation par la méthode Löwdin et la méthode Löwdin+Gram-Schmidt sont seulement de 0.05 eV sur un atome définissant la SLBO et possédant une CO. De plus, nous avons veillé à ce que la CO ne soit pas sur un des atomes définissant la SLBO lors des ionisations de cœur du carbone de systèmes biologiques.

Dans ce chapitre, nous allons décrire une nouvelle approche développée par V. Glushkov

qui permet d'orthogonaliser les OL aux orbitales variationnelles : la méthode de la projection asymptotique (AP, *Asymptotic Projection*) [357–359]. Les sections suivantes seront dédiées à la validation de cette approche sur un jeu de molécules organiques. Ensuite, nous évaluerons les ionisations de cœur d'acides aminés et finalement, nous observerons les variations de l'ionisation suite à la polymérisation de la glycine, ainsi que les effets de l'environnement lorsque la glycine est un résidu constitutif de la Serum-Albumine Humaine (HSA, *Human Serin Albumin*).

7.1 L'approche de la projection asymptotique

POUR rappel, il est possible de localiser et de geler des orbitales en LSCF et, qui plus est, une contrainte d'orthogonalité leur est imposée. Cette contrainte est réalisée en trois étapes successives :

1. L'ensemble des orbitales gelées sont orthogonalisées entre elles par la méthode de Löwdin, Eq. (4.12).
2. Chaque fonction de base atomique $|\mu\rangle$ est projetée dans le sous-espace des complémentaires des OL (OLO), Eq. (4.13).
3. Il est nécessaire d'enlever les L dépendances linéaires engendrées par les L OL et présentes dans l'étape 2, Eq. (4.16).

Les deux dernières étapes sont rassemblées en une matrice $\mathbf{B} = \mathbf{M} \cdot \mathbf{X}$ (Eq. (4.17)).

La projection asymptotique est une méthode qui traite aussi des contraintes telles que des orbitales gelées de manière totalement orthogonales aux orbitales variationnelles. Cependant, la procédure est légèrement simplifiée. Cette méthode a été utilisée auparavant pour décrire les ionisations de cœur de petits systèmes (H_2 , CO , FH , H_2O , ...). En l'implantant dans notre version modifiée de Gaussian [360], nous allons pouvoir réaliser des applications sur de plus grands systèmes et l'employer sur des systèmes protéiques traités en QM-MM.

Les orbitales localisées (OL) sont toujours décrites comme une combinaison linéaire des orbitales atomiques :

$$|l_i\rangle = \sum_{\lambda}^N a_{\lambda i} |\lambda\rangle \quad (7.2)$$

Plaçons nous dans le cadre des ionisations de cœur, un électron est enlevé de la spin-orbitale de cœur β $|\psi_{0k}^{\beta}\rangle$ dans l'état ionisé. Nous nous trouvons donc dans le formalisme *Unrestricted*. La

première contrainte à imposer est que les orbitales variationnelles (ψ_j) soient orthogonales aux OL :

$$\begin{aligned} \langle \psi_j^\alpha | l_i \rangle &= 0 & j = 1, 2, 3, \dots, n^\alpha \\ \langle \psi_j^\beta | l_i \rangle &= 0 & j = 1, 2, 3, \dots, n^\beta - 1 \end{aligned} \quad (7.3)$$

Deuxièmement, toutes les spin-orbitales β de l'état ionisé k ($\langle \psi_j^\beta |$) doivent être orthogonales à celle de l'état fondamental ($|\psi_{0k}^\beta\rangle$) :

$$\langle \psi_j^\beta | \psi_{0k}^\beta \rangle = 0 \quad j = 1, 2, 3, \dots, n^\beta - 1 \quad (7.4)$$

Ces deux contraintes peuvent être réécrites à l'aide d'un projecteur P_r défini comme $\sum_i^N |l_i\rangle\langle l_i|$.

$$\begin{aligned} \sum_j^{n^\alpha} \langle \psi_j^\alpha | P_r | \psi_j^\alpha \rangle &= 0 \\ \sum_j^{n^\beta-1} \langle \psi_j^\beta | P_r | \psi_j^\beta \rangle &= 0 \end{aligned} \quad (7.5)$$

et

$$\sum_j^{n^\beta-1} \langle \psi_j^\beta | P_k | \psi_{0k}^\beta \rangle = 0 \quad (7.6)$$

avec le projecteur $P_k = |\psi_{0k}\rangle\langle \psi_{0k}|$. Suite à cela, on peut décrire un Lagrangien qui prend en compte ces contraintes :

$$L = \langle \Psi | H | \Psi \rangle + \lambda_1 \left(\sum_j^{n^\alpha} \langle \psi_j^\alpha | P_r | \psi_j^\alpha \rangle + \sum_j^{n^\beta-1} \langle \psi_j^\beta | P_r | \psi_j^\beta \rangle \right) + \lambda_2 \sum_j^{n^\beta-1} \langle \psi_j^\beta | P_{0k} | \psi_{0k}^\beta \rangle \quad (7.7)$$

Il suffit maintenant de résoudre le système :

$$\partial L = 0 \quad (7.8)$$

où l'approximation algébrique $|\psi_i\rangle = P|\psi_i\rangle$ est utilisée pour définir la variation des orbitales atomiques :

$$|\delta\psi_j\rangle = P_n|\delta\psi_j\rangle + (P - P_n)|\psi_j\rangle \quad (7.9)$$

P est un orthoprojecteur¹ défini par un jeu de bases de dimension M et P_n est défini dans le sous-espace des orbitales occupées. Le premier terme de cette dernière expression n'induit

¹Un orthoprojecteur permet de réduire un espace de dimension infinie X à un espace de dimension finie M compris dans l'espace X selon la relation $M = PX$

aucun changement sur l'énergie totale du système vu qu'il est invariant à la transformation de l'orthogonalité des orbitales ayant un même spin. Les modifications de l'énergie proviendront principalement du second terme. L'Eq. (7.8) à l'aide des Eq. (7.7) et (7.9) devient après quelques manipulations :

$$\sum_j^{n^\alpha} \langle \delta\psi_j^\alpha | (P - P_n) (F^\alpha + \lambda_1 P_{lr}) | \psi_j^\alpha \rangle + \sum_j^{n^\beta} \langle \delta\psi_j^\beta | (P - P_n) (F^\beta + \lambda_1 P_{lr} + \lambda_2 P_{0k}) | \psi_j^\beta \rangle + \text{c.c.} = 0$$

où F^α et F^β sont les opérateurs UHF habituels. À partir de cette équation, on peut en tirer le système d'équations aux valeurs propres suivant :

$$P (F^\alpha + \lambda_1 P_{lr}^\alpha) | \psi_j^\alpha \rangle = \epsilon_j | \psi_j^\alpha \rangle \quad (7.10)$$

$$P (F^\beta + \lambda_1 P_{lr}^\beta + \lambda_2 P_{0k}^\beta) | \psi_j^\beta \rangle = \epsilon_j | \psi_j^\beta \rangle \quad (7.11)$$

Un Fockien modifié (F_{mod}) peut ainsi être défini pour les spin-orbites α et β :

$$F_{\text{mod}}^\alpha = F^\alpha + \lambda_1 P_{lr}^\alpha \quad (7.12)$$

$$F_{\text{mod}}^\beta = F^\beta + \lambda_1 P_{lr}^\beta + \lambda_2 P_{0k}^\beta \quad (7.13)$$

Les multiplicateurs de Lagrange, outils purement mathématiques, ne sont toutefois pas encore déterminés. La méthode utilisée pour résoudre ce problème sous contrainte est la projection asymptotique dont voici le théorème [361] :

If it is assumed that the normalized vector $|u\rangle$ is not an eigenstate of FPF operator, then the constraint vector $|u\rangle$ tends to an eigenvector of the modified operator

$$F_{\text{mod}} = P(F + \lambda P_u)P \text{ if and only if } \lambda \rightarrow \infty, P_u = |u\rangle\langle u|.$$

Ce théorème peut-être démontré d'une manière rapide et efficace à l'aide de l'exemple suivant :

Soit le vecteur $|u^1\rangle$ défini par les coefficients d'une orbitale localisée, multipliés par l'opérateur modifié F_{mod} :

$$\begin{pmatrix} (F_{11} + \lambda_1) & F_{12} & \cdots & F_{1n} \\ F_{21} & F_{22} & \cdots & F_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ F_{n1} & \cdots & \cdots & F_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{pmatrix} = \lambda_1 \begin{pmatrix} \frac{F_{11}}{\lambda_1} + 1 \\ \frac{F_{21}}{\lambda_1} \\ \cdots \\ \frac{F_{n1}}{\lambda_1} \end{pmatrix} \quad (7.14)$$

avec $\lambda_1 = \lambda \langle \mu | P | \mu \rangle$. Cette équation montre clairement, que le vecteur $|u^1\rangle$ et donc l'orbitale localisée sera bien un vecteur propre de F_{mod} si et seulement si λ tend vers l'infini (souvent de l'ordre de 10^6).

7.1.1 L'implantation

Nous avons réalisé l'implantation de l'approche AP dans la version de Gaussian03 déjà modifiée pour le LSCF. Sachant que le projecteur P_u peut être défini comme suit avec $|\nu\rangle$ et $\langle\mu|$ des orbitales atomiques :

$$P_u = \sum_i^N |l_i\rangle \langle l_i| \quad (7.15)$$

$$= \sum_i c_{\mu i} c_{\nu i} |\nu\rangle \langle \mu|, \quad (7.16)$$

et les éléments du Fockien deviennent sous forme matricelle :

$$F_{pq \text{ mod}} = F_{pq} + \lambda \langle \phi_p | P_u | \phi_q \rangle \quad (7.17)$$

$$= F_{pq} + \lambda \left(\sum_i c_{\mu i} c_{\nu i} \langle \phi_p | \nu \rangle \langle \mu | \phi_q \rangle \right) \quad (7.18)$$

Finalement, le terme qui faisait intervenir le projecteur est calculé comme le recouvrement orbitalaire entre les OA.

$$F_{pq \text{ mod}} = F_{pq} + \lambda \sum_{\nu, \mu} c_{\mu i} c_{\nu i} S_{\phi_p \nu} S_{\mu \phi_q} \quad (7.19)$$

Ce recouvrement est aussi calculé en LSCF pour l'étape 2. (cf. Eq. (4.15)). Le point commun entre l'approche LSCF et AP est, en fait, l'utilisation d'un même projecteur mais utilisé de manière différente. En effet, en LSCF, un projecteur défini par les orbitales localisées modifie le jeu d'orbitales atomiques pour les rendre orthogonales aux OL. Dans l'AP, ce même projecteur est simplement ajouté au Fockien.

7.2 Validation de la projection asymptotique

A APRÈS l'implantation de l'AP dans la version B.05 de Gaussian03 [360], nous avons testé et validé l'approche en calculant les énergies d'ionisation de l'orbitale 1s du carbone d'un ensemble de petites molécules organiques (les Tables 7.2 et 7.3). Ces molécules organiques ont

été utilisées préalablement dans une étude de P.-F. Loos à l'aide de la méthode LSCF [356]. Ceci permettra de valider l'AP en la comparant à l'approche LSCF. Deux groupes constituent ces molécules organiques : un groupe d'hydrocarbure et un groupe de molécules contenant des hétéroatomes.

Dans le travail préliminaire de P. F. Loos, le choix de la base et de la fonctionnelle a été étudié respectivement sur le méthane et sur l'ensemble des molécules organiques; nous reprendrons le même jeu de base optimal obtenu (6-311++G(d,p)). Les géométries ont été optimisées au niveau de théorie B3LYP/6-311++G(d,p). La précision sur le calcul des énergies des deux états électroniques est primordiale et a été jugée optimale avec la fonctionnelle hybride PBE0 ². Quant aux critères de localisation de l'orbitale 1s de l'atome d'intérêt, deux critères ont été testés PM (Pipek-Mezey) et BF (Boys-Föster) sur la série de molécules organiques. La déviation des énergies d'ionisation par rapport à l'expérience se trouve entre -1.78 eV et 0.63 eV avec le critère de localisation de PM contre seulement 0.22 eV à 0.81 eV avec le critère de BF. Le niveau de théorie utilisé finalement est donc BF-PBE0/6-311++G(d,p)//B3LYP/6-311++G(d,p).

7.2.1 Ionisation de cœur de molécules organiques

Le premier test de validité de l'AP a été effectué en comparant les ionisations de cœur de l'orbitale 1s de l'atome de carbone du méthane. Le potentiel d'ionisation de référence est la valeur expérimentale de 290.84 eV ainsi que celle obtenue par l'approche LSCF qui nous servira de valeur théorique de référence. La Table 7.2 reprend ces ionisations de cœur obtenues par les deux approches (LSCF et AP). Nous avons posé une ou deux contraintes. Une contrainte signifie qu'une seule orbitale de cœur 1s du carbone est gelée, celle-ci sera vacante ³. Lorsque nous avons imposé deux contraintes, une orbitale de cœur vacante (pas d'électron β) et une seconde orbitale de cœur occupée par un électron α sont gelées. Nous avons nommé ces deux schémas respectivement 1 CO et 2 CO. En LSCF, l'énergie d'ionisation de l'orbitale 1s du carbone est de 290.78 eV avec 2 CO et de 289.65 eV avec 1 CO. Cela s'explique par le fait qu'il y a moins d'orbitales variationnelles lors du calcul à deux contraintes (une de moins). Nous verrons que

²Le calcul de l'énergie d'ionisation de l'orbitale 1s des atomes de carbone du pentane a été évalué avec les deux fonctionnelles hybrides PBE0 et B3LYP. Un RMSD total par rapport à l'expérience de 0.54 eV a été obtenu avec B3LYP alors qu'en PBE0 cette déviation par rapport à l'expérience n'est que de 0.16 eV.

³Dans le programme Gaussian, les électrons α remplissent en premier les orbitales suivi des électrons β , il faut donc toujours vider l'orbitale 1s de l'électron β .

cette déviation est constante.

Passons maintenant à l'approche AP, cette première évaluation des énergies d'ionisation sur le méthane a pour objectif principal de déterminer le coefficient de Lagrange (λ) optimal. Nous pouvons observer que plus λ est grand, plus l'énergie d'ionisation tend vers la valeur obtenue en LSCF, ce qui était attendu. Lorsque $\lambda = 10^6$, l'écart par rapport l'énergie d'ionisation de cœur ne diffère qu'à la quatrième décimale, ce qui est amplement satisfaisant. Les erreurs entre AP et LSCF sont donc totalement similaires à l'erreur numérique près. Dans la suite, nous avons donc pris un $\lambda = 10^6$ afin que les contraintes d'orthogonalité soient respectées. D'un point de

Méthode	E_{ion} (eV)		
	2 CO	1 CO	
LSCF	290.776351	289.652904	
AP	$\lambda = 10^6$	290.776280	289.652888
	$\lambda = 10^5$	290.775638	289.652745
	$\lambda = 10^4$	290.769243	289.651314
	$\lambda = 10^3$	–	289.636783
Expérience	290.84		

Table 7.2: Comparaison entre l'énergie d'ionisation obtenue par la méthode LSCF et l'AP. Le coefficient de Lagrange de l'approche AP varie entre 10^3 et 10^6 .

vue temps de calcul entre les deux méthodes, nous observons des temps totalement similaires.

Ensuite, nous avons évalué les énergies d'ionisations de l'orbitale 1s du carbone pour une série de molécules organiques. La Table 7.3 reprend les valeurs obtenues avec une ou deux contraintes (1 CO ou 2 CO) à l'aide de l'AP. Les énergies d'ionisation de cœur obtenues en LSCF par P.-F. Loos ne sont pas montrées parce qu'elles sont identiques à une erreur numérique près. La déviation moyenne par rapport aux valeurs LSCF est de 1×10^{-5} et de 3.5×10^{-4} lorsque l'on pose respectivement 1 CO ou 2 CO.

Molécules	E_{ion} (eV)			Déviations		
	2 CO	1 CO	Exp.	2 CO	1 CO	
Méthane	C	290.78	289.65	290.84	-0.06	-1.19

Table 7.3 – suite de la page précédente

Molécules		E_{ion} (eV)			Déviations	
		2 CO	1 CO	Exp.	2 CO	1 CO
Éthane	C	290.67	289.55	290.71	-0.04	-1.16
Éthylène	C	290.70	289.59	290.82	-0.12	-1.23
Éthyne	C	291.19	290.07	291.25	-0.06	-1.18
Propane	CH ₃	290.48	289.37			
	CH ₂	290.63	289.52			
	Moyenne	290.56	289.44	290.57	-0.01	-1.13
Propène	CH ₂	290.10	289.00			
	CH	290.62	289.52			
	CH ₃	290.79	289.67			
	Moyenne	290.51	289.40	290.68	-0.17	-1.28
Propyne	CH	290.21	289.09	290.37	-0.16	-1.28
	C	290.84	289.74	290.93	-0.09	-1.19
	CH ₃	291.70	290.58	291.76	-0.06	-1.18
Butane	CH ₃	290.44	289.32			
	CH ₂	290.46	289.35			
	Moyenne	290.45	289.33	290.48	-0.03	-1.15
Pentane	CH ₃ (C1)	290.41	289.29			
	CH ₂ (C2)	290.41	289.30			
	CH ₂ -c (C3)	290.29	289.19			
	Moyenne	290.39	289.28	290.42	-0.03	-1.14
RMSD hydrocarbures					0.09	1.19
Formaldéhyde	C	294.46	293.36	294.47	-0.01	-1.11
Acétone	C(O)	293.72	292.62	293.71	0.01	-1.09
				293.88	-0.16	-1.26
	CH ₃	291.08	289.97	291.15	-0.07	-1.18
Méthanol	C			291.23	-0.15	-1.26
		292.44	291.32	292.30	0.14	-0.98
				292.80	-0.36	-1.48

Table 7.3 – suite de la page précédente

Molécules		E_{ion} (eV)			Déviations	
		2 CO	1 CO	Exp.	2 CO	1 CO
Éthanol	CH ₂	292.30	291.19	292.50	-0.20	-1.31
	CH ₃	290.69	289.57	291.10	-0.41	-1.53
Acide éthanoïque	CH ₃	291.85	290.73	291.55	0.03	-0.82
				291.60	0.25	-0.87
	COOH	295.35	294.26	295.38	-0.03	-1.12
Diméthyléther	CH ₃	292.17	291.06	292.17	0.00	-1.11
				292.55	-0.38	-1.49
Acide formique	C	296.33	295.46	295.80	0.53	-0.34
				295.90	0.43	-0.44
Méthylamide	CH ₃	291.55	290.43	291.60	-0.05	-1.17
RMSD hétérocarbures					0.26	1.15
RMSD total					0.175	1.165

Table 7.3: Comparaison des énergies d'ionisation de cœur du carbon 1s d'une série de composés organiques en employant la méthode LSCF et AP. L'optimisation et les calculs d'énergies ont été obtenus au niveau de théorie : PBE0/6-311++G**//B3LYP/6-311++G**.

Cette Table nous montre de manière très claire que la méthode 2 CO reproduit très bien les énergies d'ionisation de cœur que ce soit pour des composés entièrement carbonés ou pour des composés constitués d'azote ou d'oxygène. En effet, le RMSD calculé pour les énergies d'ionisation du C 1s avec 2 CO des composés totalement carbonés est de 0.09 eV. Le RMSD pour les autres composés est légèrement plus grand (0.26 eV).

Concernant les hydrocarbures, les énergies d'ionisation de cœur sont expérimentalement proches de 290.5 eV pour les composés contenant des liaisons simples ou doubles. Par contre, la présence d'une liaison triple induit une légère augmentation de l'énergie d'ionisation : 291.25 eV (éthyne) et 291.76 eV (propyne). Ainsi, l'énergie d'ionisation théorique du C 1s du -CH₃ passant du propane au propène et au propyne est augmentée respectivement de +0.31 eV et +1.18 eV.

En comparant cette catégorie avec les composés contenant des hétéroatomes, nous observons

une augmentation de l'énergie d'ionisation, principalement sur les atomes de carbone liés à un alcool, de +1.66 eV (+1.46/+1.96 eV) passant du méthane au méthanol théoriquement (expérimentalement). Le fait de passer de l'éthanol à l'acide éthanoïque induit aussi un déplacement du spectre expérimental de +2.88/+3.1 eV pour le carbone du $-\text{CH}_2\text{-OH}$ et du groupement $-\text{C}(=\text{O})\text{-OH}$. Ce déplacement est aussi bien reproduit par la théorie (+3.05 eV). Cet effet de légère augmentation de l'énergie d'ionisation en présence d'un atome électronégatif a déjà été observé d'un point de vue expérimental aussi. Cela nous montre donc que les énergies d'ionisation sont très sensibles à leur environnement proche et à la nature des atomes mais aussi que notre approche théorique (2 CO) reproduit très bien les données expérimentales.

Tout comme pour le méthane, le calcul du potentiel d'ionisation de cœur avec 1 CO diminue constamment l'énergie d'ionisation d'environ 1.12 eV alors que cette situation devrait être plus proche de la réalité vu qu'il y a plus de relaxation orbitalaire que dans 2 CO. La Figure suivante reprend les valeurs de la Table 7.3 et montre clairement cette tendance.

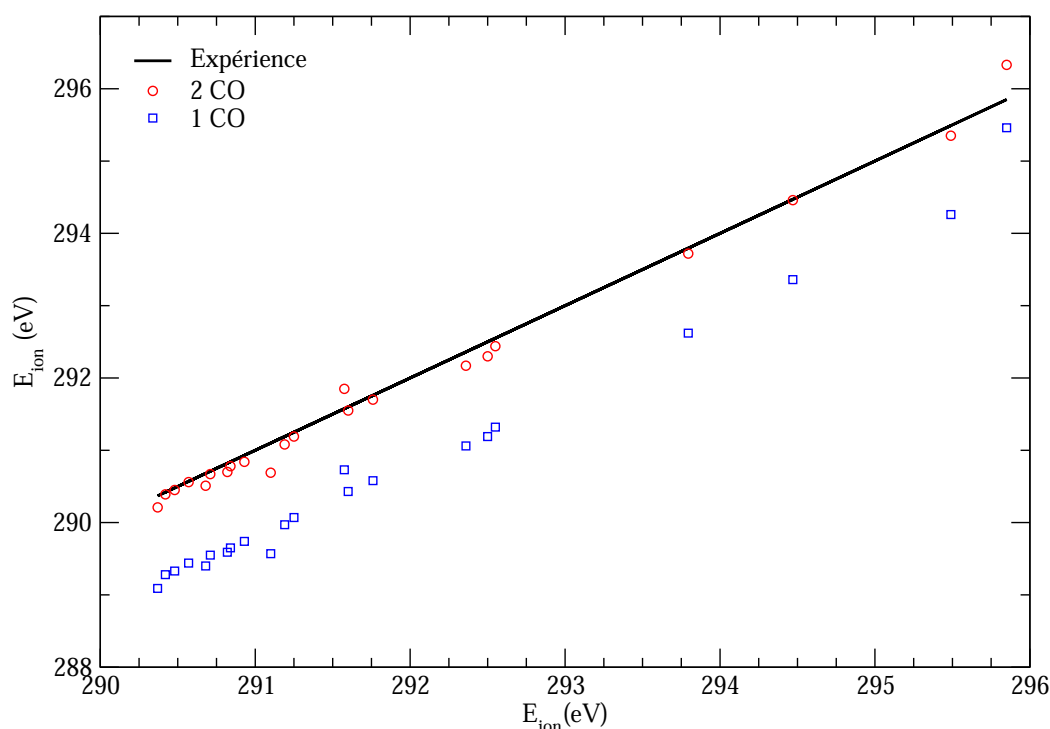


Figure 7.1: Graphe comparatif des énergies d'ionisation obtenues lorsque nous imposons une ou deux contraintes par rapport aux valeurs expérimentales issues des références [341,342]. L'optimisation et l'énergie des états électroniques ont été obtenus au niveau de théorie BF-PBE0/6-311++G**//B3LYP/6-311++G**.

Cet ensemble de résultats nous a permis d'établir et ensuite d'appliquer une régression linéaire sur les valeurs obtenues avec 1 CO pour qu'elles se juxtaposent aux valeurs obtenues avec 2 CO. Cette relation, nommée LR-AP, a pour équation : $E_{\text{ion}}^{\text{LR}} = 0.98E_{\text{ion}}^{\text{1CO}} + 6.75$ et sera appliquée dans la section suivante relative aux acides aminés.

7.2.2 Application aux acides aminés

Après la validation de l'approche AP sur un grand nombre de molécules organiques, nous avons calculé les ionisations de cœur d'une série d'acides aminés à laquelle nous avons appliqué la régression linéaire. Actuellement, des spectres NEXAFS permettent d'associer une signature spectrale à une classe d'acides aminés. Par exemple, les acides aminés à chaîne latérale aromatique ont une signature différente⁴. Nous disposons des données expérimentales pour chacun des acides aminés et principalement pour les énergies d'ionisation des couches 1s du carbone α (C_α) et du carbone du carboxyle ($C(=O)OH$) du squelette des acides aminés [362–364]. Les données sur l'azote pour les acides aminés sont plus rares. Les 20 acides aminés peuvent être classés en quatre groupes qui se différencient par la nature de leur chaîne latérale :

1. Les acides aminés possédant des chaînes latérales aliphatiques telles que l'alanine (Ala), la valine (Val), la leucine (Leu), l'isoleucine (Ile) sont non-polaires.
2. Les acides aminés possédant des groupements fonctionnels comme -OH, -SH, -SCH₃ se trouvent dans la sérine (Ser), la thréonine (Thr), la cystéine (Cys) et la méthionine (Met).
3. Les résidus ayant une chaîne latérale avec des acides carboxyliques ou amines comme l'aspartate (Asp) et le glutamate (Glu), et l'asparagine (Asn) et la glutamine (Gln).
4. Le groupe des acides aminés possédant des chaînes latérales fortement basiques comme la lysine (Lys) et l'arginine (Arg). Leur spectre NEXAFS sont très similaires au groupe précédent.
5. Les acides aminés ayant une chaîne latérale aromatique : la phénylalanine (Phe), le tryptophane (Trp), la tyrosine (Tyr) et l'histidine (His).

Seuls les acides aminés des groupes 1., 2. et 5. ont été considérés. La Table 7.4 reprend les énergies d'ionisation de cœur de l'orbitale 1s de l'azote, du carbone α et du carbone du carboxyle. Ces valeurs sont des moyennes parce que pour chaque acide aminé deux conformations ont été

⁴Celle-ci n'est pas modifiée lors de la formation d'une liaison peptidique

considérées⁵. Nous verrons dans la section suivante dédiée à la polymérisation de la glycine que quatre conformations de son monomère ont ainsi été étudiées.

De manière générale, nous pouvons dire que les valeurs obtenues théoriquement, reprises dans la Table 7.4 sont en excellent accord avec l'expérience que ce soit pour l'ionisation du carbone 1s en position α ou celui du carboxyle avec respectivement un RMSD de 0.29 eV et 0.36 eV. La régression linéaire appliquée fonctionne donc de manière très satisfaisante sur ce type de composé. Les énergies d'ionisation de cœur du C_α expérimentales se trouvent entre 292.0 eV et 292.3 eV quels que soient les acides aminés et entre 292.15 et 292.89 eV avec l'approche théorique LR-AP. De la même manière la valeur expérimentale (théorique) de l'ionisation de l'orbitale 1s du carbone du $-C(=O)OH$ est comprise entre 295.0 (294.20) eV et 295.3 (295.05) eV. Nous pouvons donc dire que les valeurs obtenues théoriquement sont en bon accord avec l'expérience et que nous pouvons distinguer sans ambiguïté les deux types de carbone étudié.

Concernant les ionisations de cœur de l'orbitale 1s du l'azote, nous observons que pour la glycine (seule donnée expérimentale) un écart de 2.93 eV est obtenu entre la valeur expérimentale et la valeur théorique. L'énergie d'ionisation de cœur de l'azote se trouve toujours entre 402.0 et 403.0 sachant que expérimentalement elle doit se trouver vers 405.58 ± 0.15 . En utilisant le schéma 2 CO, l'écart est sensiblement plus faible (0.94 eV). La relation établie pour les atomes de carbone n'est donc sans doute pas la meilleure pour reproduire les énergies d'ionisation de cœur de l'azote, cependant l'utilisation des deux contraintes induit aussi un écart assez important par rapport à l'expérience.

Acides aminés	1s	E_{ion} (eV)		Écarts
		LR-AP	Exp.	
Alphatiques				
Glycine	N(H)	402.65	405.6	2.95
	C_α	292.49	292.3	-0.19
	C(=O)OH	295.05	295.1	0.05
Alanine	N(H)			
	C_α	292.31	[292.2, 292.1]	

⁵Les conformations ne diffèrent de la position de l'hydrogène du groupement terminal C(=O)OH et par l'orientation du groupement terminal amine.

Table 7.4 – suite de la page précédente

Acides aminés	1s	E_{ion} (eV)		Écart
		LR-AP	Exp.	
Valine	C(=O)OH	294.76	[295.0, 295.1]	
	N(H)	402.29		
	C_{α}	292.08	[292.2, 292.1]	[0.12, 0.22]
Isoleucine	C(=O)OH	294.76	[295.0, 295.1]	[0.24, 0.34]
	N(H)	402.22		
	C_{α}	292.89	[292.2, 292.1]	[-0.6, 0.7]
Sérine	C(=O)OH	294.20	[295.0, 295.1]	[0.8, 0.9]
	N(H)	402.29		
	C_{α}	292.34	[292.2, 292.3]	[-0.14, -0.04]
Thréonine	C(OOH)	294.85	[295.0, 295.3]	[0.15, 0.25]
	N(H)	402.41		
	C_{α}	292.24	[292.2, 292.3]	[-0.04, -0.06]
Cystéine	C(=O)OH	294.96	[295.0, 295.1]	[0.04, 0.14]
	N(H)	402.39		
	C_{α}	292.49	[292.2, 292.1]	[-0.29, -0.39]
Méthionine	C(=O)OH	294.99	[295.0, 295.1]	[0.01, 0.11]
	N(H)			
	C_{α}		292.2(1)	
	C(=O)OH		295.0(1)	
Aromatiques				
Phénylalanine	N(H)	402.96		
	C_{α}	292.45	[292.2, 292.1]	[-0.25, -0.35]
	C(=O)OH	295.05	[295.0, 295.1]	[-0.05, 0.05]
Tyrosine	N(H)	402.18		
	C_{α}	292.17	[292.2, 292.1]	[0.03, -0.07]
	C(=O)OH	294.57	[295.0, 295.1]	
Histidine (δ)	N(H)	401.88		
	C_{α}	292.35	[292.2, 292.1]	[-0.15, -0.25]

Table 7.4 – suite de la page précédente

Acides aminés	1s	E_{ion} (eV)		Écart
		LR-AP	Exp.	
Histidine (ϵ)	C(=O)OH	294.90	[295.0, 295.1]	[0.10, 0.20]
	N(H)	402.01		
	C_{α}	291.96		[0.24, 0.14]
Histidine (protonée)	C(=O)OH	294.49		[0.51, 0.61]
	N(H)	407.20		
	C_{α}	297.15		[-4.95, -5.05]
Proline	C(=O)OH	299.68		[-4.68, -4.58]
	N(H)	402.03		
	C_{α}	292.15		
	C(=O)OH	294.68		

Table 7.4: Comparaison des énergies d’ionisation de cœur du carbon 1s d’une série de composés organiques en employant la méthode LSCF et AP. L’optimisation et les calculs d’énergies ont été obtenus au niveau de théorie : PBE0/6-311++G**//B3LYP/6-311++G**. Les valeurs expérimentales proviennent des références [362–364].

Parmi les acides aminés aliphatiques, la chaîne latérale des acides aminés n’induit pas expérimentalement de modifications dans les énergies d’ionisation. Cela montre que les énergies d’ionisation de cœur du C_{α} et du C terminal ne sont que très peu affectées par la nature des atomes se trouvant à deux liaisons ou plus en phase gazeuse et lorsqu’il s’agit de composés neutres. Cela est aussi valide sur la série de composés organiques. Concernant les résidus aromatiques, nous pouvons observer que le résultat expérimental pour l’histidine est effectué sous sa forme neutre. En effet, la déviation entre l’expérience et l’histidine protonée est supérieure à 4 eV d’écart que ce soit pour les carbones ou l’azote. Cela n’est pas surprenant parce qu’il est bien connu que les énergies d’ionisation de cœur sont fortement affectées par les états de protonation et ainsi du pH. Les acides aminés existent dans différents états de protonation en fonction de leur environnement local ou du pH. Hitchcock et collaborateurs ont montré que le spectre NEXAFS du carbone 1s est très sensible à la protonation/déprotonation des groupement terminaux carboxyle et amine [345]. Le cas de l’histidine illustre bien le fait que les états de protonation ont beaucoup

d'influence et ne peuvent être négligés. La présence d'une charge dans l'environnement proche de l'atome ionisé jouera donc un rôle important.

Dans cette partie sur les énergies d'ionisation de cœur des acides aminés, nous sommes parvenus à reproduire raisonnablement les énergies d'ionisation du squelette des acides aminés aliphatiques et aromatiques. Cependant, les énergies d'ionisation 1s de l'azote sont assez fortement sous-estimées, cela pourrait être dû à la base utilisée. Une base plus étendue serait sans doute plus judicieuse pour les atomes d'azote⁶. La présence de hétéroatomes ou de cycles aromatiques à plus de deux liaisons du carbone étudié ne déplace ni le spectre expérimental d'absorption NEXAFS, ni les pics d'absorption théorique. Nous avons confirmé que l'état de protonation modifie fortement les énergies d'ionisation de cœur avec les deux états de protonation de l'histidine. Nous allons maintenant observer l'effet de la polymérisation de glycine et ensuite les variations de l'énergie d'ionisation du C $_{\alpha}$ des glycines de la HSA en fonction de son environnement proche.

7.3 Les effets de l'environnement sur la Glycine

LE dipeptide le plus simple impliquant une liaison peptidique est la glycil-glycine ((Gly) $_2$). Hitchcock et collaborateurs ont mesuré les spectres des C 1s, N 1s et O 1s des deux espèces gazeuses (Gly) et (Gly) $_2$ ainsi que les trois espèces solides, Gly, (Gly) $_2$, (Gly) $_3$ [345]. Seulement quelques études de chimie quantique ont été réalisées sur plusieurs dipeptides (Gly-Thr, His-Phe, Gly-Trp, Arg-Gly et Gly-Gly-Gly) [343].

7.3.1 La mono-, di- et triglycine

Cette étude consiste principalement à, dans un premier temps, évaluer les énergies d'ionisation de la poly-glycine afin d'observer les effets conformationnels et ceux de la formation de liaisons peptidiques⁷.

Il y a plusieurs conformères stables du monomère de la glycine en phase gazeuse. Des études théoriques ont montré qu'il en existait huit qui sont nommées Ip, IIp, III, IVp, IVn, Vn, VIp, VIIp et VIIIIn (les conformères de symétrie C $_s$ (C $_1$) sont notées p (n)) [365]. Cependant, que certaines d'entre elles ont été observées expérimentalement : Ip, IIp, IIIp et IVn. Ces quatre

⁶La base utilisée étant optimale pour caractériser les ionisations de cœur d'atomes de carbone.

⁷Dans la référence [356] se trouve l'analyse concernant la polymérisation de l'alanine avec l'approche LSCF.

conformères se trouvent sur la Figure 7.2. Les conformations Ip et IIIp sont majoritaires à 53 % et de 30 % [345]. Les deux autres conformations sont possibles (IIp et IVn) mais sont moins abondantes (9% et 7%). Nous avons considéré, ensuite, deux conformations pour le dipeptide, la Figure 7.2 reprend celles-ci. Ces deux conformations diffèrent entre elles par l'angle dièdre $C_{\alpha}2-N(H)-C(=O)-O$ passant de 0° à 180° . Ces deux conformations sont étudiées afin d'observer les modifications qu'elles entraînent sur les énergies d'ionisation. Finalement, nous avons calculé les énergies d'ionisation de cœur de la glycine centrale du tri-glycine ((Gly)₃) en conformation hélice α et feuillet β .

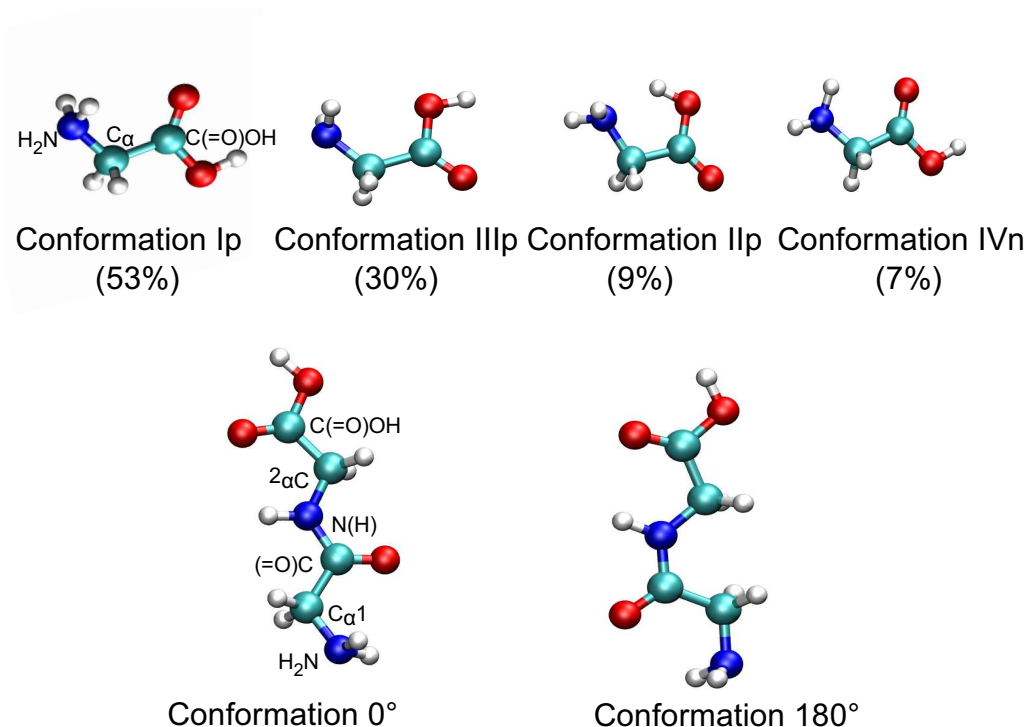


Figure 7.2: Conformations de la glycine monomère et dimère explorée lors de cette étude.

Les valeurs expérimentales pour l'ionisation de la couche 1s du carbone de la Gly, (Gly)₂ et (Gly)₃ sont similaires en phase solide ou gazeuse [345]. Un déplacement de ~ 0.3 eV pour l'ionisation du $C(=O)$ entre le monomère et le di- ou trimère en phase solide est observé. Il est dû à la formation de la liaison peptidique. En effet, passant du monomère au di ou tri-peptide, le groupement formyl devient le groupement amide de la liaison peptidique. L'effet de la formation de la liaison peptidique devrait être aussi observé pour l'ionisation de l'atome d'azote.

La Table 7.5 reprend les énergies d'ionisation de cœur de la glycine et de la glycil-glycine.

Les conformations du monomère Ip et IVn sont plus stables de $1.5 \text{ kcal.mole}^{-1}$ par rapport aux conformations IIp et IIIp. Nous observons que les énergies d'ionisation de la conformation Ip et IIIp sont très similaires quel que soit l'atome ionisé. Ces conformations sont d'ailleurs très similaires, une simple rotation du groupement carboxyle. Les énergies d'ionisation de l'azote et du carbone du groupement $-C(=O)OH$ de la conformation IIp sont quant à elles modifiées de -0.2 eV et de $+1.1 \text{ eV}$, par rapport à Ip et IIIp ; cela est dû à la présence d'une liaison hydrogène entre l'amine et le carboxylate. Le groupement donneur est donc plus affecté par la formation de liaisons hydrogène. La conformation IVn ne diffère de la conformation Ip que par sa symétrie, l'énergie d'ionisation de cœur du $-N(H)$ est fortement modifiée avec un déplacement de $+1.0 \text{ eV}$ par rapport à la conformation Ip. Cette dernière observation diffère de l'interprétation de Gordon qui montrait que la conformation IIp possédait la plus grande différence par rapport aux trois autres conformations avec seulement un pic expérimental légèrement plus large pour l'ionisation de l'azote, mais pas de déplacements significatifs.

Concernant la glycyl-glycine, les modifications des énergies d'ionisation sont très faibles en fonction de la conformation 0° et 180° , au maximum 0.1 eV . La conformation 0° est cependant plus favorable de $3.8 \text{ kcal.mole}^{-1}$, principalement pour des raisons stériques. En comparant les énergies d'ionisation du monomère Ip et de la conformation 0° , le passage de $C(=O)OH$ à $C(=O)N(H)$ est marqué par un déplacement de $+0.39 \text{ eV}$, ce qui est en total accord avec l'expérience ($\sim 0.3 \text{ eV}$). De la même manière, le passage de l'amine à la liaison peptidique induit sur l'énergie d'ionisation de l'azote un déplacement assez important de $+1.02 \text{ eV}$.

La Table 7.6 reprend les énergies d'ionisation de cœur du squelette de la glycine centrale de la glycyl-glycyl-glycine ($(\text{Gly})_3$) dans les conformations hélice α et feuillet β . Une étude similaire de P.F. Loos a été menée sur le trimère de l'alanine. Comme dans cette étude, les différences entre les deux conformations sont, comme dans le cas du dimère, relativement faibles passant de la conformation en hélice α à la conformation en feuillet β : $+0.87 \text{ eV}$ pour l'ionisation $1s \text{ N}$, -0.41 eV $1s \text{ C}_\alpha$ et $+0.21 \text{ eV}$ pour le $1s \text{ C}(=O)$. Par rapport au trimère de l'alanine, les déplacements entre les deux conformations de la glycine sont légèrement plus importants et ne sont pas nécessairement dans le même sens : $+0.13 \text{ eV}$ pour le $1s \text{ C}_\alpha^{Ala}$ et $+0.18$ pour le $1s \text{ C}(=O)$ [356]. L'effet de la polymérisation sur le C_α est marqué lorsque la conformation du trimère est l'hélice α ($+0.4 \text{ eV}$) ce qui n'est pas surprenant parce que la géométrie de la glycine centrale dans le feuillet β est très proche de la géométrie de la glycine isolée (cf. Figure 7.3).

Monomère	E_{ion} (eV)					Expérience
	Conformations					
Ionisation 1s	Ip	IIIp	IIp	IVn	Moyenne	
NH ₂	402.23	402.26	402.08	403.22	402.65	405.58
C _α	292.36	292.37	292.35	292.63	292.49	292.3
C(=O)OH	294.16	294.15	295.26	294.85	295.05	295.1
Dimère	Conformations					
Ionisation 1s	0°		180°			
NH ₂	401.90		401.91			
C _α 1	291.92		291.95			
C(=O)	293.77		293.78			
N(H)	403.25		403.23			
C _α 2	292.71		292.81			
C(=O)OH	295.68		295.68			

Table 7.5: Énergie d'ionisation de cœur en eV des quatre conformations du monomère de la Glycine et des deux conformations du dimère de la Glycine (voir Figure 7.2) en employant la régression linéaire sur les résultats AP avec 1 CO. Les valeurs expérimentales viennent de la référence [345].

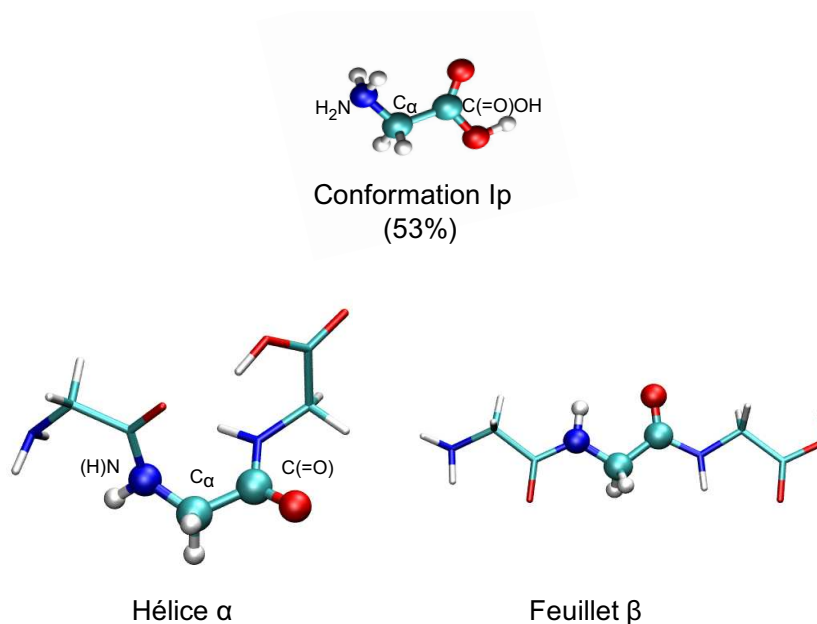


Figure 7.3: Les deux conformations de la triglycine. La glycine centrale est représentée en boule et bâton.

Ionisation 1s	E_{ion} (eV)		
	Monomères (Ip)	Trimères	
		hélice α	feuillet β
N(H)	402.23	403.88	403.01
C $_{\alpha}$	292.36	292.78	292.37
C(=O)	294.16	294.20	294.41

Table 7.6: Comparaison des énergies d’ionisation de cœur 1s sur les trois atomes du squelette de la glycine centrale du tripeptide (voir Figure 7.3) après régression linéaire. Les valeurs expérimentales viennent des références [345].^a sont les écarts à partir des valeurs calculées moyennes.

La suite de ce travail sur les glycines consiste à placer la glycine dans un environnement plus complexe comme une hélice α ou une protéine. Les spectres NEXAFS pour des biomolécules sont cependant assez rares. Seuls deux protéines sont, à ce jour, caractérisées par un spectre NEXAFS, la fibrinogène et la Sérum-Albumine Humaine ainsi que quelques polypeptides (entre 10 et 20 acides aminés) [344, 366].

7.3.2 La Sérum-Albumine Humaine (HSA)

L’énergie d’ionisation de cœur étudiée dans cette partie est celle du C $_{\alpha}$ de la glycine. La Figure suivante reprends ces énergies calculées pour le monomère, le trimère et une hélice α de 17 glycines⁸. Une première chose à considérer dans des systèmes protéiques est de valider l’approche LSCF-AP/MM, et plus précisément de vérifier que la présence de SLBO n’induit pas d’importants artefacts sur la propriété étudiée. Cette validation a été réalisée sur le trimère de la glycine. La valeur entre parenthèse sur la Figure 7.4 correspond au calcul LSCF-AP/MM (Amber99 comme champ de forces). La différence entre l’approche QM et l’approche LSCF-AP/MM n’est que de -0.27 eV. P.-F. Loos avait au préalable montré que l’effet de la SLBO n’influençait que très peu sur l’ E_{ion} du pentane (~ 0.1 eV).

Comme pour l’alanine, l’énergie d’ionisation de cœur obtenue en LSCF-AP/MM est plus faible pour l’hélice α que pour le trimère en conformation hélice α (-0.39). L’effet d’un envi-

⁸Lors de l’optimisation de la structure hélice α , le squelette de l’hélice est resté gelé. Deux SLBO ont été positionnées de par et d’autre de la glycine centrale.

ronnement en hélice α non polaire sur la glycine ou sur l'alanine est très faible. N. Ferré et X. Assfeld avait observé dans leur travail sur la crambine que l'influence de l'environnement jouait un rôle très important sur l'énergie d'ionisation de cœur du carbone 1s du glutamate dans la crambine. Un déplacement de 4 eV entre le glutamate en phase gazeuse et dans la protéine avait été observé [355]. Il est donc possible d'observer des déplacements très importants

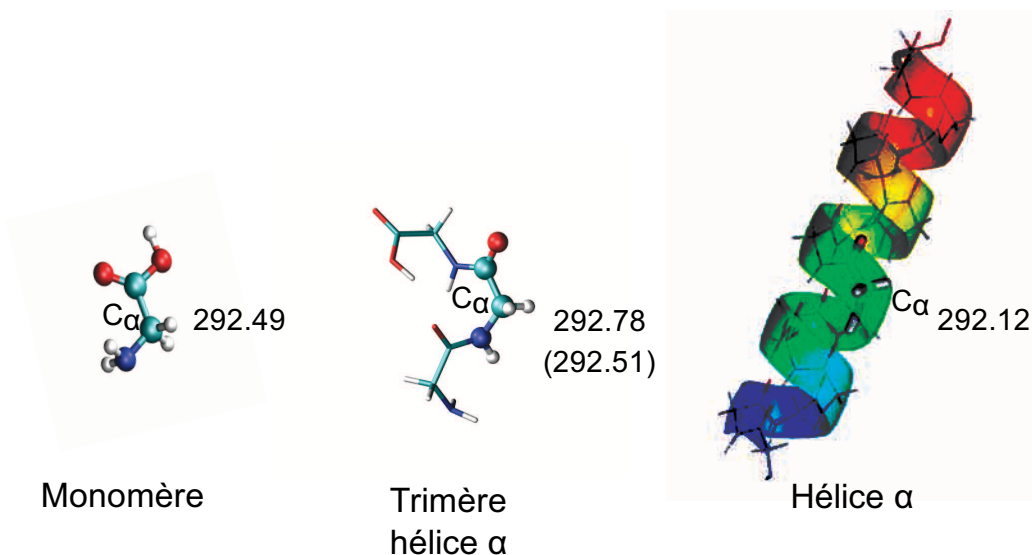


Figure 7.4: Énergie d'ionisation en eV pour le C_{α} du monomère, de trimère en conformation α et d'une polyglycine en conformation hélice α de la glycine. La valeur entre parenthèse pour le trimère est l'énergie d'ionisation de cœur obtenue en LSCF-AP/MM. La valeur pour le polypeptide est aussi obtenue en LSCF-AP/MM

dûs à l'environnement. Similairement à l'étude de N. Ferré, l'énergie d'ionisation du carbone α de la glycine va être étudiée dans la protéine Sérum-Albumine Humaine (HSA, *Human Serine Albumine*). Cette protéine produite par le foie, comporte 550 acides aminés, répartis en deux chaînes (A et B) et formant une structure secondaire constituée majoritairement d'hélices α et de quelques hélice 3_{10} ⁹. Un spectre NEXAFS de cette protéine est disponible. Cependant, vu le grand nombre d'acides aminés, il est très difficile de caractériser les énergies d'ionisation de chaque acide aminé [344].

Nous avons vu que l'état de protonation d'acides aminés comme l'histidine modifie fortement l'énergie d'ionisation du carbone α . D'autres études, similaires, ont exploré l'effet de protona-

⁹Dans une hélice α , l'acide aminé n est relié par une liaison hydrogène à l'acide aminé $n + 4$. Lorsque la liaison hydrogène se fait entre l'acide aminé n et l'acide aminé $n + 3$, on parle d'hélice 3_{10} .

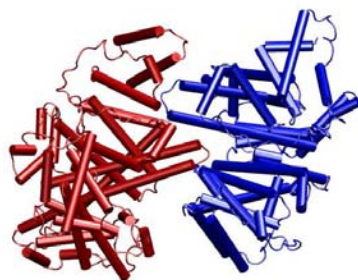


Figure 7.5: Représentation de la chaîne A (en rouge) et de la chaîne B (en bleu) en *Ribbons* de la Sérum-Albumine Humaine (code pdb 1AO6). Les bâtons représentent les hélices α ou les hélices 3_{10} .

tion/pH de la lysine [367]. Les états de protonation dans des protéines ou polypeptides modifient considérablement les spectres expérimentaux. Une étude réalisée par Zubavichus et collaborateurs a montré que le spectre de la glycine était fortement modifié en fonction du pH (forme zwitterionique, cationique, anionique de la glycine) et de son environnement proche (ajout de Li^+ et de Cl^-) [368]. Par contre, l'effet des résidus chargés dans l'environnement proche de la glycine est nettement moins étudié. Nous avons donc décidé d'explorer cet effet sur les glycines présentes dans la HSA. Dans cette étude, nous ne considérerons que la chaîne A de la HSA (code pdb 1AO6, résolution de 2.5 Å). La chaîne A est constituée d'une vingtaine d'hélices α et de 17 liaisons disulfures qui maintiennent deux à deux certaines hélices α .

À partir de la structure cristallographique de la chaîne A, nous avons utilisé propKa [300] afin de poser les états de protonation pour un pH neutre. Ainsi, tous les glutamates et les aspartates (lysines et arginines) sont déprotonés (déprotonées) à pH neutre. Seul le glutamate 244 a été protoné, ainsi que l'histidine 288. L'environnement des résidus histidines a été visualisé afin de positionner l'hydrogène sur l'azote δ ou ϵ ¹⁰. Ensuite, nous avons ajouté les atomes d'hydrogènes et neutraliser le système par 13 contre-ions Na^+ à l'aide du programme *tleap* de la suite de programme Amber9 [297]. Après cette construction du système, nous avons minimisé l'ensemble de la structure avec Tinker4.2 [298].

Les étapes d'optimisation QM/MM et de calculs d'énergie des deux états électroniques sont réalisés avec le même niveau de théorie que précédemment. Nous allons étudier les ionisations

¹⁰En δ , les histidines 9, 105, 128, 146, 338, 510, 535 et en ϵ les histidines 39, 67, 242, 367, 440, 464.

de cœur du carbone α des glycines présentes dans le système. Nous avons optimisé une par une la géométrie des onze glycines en présence de leur environnement, les deux acides aminés voisins de chaque glycine (-Xxx-Gly-YYY-) se relaxant, le reste de l'environnement est gelé. La coupure s'effectue de part et d'autre de la glycine d'intérêt au niveau de la liaison peptidique. La SLBO a été localisée avec le critère de localisation de BF sur le composé modèle qu'est la N-Méthylacétamide.

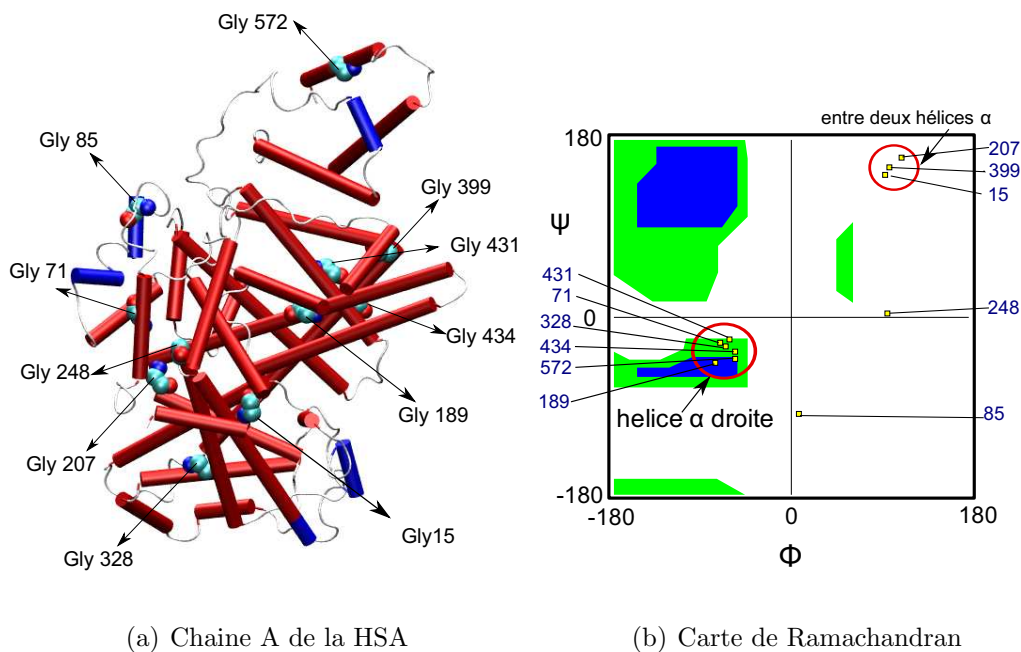


Figure 7.6: À gauche, représentation de la chaîne A en Ribbons et des onze glycines représentées par des sphères de vdW ; les hélices α sont représentées par les bâtons rouges et les hélices 3_{10} par les bâtons bleus. À droite, se trouve la carte de Ramachandran où les onze carrés jaunes représentent les glycines de la protéine HSA. La position des glycines dans la séquence de la chaîne A est indiquée en bleu.

Une première hypothèse serait que l'énergie d'ionisation du carbone α de la glycine est dépendante de la position de la glycine au sein de la structure secondaire de la chaîne A de la HSA. Pour cela, une carte de Ramachandran va nous permettre de déterminer leur position (cf. Figure 7.6). En effet, la conformation de l'ossature d'un acide aminé peut être déterminée à partir de deux angles dièdres, Φ et Ψ , caractérisés respectivement par les atomes $C(=O)-NH-C_{\alpha}-C(=O)$ et $NH-C_{\alpha}-C(=O)-NH$ d'un acide aminé. Toutes les valeurs des angles Φ et Ψ ne sont pas possibles parce que certaines sont stériquement très peu probables et, par conséquent, énergétiquement

très défavorables. Cette carte comprend trois zones énergétiquement favorables. Dans une protéine, la grande partie des acides aminés a des combinaisons d'angles (Φ , Ψ) qui se trouvent à l'intérieur de ces zones. Deux zones caractérisent les hélices α droites (la zone -57° et -47°) et le feuillet β (parallèle -119.0° et $+113.0^\circ$ et antiparallèle -139.0° et $+135.0^\circ$). La troisième région, plus petite, correspond à une conformation en hélice α gauche ($\Phi > 0$). Ces trois zones (en vert ou bleu) sont représentées sur la Figure 7.6-(b). La glycine étant un acide aminé sans chaîne latérale, elle pourra adopter des combinaisons d'angle hors de ces trois zones.

Les glycines 71, 85, 189, 328, 431, 434 et 572 sont des résidus constitutifs d'hélices α ou hélices 3_{10} . Les Glycines 15, 207, 248 et 399 ont comme caractéristique de se trouver dans les trois acides aminés reliant deux hélices α (coude ou γ -turn); trois de ces résidus sont orientés vers l'extérieur de la protéine (Gly15, Gly207 et Gly 399), la dernière (Gly248) vers l'intérieur de la protéine (*burried*).

La Table 7.7 reprend les énergies d'ionisation de l'orbitale 1s du C_α ainsi que les caractéristiques liées à la conformation et les voisins Xxx et Yyy du tripeptide -Xxx-Gly-Yyy-. Les déviations ont été calculées à partir de l'hélice α (trimère) lorsque la glycine se trouve dans une hélice (coude). Nous pouvons observer que les énergies d'ionisation se trouvent entre 285.57 (Gly572) à 295.05 (Gly399), ce qui est une région très large. La Gly15, la Gly85 et la Gly71 possèdent les plus petits écarts, inférieurs à ~ 1 eV. Ces trois résidus n'ont rien en commun, que ce soit leur structure secondaire ou même leur environnement proche. La Table 7.7 est, en fait, relativement difficile à analyser parce que plusieurs facteurs peuvent influencer : non seulement les deux acides aminés voisins (Xxx et Yyy) sont de nature très différentes (neutres ou chargés), exceptés pour la Gly15 et la Gly399 et pour la Gly189 et la Gly572 qui ont les mêmes voisins, mais aussi parce que l'environnement à 6 Å autour de chaque glycine est très diversifié.

Pour essayer de rationaliser les effets de l'environnement sur l'énergie d'ionisation, une étude systématique a été réalisée sur chaque glycine. Celle-ci consiste à extraire les onze tripeptides Xxx-Gly-Yyy de la protéine et à calculer l'énergie d'ionisation de cœur du C_α de la glycine dans la géométrie qu'ils possèdent dans la protéine. Ce tripeptide sera traité en LSCF-AP/MM en phase gazeuse. Grâce à cela, nous pourrions tirer des conclusions sur i) l'influence des deux voisins et sur ii) l'impact de l'environnement protéique.

La Figure 7.7 montre les variations de l'énergie d'ionisation du carbone α de chaque glycine des onze tripeptides en phase gazeuse par rapport à celle de la triglycine, prise comme valeur de

Glycines	E_{ion}	Écarts ^a	Conf.	Xxx et Yyy
15	292.80	+0.29	coude	Leu et Glu
71	291.16	-0.96	α	Phe et Asp
85	291.53	-0.59	α	Tyr et Glu
189	294.26	+2.14	3_{10}	Glu et Lys
207	293.58	+1.07	coude	Phe et Glu
248	291.31	-1.20	coude	His et Asp
328	290.60	-1.52	α	Leu et Met
399	295.05	+2.54	coude	Leu et Glu
431	294.40	+2.28	α	Leu et Lys
434	294.33	+2.21	α	Val et ser
572	285.57	-6.55	α	Glu et Lys
(Gly) ₃ hélice α	292.51	–	coude	Gly et Gly
Hélice α	292.12	–	α	Gly et Gly

Table 7.7: Énergies d'ionisation en eV d'un électron de l'orbitales 1s du C_α de la glycine. ^aLes écarts sont calculés en fonction de la conformation : les quatre glycines en conformation coude sont comparées à l'énergie d'ionisation de la triglycine en conformation hélice α , la déviation des sept autres glycines est calculée par rapport à l' E_{ion} glycine au centre de l'hélice α de glycine.

référence (La Table des énergies d'ionisation propre à ces tripeptides est disponible en Annexe C). L' E_{ion} du composé de référence (glycyl-glycyl-glycine) est calculée en LSCF-AP/MM (292.51 eV). Ce choix de référence se justifie parce que les deux résidus voisins de la glycine centrale sont des acides aminés non-polaires. Les effets des deux voisins Xxx et Yyy des glycines peuvent ainsi être rationalisés. Dans un premier temps, nous allons considérer les glycines liées à un acide aminé non polaire. Nous verrons que selon la nature du second acide aminé l' E_{ion} se verra modifiée. La présence d'un acide aminé possédant un atome fortement électronégatif dans sa chaîne latérale (oxygène ou soufre) comme la sérine ou la méthionine induit un déplacement faible du spectre de +0.5 eV, c'est le cas pour Leu-Gly328-Met et Val-Gly434-Ser. La présence d'un résidu chargé positivement dans le tripeptide entraîne une augmentation de l' E_{ion} plus importante de l'ordre de +1.8 eV, c'est le cas pour la Gly431 qui a pour voisin une lysine. Cette

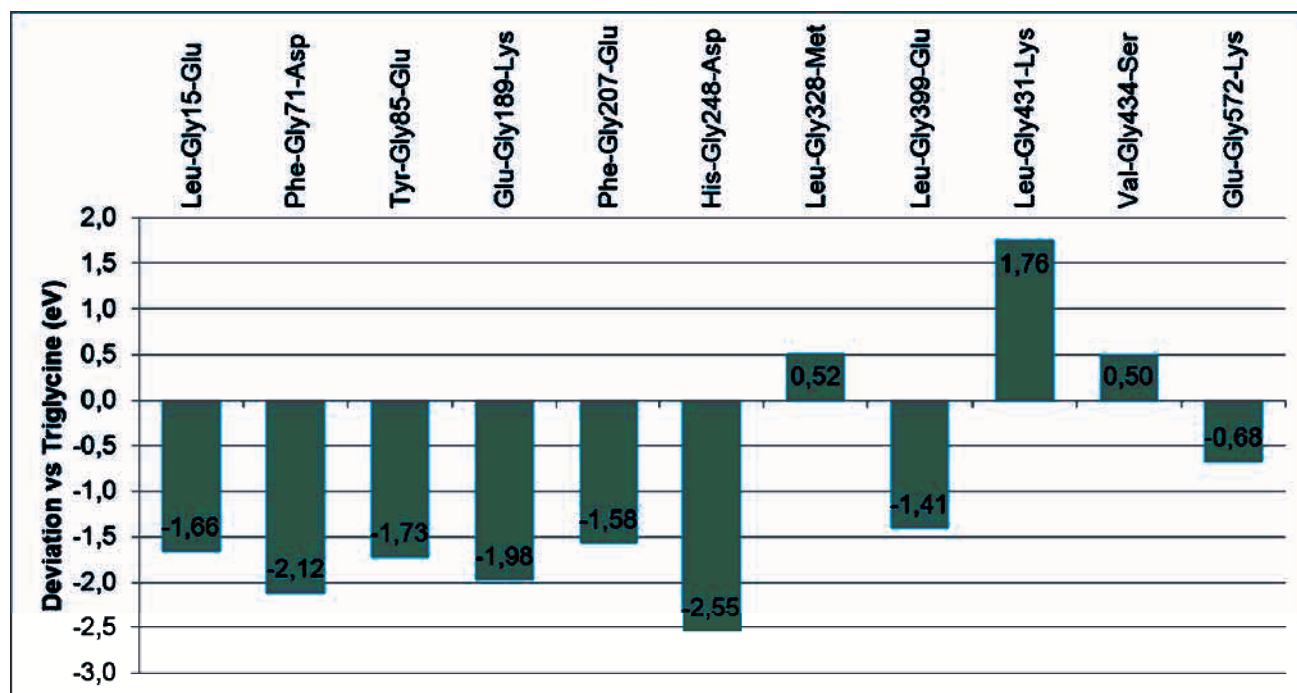


Figure 7.7: Variation de l'énergie d'ionisation du carbone α des onze glycines en présence des deux acides aminés voisins Xxx et Yyy en phase gazeuse par rapport à la triglycine (conformation hélice α).

observation semble assez logique parce que la charge positive de la lysine déstabilise l'état ionisé de cœur, son énergie s'en voit augmentée. *A contrario*, la présence d'un acide aminé chargé négativement comme un glutamate ou un aspartate stabilise l'état ionisé de cœur, diminuant son énergie. Par conséquent, l'énergie d'ionisation sera moindre (en moyenne -1.8 eV) par rapport à la référence qu'est la triglycine, c'est le cas pour la Gly15, la Gly71, la Gly207 et la Gly399. Les effets dûs à la présence d'une charge positive ou négative ont donc quantitativement la même amplitude mais de signes opposés.

Dans le cas où les deux résidus voisins de la glycine sont chargés (Gly189 et Gly572), un résidu positif et un négatif, les effets devraient s'annuler et tendre vers une valeur proche du trimère (292.51 eV). Or, ce n'est pas le cas. La Figure 7.8 représente ces deux tripeptides. Clairement, la distance et donc la géométrie entre le groupement chargé et le C_α de la glycine joue un rôle non négligeable sur l'ionisation de cœur en présence d'espèce chargée. Ainsi pour la Gly189, le groupement chargé du glutamate est 3.58 Å plus proche du C_α que du NH_3^+ de la lysine. Par conséquent, l'influence du glutamate est plus forte et l'emporte sur l'effet que peut jouer la lysine.

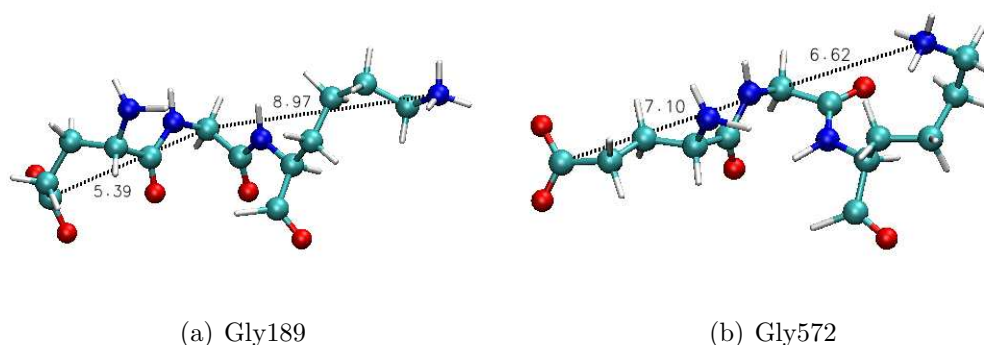


Figure 7.8: Représentation du tripeptide Glu-Gly-Lys. Les distances entre les groupements chargés et le carbone α de la glycine sont indiquées.

L'énergie d'ionisation obtenue est donc très faible (290.53 eV). Pour la Gly572, les groupements chargés du glutamate et de la lysine sont à une distance assez similaire du carbone α de la glycine, respectivement 7.10 Å et 6.62 Å. Les effets dus à la charge sur l'énergie d'ionisation s'annulent quasiment : 291.83 eV contre 292.51 pour le trimère en phase gazeuse.

Deux autres glycines (la Gly15 et la Gly399) possèdent des acides aminés voisins identiques : leucine et glutamate. Elles se trouvent toutes deux en conformation coude dans la protéine. D'un point de vue géométrique, les trois acides aminés se superposent parfaitement (RMSD = 0.53 Å). Le groupement chargé du glutamate se trouve à une distance de 6.4 Å du C_α de la glycine. Leur énergie d'ionisation de cœur est relativement similaire à 0.25 eV près : 290.85 eV et 291.10 eV.

Cette première approche nous indique de sérieux renseignements quant à l'importance des résidus chargés au voisinage des glycines étudiées. La suite de l'étude est de replacer ces tripeptides dans leur environnement protéique afin d'observer les effets liés à cet environnement. La Figure 7.9 illustre cet effet (La Table des énergies d'ionisation propres à ces tripeptides dans leur environnement est disponible en Annexe C). Majoritairement, une augmentation de l'énergie d'ionisation de cœur du carbone α de la glycine est observée, allant de +0.75 eV à +3.95 eV. Seules la Gly328 et la Gly572 voient leur énergie d'ionisation diminuer en présence de l'environnement. L'augmentation de l'énergie d'ionisation est principalement due à la présence dans l'environnement proche (6 Å) de la glycine de résidus polaires et chargés positivement (basique). La Figure 7.10 illustre l'environnement de la Gly328 et de la Gly572 dont l'énergie d'ionisation diminue lors du passage du tripeptide de la phase gazeuse vers son propre environnement.

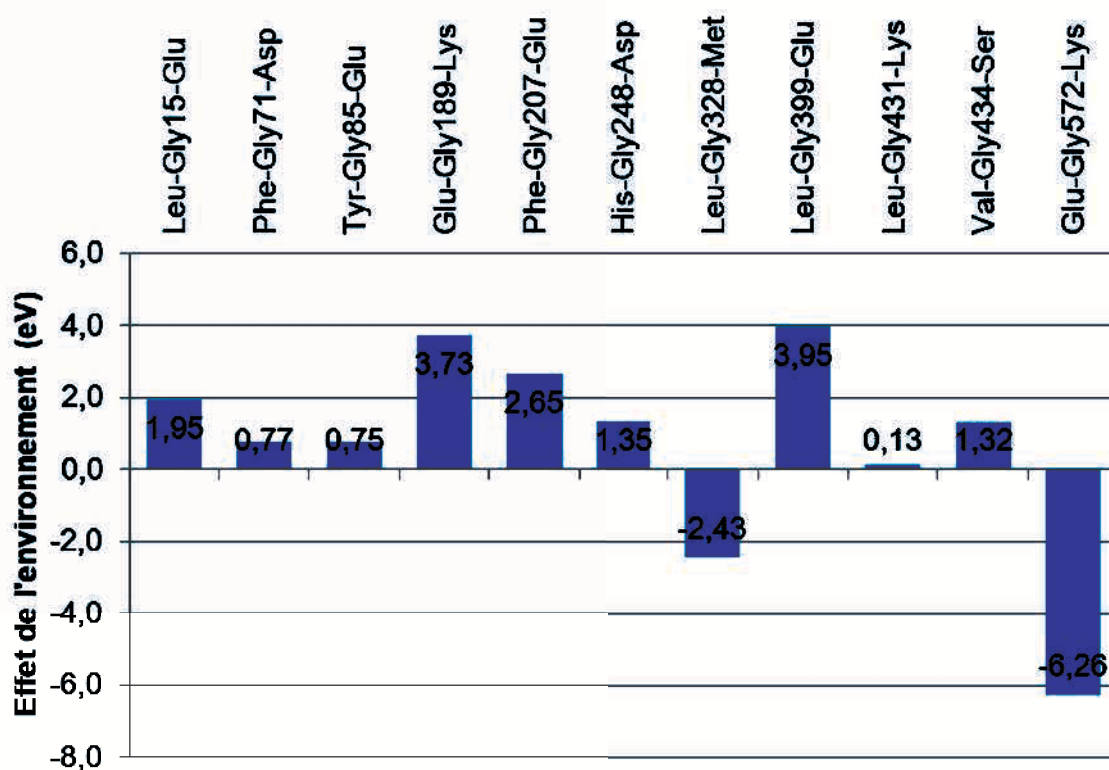


Figure 7.9: Variation de l'énergie d'ionisation du carbone α des onze glycines en présence des deux acides aminés voisins Xxx et Yyy en phase gazeuse et dans la protéine.

Le plus impressionnant est la diminution de l'énergie d'ionisation de la Gly572 (-6.26 eV), son environnement proche contient effectivement un résidu basique (Asp557) à 8.60 Å, de la même manière pour la Gly328, un aspartate est aussi présent à 7.4 Å. Cependant dans les deux cas sont aussi présents légèrement plus loin (environ 10Å de la glycine) des résidus chargés positivement. Ces deux exemples montrent clairement combien il est toujours difficile de quantifier précisément l'effet de l'environnement malgré la quantification que nous avons établie concernant les résidus chargés. De plus, ces effets électrostatiques sont modifiés selon la distance qui sépare la partie chargée de l'acide aminé de l'atome étudié.

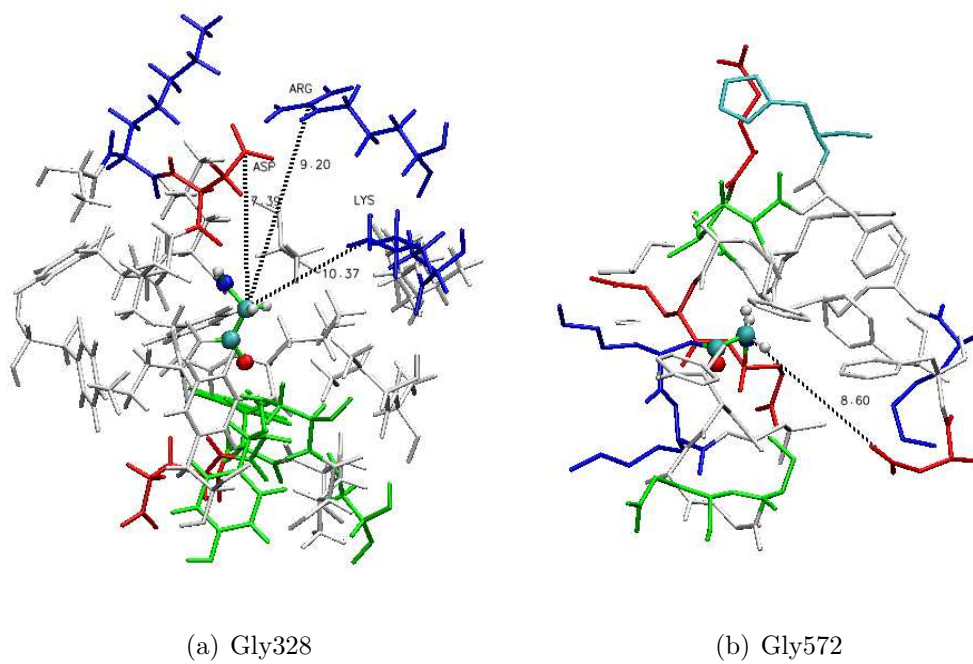


Figure 7.10: Représentation de l'environnement de la Gly328 et de la Gly557. La convention couleur est la suivante : les résidus en blanc, vert, bleu et rouge sont respectivement non-polaires, polaires, basiques et acides.

7.4 Conclusions

Nous avons montré dans cette dernière partie une nouvelle méthode, la projection asymptotique, qui orthogonalise les orbitales localisées (SLBO ou CO) aux orbitales variationnelles. Cette approche a été implémentée et a ensuite été testée et validée en évaluant les énergies d'ionisation de cœur sur un ensemble de molécules organiques. La méthode a été approuvée avec des erreurs similaires à celles obtenues en LSCF et un RMSD de 0.18 eV. À partir des molécules organiques, nous avons observé l'effet de simple, double et triple liaisons sur l'énergie d'ionisation (> 1 eV) ainsi que la présence d'atomes électronégatifs (> 2 eV) dans des groupements fonctionnels tels que l'hydroxyle, le carbonyle et le carboxyle.

Ensuite, nous avons calculé les énergies d'ionisation d'un certain nombre d'acides aminés. L'approche AP reproduit aussi très bien les énergies d'ionisation des carbones du squelette; cependant, elle ne permet pas une description assez précise de l'énergie d'ionisation 1s de l'azote peptidique. Cette étude a mis en évidence, à l'aide de l'histidine, que l'état de protonation jouait un rôle très important (> 4 eV) et que nous parvenions à évaluer cet effet. Une étude plus poussée a ensuite été réalisée sur la glycine. Expérimentalement, la formation d'une liaison peptidique induit un déplacement de l'énergie d'ionisation du carbone C(=O) de 0.3 à 0.4 eV. Nous sommes parvenus à reproduire cet effet lors du passage du monomère au trimère (0.39 eV). Les deux conformations du dimère de la glycine n'induisent pas de modifications importantes (< 0.1 eV) sur l'énergie d'ionisation de chacun des atomes.

La suite du travail a été concentrée sur l'énergie d'ionisation du carbone α de la glycine dans un milieu protéique. L'analyse a tout d'abord porté sur les onze tripeptides (Xxx-Gly-Yyy) extraits de la protéine de Sérum-Albumine Humaine contenant une glycine. À partir de ces tripeptides en phase gazeuse, nous avons montré que la présence de résidus voisins contenant des atomes électronégatifs dans leur chaîne latérale et basique induisaient une augmentation de l'énergie d'ionisation de respectivement +0.5 eV et +1.8 eV. La présence de résidus chargés négativement montre une diminution de cette énergie d'ionisation de 1.8 eV par rapport au trimère de la glycine. Une charge positive (négative) déstabilise (stabilise) l'état ionisé.

Il existe aussi une relation coulombique entre la distance entre l'atome ionisé et le groupement chargé. Au plus cette charge est proche, au plus elle aura de l'influence sur l'énergie d'ionisation. En plaçant ces tripeptides dans leur environnement protéique, une augmentation de l'énergie est due à la présence de résidus basiques. Seuls deux tripeptides voient leur énergie

d'ionisation diminuer en présence de l'environnement protéique. Les effets de l'environnement restent cependant assez difficile à déterminer. Une analyse systématique sur les résidus entourant la glycine pourrait sans doute apporter un éclaircissement notamment en éteignant successivement les charges de la chaîne latérale (cf. Etude sur kalata B1). Cependant, cela demanderait un travail considérable.

Conclusions et perspectives

L'OBJECTIF principal de cette thèse a été d'étudier les phénomènes électroniques sur différents systèmes biologiques à l'aide de l'approche QM/MM développée dans le laboratoire de Nancy, le *Local Self-Consistent Field* (LSCF). Nous avons montré que les récents développements concernant la coupure d'une liaison covalente mais aussi le couplage LSCF/MM et PCM développés par P.-F. Loos s'adaptent parfaitement dans le traitement de systèmes biologiques.

La première partie est dédiée à la présentation des méthodes de chimie quantique. Le premier chapitre présente les approches *ab initio* indépendantes du temps : Hartree-Fock (HF), les théories des perturbations et de la fonctionnelle de la densité (DFT). Nous avons ensuite décrit les méthodes qui évaluent les transitions électroniques et plus particulièrement, les méthodes HF et DFT dépendantes du temps. Nous sommes revenus sur les approches de solvatation qui sont nécessaires pour modéliser des phases condensées.

Dans la seconde partie, nous avons cité des outils qui traitent les systèmes biologiques. Les principes de base de la mécanique moléculaire ou champ de forces ont été explicités brièvement pour rassembler les théories qui interviennent dans les méthodes hybrides QM-MM. Dans le chapitre 4, nous avons développé de façon plus approfondie les méthodes QM-MM et plus spécifiquement, le LSCF/MM développé à Nancy. Nous avons décrit les récents développements de P.-F. Loos (LSCF/+3, LSCF/+3,+5) ; pour, dans la troisième partie, les utiliser pour traiter des systèmes biologiques. La partie centrale de la thèse comprend un certain nombre d'applications sur les phénomènes électroniques.

Au chapitre 5, nous avons adapté une méthodologie sur la base des développements de P.-F. Loos pour intégrer la réponse électronique de l'environnement dans les spectres d'absorption, lorsque le chromophore absorbe un photon. Cette approche est basée sur le couplage LSCF/MM avec un continuum caractérisé par la constante diélectrique dynamique, le LSCF/MM:ERS. La réponse électronique de l'environnement est habituellement négligée ; or, dans le cadre d'une collaboration avec l'ENSCP et l'Unité de Chimie Physique Théorique et Structurale, nous avons montré qu'elle est essentielle pour le complexe de squaraine-tétralactame. Grâce à l'ajout de la réponse électronique de la cage tétralactame, l'effet de la complexation du chromophore de squaraine approche l'expérience, ce qui n'était pas le cas en l'absence de l'ERS. Toujours dans le cadre de l'étude de la réponse électronique de l'environnement, nous avons similairement étudié les chromophores d'une protéine fluorescente. Sur ceux-ci, nous avons testé et validé la coupure

entre la partie QM et la partie MM au niveau de la liaison peptidique (LSCF/+3,+5). Deux conformations de la protéine ECFP ont été étudiées en absence ou en présence de molécules d'eau. Nous avons vu que la présence des molécules d'eau induisait d'importantes modifications sur la géométrie du chromophore et, par conséquent, sur son spectre d'absorption UV/visible. La contribution de la réponse électronique de l'environnement est relativement similaire aux contributions électrostatiques et géométriques (environ 10 nm). Cela montre encore une fois que celle-ci ne peut être négligée. Nous avons terminé ce chapitre en observant les effets de la substitution ou de la mutation du chromophore. Nous avons, à nouveau, utilisé notre modèle pour prendre en compte la réponse électronique de l'environnement. Avec ce modèle (LSCF/MM:ERS), nous obtenons le même ordre de grandeur dans la déviation (entre -10 nm et +10 nm) par rapport à des résultats expérimentaux possédant des mutations similaires. Au terme de ce chapitre, nous avons montré que les effets géométriques, électrostatiques et électroniques ont le même ordre de grandeur sur le spectre d'absorption des chromophores. Cela prouve qu'il est important de bien considérer ces trois, et non deux, contributions lorsqu'on traite un chromophore en présence d'une protéine. Cette méthodologie, simple et efficace, pourra par la suite être réutilisée dans le cadre de la thèse de Thibaut Very, supervisée par Xavier Assfeld. Cette thèse s'inscrit dans le cadre d'une ANR et porte sur l'insertion de complexes de ruthénium qui présentent des propriétés spectroscopiques très différentes une fois intercalés dans l'ADN. De plus, il sera nécessaire d'introduire les effets de dispersion pour correctement décrire les interactions de systèmes π - π , qui sont notamment étudiés par Soraya Abdouche. D'autre part, une collaboration avec un groupe qui développe des champs de forces polarisables pourrait s'avérer être une autre perspective pour traiter la polarisation électronique dans des systèmes biologiques. En effet, Fabien Archambault, en thèse dans l'équipe de Dynamique des Assemblages Membranaires de Nancy, a développé un champ de forces polarisables pour les systèmes biologiques.

Comme seconde application, au chapitre 6, nous avons collaboré avec Elise Dumont (ENS, Lyon) sur des macrocycles contenant des liaisons disulfures. Ce travail s'inscrit dans une série de plusieurs publications réalisées sur des petits systèmes aliphatiques contenant un pont disulfure. L'objectif était d'étudier la capture électronique sur des systèmes plus conséquents et principalement, de rationaliser les effets dus à une compétition interne au système, induite par la présence de trois ponts disulfures. L'étude nous a permis d'identifier clairement les liaisons disulfures les plus favorables à la capture électronique et de déterminer les effets géométriques

induits par la capture électronique. Une étude systématique nous a aidés à quantifier les effets électrostatiques de l'environnement sur chacune des captures électroniques. Une suite de plusieurs ordres pourrait être donnée à ce travail. En effet, il serait intéressant d'observer les effets dynamiques que provoque la capture électronique en calibrant un champ de forces spécifique des liaisons à 2c-3e. Jusqu'à maintenant, nous avons étudié sur ces systèmes particuliers les affinités électroniques adiabatiques. Nous pourrions aussi utiliser d'autres méthodes comme la mécanique moléculaire *ab initio* pour décrire des quantités thermodynamiques et s'approcher de l'évaluation des potentiels rédox.

Dans le dernier chapitre de cette partie, nous avons étudié l'effet de l'environnement sur les ionisations de cœur. En collaboration avec Vitaly N. Glushkov, nous avons implémenté la projection asymptotique, qui orthogonalise les orbitales localisées (SLBO ou CO) aux orbitales variationnelles nécessaires à rendre l'état fondamental orthogonal aux états ionisés. Elle a été testée et validée en évaluant les énergies d'ionisation de cœur d'un ensemble de molécules organiques. La méthode a été approuvée avec des erreurs similaires à celles obtenues en LSCF. À partir de ces résultats, nous avons observé que les atomes électronégatifs modifient l'énergie d'ionisation. Une étude des énergies d'ionisation de cœur des atomes du squelette des acides aminés a été réalisée. Cependant, elle ne permet pas une description assez précise de l'énergie d'ionisation de l'orbitale 1s de l'azote peptidique. Celle-ci a mis en évidence, à l'aide de l'histidine, que l'état de protonation jouait un rôle très important (> 4 eV) et que nous parvenions à évaluer cet effet. Une étude plus poussée a ensuite été réalisée sur la glycine. Nous nous sommes, en effet, concentrés sur l'énergie d'ionisation du carbone α de la glycine dans un milieu protéique. L'analyse a tout d'abord porté sur les onze tripeptides (Xxx-Gly-Yyy) de la protéine de Sérum-Albumine Humaine contenant une glycine. À partir de ces tripeptides en phase gazeuse, nous avons montré que la présence de résidus voisins polaires et basiques induisait une augmentation de l'énergie d'ionisation de, respectivement, +0.5 eV et 1.8 eV. Par contre, la présence de résidus chargés négativement indique une diminution de cette énergie d'ionisation de 1.8 eV par rapport au trimère de la glycine. Il existe aussi une relation coulombique entre l'atome ionisé et le groupement chargé. En replaçant ces tripeptides dans leur environnement protéique, une augmentation de l'énergie due à la présence de résidus polaires et basiques est observée ; cependant, cette quantification est nettement plus délicate à obtenir.

Cette thèse a montré l'importance de la prise en compte de la réponse électronique de l'envi-

ronnement sur les propriétés optiques de chromophores et les nombreux effets de l'environnement sur des phénomènes électroniques comme la capture électronique et les ionisations de cœur. Pour de tels systèmes biologiques, les méthodes hybrides QM-MM offrent d'intéressantes perspectives comme indiqué ci-dessus.

Appendices

Annexe A

Quelques fonctionnelles usuelles

A.1 LDA et LSDA

Afin d'utiliser ces résultats dans les calculs de DFT, Vosko, Wilk et Nusair (VWN) [24] ont construit une forme analytique d'interpolation entre les limites de spins non polarisés ($\zeta = 0$) et polarisés ($\zeta = 1$).

$$\begin{aligned}\epsilon_c^{\text{VWN}}(r_s, \zeta) &= \epsilon_c(r_s, 0) + \epsilon_a(r_s) \left[\frac{f(\zeta)}{f''(0)} \right] [1 - \zeta^4] + [\epsilon_c(r_s, 1) - \epsilon_c(r_s, 0)] f(\zeta) \zeta^4 \\ f(\zeta) &= \frac{(1 + \zeta)^{4/3} + (1 - \zeta)^{4/3} - 2}{2(2^{2/3} - 1)}\end{aligned}\quad (\text{A.1})$$

Les fonctions $\epsilon_c(r_s, \zeta)$ et $\epsilon_a(r_s)$ sont paramétrisées.

A.2 Les GGA

Becke [25] propose pour l'échange une relation simple :

$$\begin{aligned}\epsilon_x^{\text{B88}} &= \epsilon_x^{\text{LDA}} + \Delta\epsilon_x^{\text{B88}} \\ \Delta\epsilon_x^{\text{B88}} &= -\beta\rho^{1/3}(\mathbf{r}) \frac{x^2}{1 + 6\beta \sinh^{-1} x} \\ x &= \frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})}\end{aligned}\quad (\text{A.2})$$

Le paramètre β est déterminé par la méthode des moindres carrés sur des données expérimentales (56 énergies d'atomisation, 8 affinités protoniques et 10 énergies atomiques totales d'atomes et

de molécules issus du jeu G1 de Pople et *al.*). Perdew et Wang ont aussi défini une fonction d'échange similaire à celle de Becke, PW91 [27] :

$$\epsilon_x^{\text{PW91}} = \epsilon_x^{\text{LDA}} \left(\frac{1 + xa_1 \sinh^{-1}(xa_2) + (a_3 + a_4 e^{-bx^2}) x^2}{1 + xa_1 \sinh^{-1}(xa_2) + a_5 x^2} \right) \quad (\text{A.3})$$

où a_{1-5} sont des constantes et x est défini par l'Eq. (A.2). Pour l'énergie de corrélation, la fonctionnelle de corrélation de Lee-Yang-Parr (LYP), la plus populaire, est caractérisée par :

$$\begin{aligned} \epsilon_c^{\text{LYP}} &= \frac{-a\gamma}{(1 + d\rho^{-1/3}(\mathbf{r}))} - \frac{ab\gamma e^{-c\rho^{-1/3}(\mathbf{r})}}{9(1 + d\rho^{-1/3}(\mathbf{r}))\rho^{8/3}(\mathbf{r})} \\ &\times \left[18(2^{2/3})C_F \left(\rho_\alpha^{8/3}(\mathbf{r}) + \rho_\beta^{8/3}(\mathbf{r}) \right) - 18\rho(\mathbf{r})\mathcal{T} \right. \\ &\left. + \rho_\alpha(\mathbf{r})(2\mathcal{T}^\alpha + \nabla^2\rho_\alpha(\mathbf{r})) + \rho_\beta(\mathbf{r})(2\mathcal{T}^\beta + \nabla^2\rho_\beta(\mathbf{r})) \right] \end{aligned} \quad (\text{A.4})$$

$$C_F = \frac{3}{10} (3\pi^2)^{2/3} \quad (\text{A.5})$$

$$\gamma = 2 \left[1 - \frac{\rho_\alpha^2(\mathbf{r}) + \rho_\beta^2(\mathbf{r})}{\rho^2(\mathbf{r})} \right]$$

$$\mathcal{T}_\sigma = \frac{1}{8} \left(\frac{|\nabla\rho_\sigma(\mathbf{r})|^2}{\rho_\sigma(\mathbf{r})} - \nabla^2\rho_\sigma(\mathbf{r}) \right)$$

La fonctionnelle \mathcal{T}_σ est la densité d'énergie cinétique locale de Weizsacker. Il est intéressant de noter que les paramètres (a , b , c , et d) sont déterminés grâce aux données expérimentales de l'atome d'hélium! Voici une autre forme de correction sur la corrélation, celle de Perdew, dite P86 [26] :

$$\epsilon_c^{\text{P86}} = \epsilon_c^{\text{LDA}} + \Delta\epsilon_c^{\text{P86}} \quad (\text{A.6})$$

$$\Delta\epsilon_c^{\text{P86}} = \frac{e^\phi C(\rho(\mathbf{r})) |\nabla\rho(\mathbf{r})|^2}{f(\zeta)\rho^{7/3}(\mathbf{r})} \quad (\text{A.7})$$

$$f(\zeta) = 2^{2/3} \sqrt{\left(\frac{1+\zeta}{2}\right)^{5/3} + \left(\frac{1-\zeta}{2}\right)^{5/3}}$$

$$\phi = a \frac{C(\infty)|\nabla\rho(\mathbf{r})|}{C(\rho(\mathbf{r}))\rho^{7/6}(\mathbf{r})}$$

$$C(\rho(\mathbf{r})) = \mathcal{C}_1 + \frac{\mathcal{C}_2 + \mathcal{C}_3 r_s + \mathcal{C}_4 r_s^2}{1 + \mathcal{C}_5 r_s + \mathcal{C}_6 r_s^2 + \mathcal{C}_7 r_s^3}$$

où a et \mathcal{C}_{1-7} sont des scalaires constants.

Annexe B

Les acides aminés

Voici la liste des acides aminés avec les abréviations.

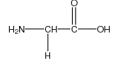
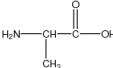
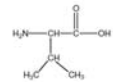
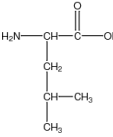
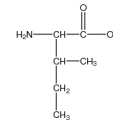
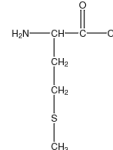
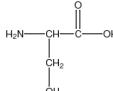
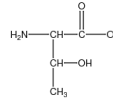
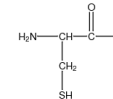
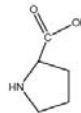
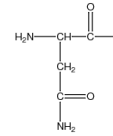
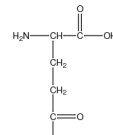
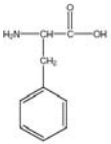
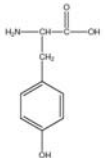
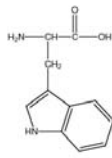
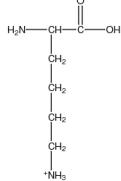
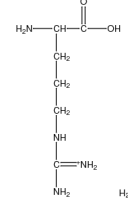
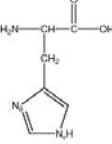
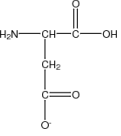
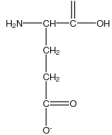
Noms	Codes à 3 lettres	Codes à 1 lettre	Structures
Glycine	Gly	G	
Alanine	Ala	A	
Valine	Val	V	
Leucine	Leu	L	
Isoleucine	Ile	I	
Méthionine	Met	M	
Sérine	Ser	S	
Thréonine	Thr	T	
Cystéine	Cys	C	
Proline	Pro	P	
Asparagine	Asn	N	
Glutamine	Gln	Q	

Table B.1 – suite de la page précédente

Noms	Codes à 3 lettres	Codes à 1 lettre	Structures
Phénylalanine	Phe	F	
Tyrosine	Tyr	Y	
Tryptophane	Trp	W	
Lysine	Lys	K	
Arginine	Arg	R	
Histidine	His	H	
Aspartate	Asp	D	
Glutamate	Glu	E	

Annexe C

Les ionisations de cœur

C.1 La Sérum-Albumine Humaine

Voici la Table des énergies d'ionisation du carbone α de chacune des glycines de la HSA. Cette Table contient donc les valeurs pour les onze tripeptides incluant une glycine en phase gazeuse, ayant la géométrie qu'ils auraient s'ils se trouvaient dans la protéine, ainsi que lorsqu'ils sont placés dans leur environnement protéique.

Tripeptides	E_{ion}		Déviation ^a
	isolé	protéine	
Leu-Gly15-Glu	290.85	292.80	+1.95
Phe-Gly71-Asp	290.39	291.16	+0.77
Tyr-Gly85-Glu	290.78	291.53	+0.75
Glu-Gly189-Lys	290.53	294.26	+3.73
Phe-Gly207-Glu	290.93	293.58	+2.65
His-Gly248-Asp	289.96	291.31	+1.35
Leu-Gly328-Met	293.03	290.60	-2.43
Leu-Gly399-Glu	291.10	295.05	+3.95
Leu-Gly431-Lys	294.27	294.40	+0.13
Val-Gly434-Ser	293.01	294.33	+1.32
Glu-Gly572-Lys	291.83	285.57	-6.26
(Gly-Gly-Gly) _{α}	292.51		

Table C.1: Énergies d'ionisation en eV de l'orbitale 1s du C _{α} de la glycine dans le tripeptide et dans la protéine. La référence est le tripeptide en phase gazeuse. ^aLa déviation est calculée entre le système en phase gazeuse et dans son environnement.

Liste des abréviations

Abréviations	Détails
ACM	Connexion Adiabatique <i>adiabatic connexion</i>
AE	Affinité Électronique
AEA	Affinité Électronique Adiabatique
AP	Projection asymptotique <i>Asymptotic Projection</i>
APC	<i>Atomic point Charges</i>
ASC	<i>Apparent Surface Charges</i>
BF	Boys-Föster
CASCI	<i>Complete Active Space Configuration Interaction</i>
CASSCF	<i>Complete Active Space Self-Consistent Field</i>
CF	Champ de Forces
CFP	Champ de Forces Polarisables
CIS(D)	<i>Configuration interaction singles (doubles)</i>
DD	Diradical dianion
DFT	<i>Density Functional Theory</i>
DFT-D	<i>Density Functional Theory with Dispersion</i>
DT	Dithiolate
ECFP	<i>Enhanced Cyan Fluorescent Protein</i>
EE	<i>Electrostatic Embedding</i>
EF	État Fondamental
EFP	<i>Effective Fragment Potential</i>
FBP	<i>Frontier Bond Potential</i>
FCO	<i>Frozen Core Orbital</i>
FO	<i>Frozen Orbital</i>
GGA	<i>Generalized Gradient Approximation</i>
GHO	<i>Generalized Hybrid Orbital</i>

GFP	<i>Green Fluorescent Protein</i>
HF	Hartree-Fock
HLA	<i>Hydrogen Link Atom</i>
HSA	Sérum-Albumine Humaine (<i>Human Serum Albumin</i>)
IEF-PCM	<i>Integral Equation Formalism of Polarizable Continuum Model</i>
IMOMM	<i>Integrated Molecular Orbital/Molecular mechanics</i>
IMOMO	<i>Integrated Molecular Orbital/Molecular Orbital</i>
LA	<i>Link Atom</i>
LJ	Potentiel de Lennard-Jones
LSCF	Champ auto-cohérent localisé <i>Local Self-consistent field</i>
L(S)DA	<i>Local (Spin) Density Approximation</i>
MCSCF	<i>Multi-Configuration Self Consistent Field</i>
ME	<i>Mechanical Embedding</i>
MEP	<i>Multipole Expansion Method</i>
MM	Mécanique Moléculaire
MO	Orbitale Moléculaire
MRCI	<i>Multi-Reference Configuration-Interaction</i>
MP2	<i>Second order of the Møller Plesset Perturbation Theory</i>
NCM	<i>Nancy Continuum Model</i>
OA	Orbitale de fonction de base atomique
OL	Orbitale Localisée
OLO	Orbitale Localisée Orthogonale
ONIOM	<i>Our N-layered Integrated Molecular Orbital and Molecular Mechanic</i>
PCM	<i>Polarizable Continuum Model</i>
PDB	Banque de données de structures protéiques (<i>Protein Data Bank</i>)
PYP	<i>Photoactive Yellow Protein</i>
QM	Mécanique Quantique (<i>quantum mechanics</i>)

RA	<i>Radical anion</i>
SAS	Surface Accessible au Solvant
SCCO	<i>Self-Consistent Core Orbital</i>
SCF	Champ auto-cohérent <i>Self-Consistent Field</i>
SCAPC	<i>Self-Consistent Atomic point Charges</i>
SCRf	Champ de réaction auto-cohérent <i>Self-Consistent Reaction Field</i>
SE	Semi-empirique
SES	Surface Excluant le Solvant
SLBO	<i>Strictly Localised Bond Orbital</i>
SS	pont Soufre-Soufre
ST	Squaraine-Tétralactame
TD-DFT	<i>Time-dependent Density Functional Theory</i>
TD-HF	<i>Time-dependent Hartree-Fock</i>
vdW	van der Waals
WP	Weinstein-Pauncz

Bibliographie

- [1] A. Warshel et M. Levitt, *Journal of molecular biology* **103**, 227 (1976).
- [2] V. Théry, D. Rinaldi, J. L. Rivail, B. Maigret, et G. G. Ferenczy, *J. Comput. Chem.* **15**, 269 (1994).
- [3] G. Monard, M. Loos, V. Thery, K. Baka, et J.-L. Rivail, *Int. J. Quantum Chem.* **58**, 153 (1996).
- [4] X. Assfeld et J.-L. Rivail, *Chem. Phys. Lett.* **263**, 100 (1996).
- [5] N. Ferré, X. Assfeld, et J.-L. Rivail, *J. Comput. Chem.* **23**, 610 (2002).
- [6] A. Fornili, Y. Moreau, M. Sironi, et X. Assfeld, *J. Comput. Chem.* **27**, 515 (2006).
- [7] A. Fornili, P.-F. Loos, M. Sironi, et X. Assfeld, *Chem. Phys. Lett.* **427**, 236 (2006).
- [8] P. Loos et X. Assfeld, *J. Chem. Theory Comput.* **3**, 1047 (2007).
- [9] P.-F. Loos et X. Assfeld, *AIP Conf. Proc.* **963**, 308 (2007).
- [10] P.-F. Loos, A. Fornili, M. Sironi, et X. Assfeld, *Comput.Lett.* **4**, 473 (2007).
- [11] N. Ferré, Ph.D. thesis, Laboratoire de Chimie Théorique, UMR CNRS-UHP 7565, Université Henri Poincaré, Nancy I, 2001.
- [12] P.-F. Loos, Ph.D. thesis, Laboratoire de Chimie Théorique, UMR CNRS-UHP 7565, Université Henri Poincaré, Nancy I, 2008.
- [13] A. Szabo et N. Ostlund, *Modern quantum chemistry* (McGraw-Hill, New York, 1989).
- [14] F. Jensen, *Introduction to Computational Chemistry* (Wiley, Chichester, 1999).
- [15] M. Sringborg, *Methods of Electronic-structure Calculations : From Molecules to Solids* (Wiley, Chichester, 2000).
- [16] C. J. Cramer, *Essentials of Computational Chemistry : Theories and Models* (Wiley, Chichester, 2002).

- [17] J. C. Tully, JCP **93**, 141061 (1990).
- [18] E. Uggerud et T. Helgaker, J. Am. Chem. Soc. **114**, 4265 (1992).
- [19] R. Car et M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
- [20] M. Tuckerman et M. Parrinello, J. Chem. Phys. **101**, 1302 (1994).
- [21] A. Laio, J. VandeVondele, et U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
- [22] C. Møller et M. S. Plesset, Physical Review **46**, 618 (1934).
- [23] B. G. Johnson, P. M. W. Gill, et J. A. Pople, J. Chem. Phys. **98**, 5612 (1993).
- [24] S. J. Vosko, L. Wilk, et M. Nusair, Can. J. Phys. **58**, 1200 (1980).
- [25] A. Becke, Phys. Rev. B **38**, 3098 (1988).
- [26] J. P. Perdew et Y. Wang, Phys. Rev. B **33**, 8800 (1986).
- [27] J. P. Perdew et Y. Wang, Phys. Rev. B **45**, 13244 (1992).
- [28] C. Lee, W. Yang, et R. G. Parr, Phys. Rev. B **37**, 785 (1988).
- [29] T. Van Hoohris et G. Scuseria, J. Chem. Phys. **109**, 400 (1998).
- [30] A. D. Boese et N. C. Handy, J. Chem. Phys. **116**, 9559 (2002).
- [31] J. Tao, J. Perdew, V. Staroverov, et G. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).
- [32] J. Sacho-Garca et J. Cornil, J. Chem. Phys. **121**, 3096 (2004).
- [33] C. Adamo et V. Barone, J. Chem. Phys. **110**, 6158 (1999).
- [34] M. Ernzerhof et G. E. Scuseria, J. Chem. Phys. **110**, 5029 (1999).
- [35] A. Becke, J. Chem. Phys. **98**, 5648 (1993).
- [36] M. A. L. Marques, U. C. A., F. Nogueira, A. Rubio, K. Burke, et E. K. U. Gross, *Time-Dependent Density Functional Theory*, Vol. 706 of *Lecture Notes in Physics* (Springer-Verlag, Heidelber, 2006).
- [37] T. Leininger, H. Stoll, H. J. Werner, et A. Savin, Chem. Phys. Lett. **275**, 151 (1997).
- [38] T. Leininger, H. Stoll, H. J. Werner, et A. Savin, Chem. Phys. Lett. **275**, 151 (1997).
- [39] J. P. Adamson, R. D. Dombroski et P. M. W. Gill, Chem. Phys. Lett. **254**, 329 (1996).
- [40] T. Yanai, D. P. Tew, et N. C. Handy, Chem. Phys. Lett. **393**, 51 (2004).
- [41] E. Rudberg, P. Salek, T. Helgaker, et H. Agren, J. Chem. Phys. **123**, 184108 (2005).

- [42] M. J. G. Peach, T. Helgaker, P. Salek, T. W. Keal, O. B. Lutnaes, D. J. Tozer, et N. C. Handy, *Phys. Chem. Chem. Phys.* **8**, 558 (2006).
- [43] H. Iikura, T. Tsuneda, T. Yanai, et K. Hirao, *J. Chem. Phys.* **115**, 3540 (2001).
- [44] O. A. Vydrov et G. E. Scuseria, *J. Phys. Chem.* **125**, 234109 (2006).
- [45] D. Jacquemin, E. A. Perpète, G. E. Scuseria, I. Ciofini, et C. Adamo, *J. Chem. Theory Comput.* **4**, 123 (2008).
- [46] D. Jacquemin, V. Wathelet, E. A. Perpète, et C. Adamo, *J. Chem. Theory Comput.* **25**, 2420 (2009).
- [47] M. Rohrdanz, K. Martins, et J. Herbert, *J. Chem. Phys.* .
- [48] P. Wiggins, J. Williams, et D. Tozer, *The Journal of Chemical Physics* **131**, 091101 (2009).
- [49] E. Perpète et D. Jacquemin, *J. Mol. Struct. (THEOCHEM)* , in press.
- [50] M. Caricato, G. Trucks, M. Frisch, et K. Wiberg, *J. Chem. Theory Comput.* **6**, 370 (2010).
- [51] E. Dumont, A. D. Laurent, A. Assfeld, et D. Jacquemin, *J. Phys. Chem. A* (2010), submitted.
- [52] S. Kristyán et P. Pulay, *Chem. Phys. Lett.* **229**, 175 (1994).
- [53] M. Allen et D. Tozer, *J. Chem. Phys.* **117**, 11113 (2002).
- [54] U. Zimmerli, M. Parrinello, et P. Koumoutsakos, *J. Chem. Phys.* **120**, 2693 (2004).
- [55] S. Grimme, *J. Comput. Chem.* **25**, 1463 (2004).
- [56] A. Tkatchenko et W. Scheffler, *Phys. Rev. Lett.* **102**, 073005 (2009).
- [57] A. Tkatchenko, R. A. DiStasio, M. Head-Gordon, et W. Scheffler, *J. Chem. Phys.* **131**, 094106 (2009).
- [58] I. Mackie et G. DiLabio, *J. Phys. Chem. A* **112**, 10968 (2008).
- [59] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, et E. Kaxiras, *J. Chem. Phys.* **114**, 5149 (2001).
- [60] J. Antony et S. Grimme, *Phys. Chem. Chem. Phys.* **8**, 5287 (2006).
- [61] T. Schwabe et S. Grimme, *Phys. Chem. Chem. Phys.* **9**, 3397 (2007).
- [62] S. Grimme, C. Diedrich, et M. Korth, *Angew Chem Int Ed Engl* **45**, 625 (2006).
- [63] V. Barone, M. Biczysko, et M. Pavone, *Chem. Phys.* **346**, 247 (2008).

- [64] A. Dreuw et M. Head-Gordon, *Chem. Rev.* **105**, 4009 (2005).
- [65] E. R. Davidson, *J. Chem. Phys.* **17**, 87 (1975).
- [66] J. B. Foresman, M. Head-Gordon, J. A. Pople, et M. J. Frisch, *J. Phys. Chem.* **96**, 135 (1992).
- [67] S. Hirata, M. Head-Gordon, et R. J. Bartlett, *J. Chem. Phys.* **111**, 10774 (1999).
- [68] J. F. Stanton, J. Gauss, N. Ishikawa, et M. Head-Gordon, *J. Chem. Phys.* **103**, 41160 (1995).
- [69] W. Thomas, *Naturwissenschaften* **13**, 627 (1925).
- [70] F. Reiche et W. Thomas, *Z. Phys* **34**, 510 (1925).
- [71] W. Kuhn, *Z. Phys.* **33**, 408 (1925).
- [72] R. Shepard, *Advances in Chemical Physics : ab initio Methods in Quantum Chemistry (Part II)* (John Wiley & Sons, Chichester, England, 1987), Vol. 69, p. 63.
- [73] H. J. Werner, *Advances in Chemical Physics :ab initio Methods in Quantum Chemistry (Part II)* (John Wiley & Sons, Chichester, England, 1987), Vol. 69, p. 1.
- [74] P. Bruna, S. Peyerimhoff, et R. Buenker, *Chem. Phys. Lett.* **72**, 278 (1980).
- [75] B. O. Roos, *Advances in Chemical Physics : ab initio Methods in Quantum Chemistry (Part II)* (John Wiley & Sons, Chichester, England, 1987), Vol. 69, p. 399.
- [76] W. A. Goddard, T. H. Dunning, W. J. Hunt, et P. J. Hey, **6**, 368 (1973).
- [77] J. Olsen, B. O. Roos, P. Jørgensen, et H. J. A. Jensen, *J. Chem. Phys.* **89**, 2185 (1988).
- [78] A. Migani, A. Sinicropi, N. Ferré, A. Cembran, M. Garavelli, et M. Olivucci, *Faraday Discuss* **127**, 179 (2004).
- [79] N. Ferre, A. Cembran, M. Garavelli, et M. Olivucci, *Theor. Chem. Acc.* **112**, 335 – (2004).
- [80] A. Strambi, P. B. Coto, N. Ferré, et M. Olivucci, *Theor. Chem. Acc.* **118**, 185 (2007).
- [81] L. M. Frutos, T. Andruniów, F. Santoro, N. Ferré, et M. Olivucci, *Proc. Natl. Acad. Sci.* **104**, 7764 (2007).
- [82] P. B. Coto, A. Strambi, N. Ferré, et M. Olivucci, *Proc. Natl. Acad. Sci.* **103**, 17154 (2006).
- [83] A. Cembran, R. Gonzalez-Luque, P. Altoè, M. Merchan, F. Bernardi, M. Olivucci, et M. Garavelli, *J. Phys. Chem. A* **109**, 6597 (2005).

- [84] T. Andruniow, N. Ferré, et M. Olivucci, *Proc. Natl. Acad. Sci.* **101**, 17908 (2004).
- [85] M. Olivucci, A. Lami, et F. Santoro, *Angew. Chem. Int. Ed. Engl.* **44**, 5118 (2005).
- [86] M. A. L. Marques et E. K. U. Gross, *Annu. Rev. Phys. Chem.* **55**, 427 (2004).
- [87] A. Sinicropi, T. Andruniow, N. Ferré, R. Basosi, et M. Olivucci, *J. Am. Chem. Soc.* **33**, 11534 (2005).
- [88] A. Sinicropi, R. Basosi, et M. Olivucci, *J. Phys. : Conf. Ser.* **101**, 012001 (2008).
- [89] A. Strambi, P. Coto, L. Frutos, N. Ferré, et M. Olivucci, *J. Am. Chem. Soc.* **130**, 3382 (2008).
- [90] G. Karlström, R. Lindh, P.-A. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogady, et L. Seijo, *Computational Material Science* **28**, 222 (2003).
- [91] D. Jacquemin, J. Preat, V. Wathelet, M. Fontaine, et E. A. Perpète, *J. Am. Chem. Soc.* **128**, 2072 (2006).
- [92] D. Jacquemin et E. A. Perpète, *Chem. Phys. Lett.* **420**, 529 (2006).
- [93] D. Jacquemin, A. D. Laurent, E. A. Perpète, et J. M. André, *Int. J. Quantum Chem.* **901**, 24 (2009).
- [94] D. Jacquemin, E. A. Perpète, I. Ciofini, et C. Adamo, *Acc. Chem. Res.* **42**, 326 (2009).
- [95] A. D. Laurent, J.-M. André, E. A. Perpète, et D. Jacquemin, *Chem. Phys. Lett.* **436**, 84 (2007).
- [96] P. Loos, J. Preat, A. Laurent, C. Michaux, D. Jacquemin, E. Perpète, et X. Assfeld, *J. Chem. Theory Comput.* **4**, 637 (2008).
- [97] J. Preat, Ph.D. thesis, Laboratoire de Chimie Théorique Appliquée, faculté Universitaire Notre Dame de la Paix, Académie de Louvain, 2008.
- [98] A. D. Laurent, J.-M. André, E. A. Perpète, et D. Jacquemin, *J. Photochem. Photobiol. A : Chem.* **192**, 211 (2007).
- [99] A. D. Laurent, D. Jacquemin, X. Assfeld, et E. A. Perpète, *Molecular Simulation* **36**, 74 (2009).
- [100] A. Perrier, F. Maurel, et J. Aubard, *J. Phys. Chem. A* **111**, 167 (2007).

- [101] E. A. Perpète et D. Jacquemin, *J. Photochem. Photobiol. A : Chem.* **187**, 40 (2007).
- [102] E. A. Perpète, F. Maurel, et D. Jacquemin, *J. Phys. Chem. A* **111**, 5528 (2007).
- [103] D. Jacquemin et E. A. Perpète, *Chem. Phys. Lett.* **429**, 147 (2006).
- [104] B. Mennucci, M. Caricato, F. Ingrosso, C. Cappelli, R. Cammi, J. Tomasi, G. Scalmani, et M. Frisch, *J. Phys. Chem. B* **112**, 414 (2008).
- [105] B. Mennucci, C. Cappelli, C. A. Guido, R. Cammi, et J. Tomasi, *J. Phys. Chem. A* **113**, 3009 (2009).
- [106] V. Russo, C. Curutchet, et B. Mennucci, *J. Phys. Chem. B* **111**, 853 (2007).
- [107] G. Scalmani, M. J. Frisch, B. Mennucci, J. Tomasi, R. Cammi, et V. Barone, *J. Chem. Phys.* **124**, 094107 (2006).
- [108] M. A. L. Marques, X. López, D. Varsano, A. Castro, et A. Rubio, *Phys. Rev. Lett.* **90**, 258101 (2003).
- [109] X. Lopez, M. Marques, A. Castro, et A. Rubio, *J. Am. Chem. Soc.* **127**, 12329 (2005).
- [110] M. Wanko, M. Hoffmann, T. Frauenheim, et M. Elstner, *J. Phys. Chem. B* **112**, 11462 (2008).
- [111] J. Y. Hasegawa, T. Ise, K. J. Fujimoto, A. Kikuchi, E. Fukumura, A. Miyawaki, et Y. Shiro, *J. Phys. Chem. B* (2010), in press.
- [112] D. Jacquemin, E. A. Perpète, A. D. Laurent, X. Assfeld, et C. Adamo, *Phys. Chem. Chem. Phys.* **11**, 1258 (2009).
- [113] A. Laurent et X. Assfeld, *Interdiscip. Sci. Comput. Life Sci.* **2**, 38 (2010).
- [114] A. D. Laurent et X. Assfeld, *J. Chem. Theory Comput.* (2009), in revision.
- [115] A. L. Fetter et J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
- [116] P. Ring et P. Schuck, *The nuclear many-body problem.* (Springer-Verlag, Heidelberg, Germany, 1980).
- [117] A. D. McLachlan et M. A. Ball, *Rev. Mod. Phys.* **36**, 844 (1964).
- [118] E. Runge et E. K. U. Gross, *Phys. Rev. Lett.* **52**, 997 (1984).
- [119] R. van Leeuwen, *Int. J. Mod. Phys. B* **15**, 1969 (2001).

- [120] J. Preat, A. D. Laurent, C. Michaux, E. A. Perpète, et D. Jacquemin, *J. Mol. Struct. (THEOCHEM)* **901**, 24 (2009).
- [121] V. Barone, R. Improta, et N. Rega, **41**, 605 (2008).
- [122] J. Bloino, M. Biczysko, O. Crescenzi, et V. Barone, *J. Chem. Phys.* **128**, 244105 (2008).
- [123] V. Barone, J. Bloino, M. Biczysko, et F. Santoro, *J. Chem. Theory Comput.* **5**, 540 (2009).
- [124] V. Barone, M. Casarin, D. Forrer, M. Pavone, M. Sambri, et A. Vittadini, *J. Comput. Chem.* **30**, 934 (2009).
- [125] M. Dierksen et S. Grimme, *J. Chem. Phys.* **120**, 3544 (2004).
- [126] C. J. Cramer et D. G. Truhlar, *Chem. Rev.* **99**, 2161 (1999).
- [127] I. Geronimo, N. Chéron, P. Fleurat-Lessard, et E. Dumont, *Chem. Phys. Lett.* **481**, 173 (2009).
- [128] C. Michaux, J. Wouters, E. A. Perpète, et D. Jacquemin, *J. Phys. Chem. B* **112**, 9896 (2008).
- [129] C. Michaux, J. Wouters, E. A. Perpète, et D. Jacquemin, *J. Phys. Chem. B* **112**, 7702 (2008).
- [130] C. Michaux, J. Wouters, E. A. Perpète, et D. Jacquemin, *J. Phys. Chem. B* **112**, 2430 (2008).
- [131] J. Tomasi, B. Mennucci, et R. Cammi, *Chem. Rev.* **105**, 2999 (2005).
- [132] J. Tomasi et M. Persico, *Chem. Rev.* **94**, 2027 (1994).
- [133] D. Rinaldi et J.-L. Rivail, *Theor. Chim. Acta.* **32**, 57 (1973).
- [134] J.-L. Rivail et D. Rinaldi, *Chem. Phys.* **18**, 233 (1976).
- [135] D. Rinaldi, M. F. Ruiz-Lopez, et J.-L. Rivail, *J. Chem. Phys.* **78**, 834 (1983).
- [136] V. Dillet, D. Rinaldi, et J.-L. Rival, *J. Phys. Chem.* **98**, 5034 (1994).
- [137] D. Rinaldi, A. Bouchy, et J.-L. Rivail, *J. Chem. Phys.* **120**, 2343 (2004).
- [138] D. Rinaldi, A. Bouchy, et J.-L. Rivail, *Theor. Chim. Acta.* **116**, 664 (2006).
- [139] M. Mennucci, R. Cammi, et J. Tomasi, *J. Phys. Chem.* **110**, 6858 (1999).
- [140] M. Mennucci, E. Cancès, et J. Tomasi, *J. Phys. Chem.* **101**, 10506 (1997).
- [141] J. Tomasi, B. Mennucci, et E. Cancès, *J. Mol. struct.(THEOCHEM)* **464**, 2999 (1999).

- [142] A. Bondi, *J. Chem. Phys.* **68**, 441 (1964).
- [143] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, 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, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, et J. A. Pople, *Gaussian 03, Revision C.02*, Gaussian, Inc., Wallingford, CT, 2004.
- [144] E. Cancès, B. Mennucci, et J. Tomasi, *J. Chem. Phys.* **107**, 3032 (1997).
- [145] E. Cancès et B. Mennucci, *J. Math. Chem.* **23**, 309 (1998).
- [146] J. Preat, P.-F. Joos, X. Assfeld, D. Jacquemin, et E. A. Perpète, *Int. J. Quantum Chem.* **107**, 574 (2007).
- [147] H. H. J. Uhlig, *Phys. Chem.* **41**, 1215 (1937).
- [148] I. Tunon, E. Silla, et J. L. Pascual-Ahuir, *Chem. Phys. Lett.* **203**, 289 (1993).
- [149] F. London, *Z. Phys.* **60**, 245 (1930).
- [150] R. McWeeny, *Methods of Molecular Quantum Mechanics* (Academic Press, London, 1992).
- [151] C. Amovilli et B. Mennucci, *J. Phys. Chem.* **101**, 1051 (1997).
- [152] E. Condon, *Phys. ReV.* **28**, 1182 (1926).
- [153] J. Franck, *Trans. Faraday Soc.* **21**, 536 (1926).
- [154] E. Condon, *Phys. ReV.* **32**, 858 (1928).
- [155] M. F. Iozzi, B. Mennucci, J. Tomasi, et R. Cammi, *J. Chem. Phys.* **120**, 7029 (2004).
- [156] M. F. Ruiz-López, D. Rinaldi, et J. Bertrán, *J. Chem. Phys.* **103**, 9249 (1995).

- [157] X. Assfeld, J. Garapon, D. Rinaldi, M. Ruiz-López, et J. Rivail, *J. Mol. Struct. (THEOCHEM)* **371**, 107 (1996).
- [158] Y. Shiratori et S. J. Nakagawa, *J. Comput. Chem.* **12**, 717 (1991).
- [159] E. Project, E. Nachliel, et M. Gutman, *J. Comput. Chem.* **29**, 1163 (2008).
- [160] P. L. Freddolino, F. Liu, M. Gruebele, et K. Schulten, *Biophys. J.* **94**, 75 (2008).
- [161] P. L. Freddolino, S. Park, B. Roux, et K. Schulten, *Biophys. J.* **96**, 3772 (2009).
- [162] C. Czaplewski, S. Oldziej, A. Liwo, et H. Scheraga, *Protein Engineering Design and Selection* **17**, 29 (2004).
- [163] A. Van Duin, S. Dasgupta, F. Lorant, et W. Goddard III, *J. Phys. Chem. A* **105**, 9396 (2001).
- [164] W. D. Cornell, P. Cieplak, C. Bayly, I. Gould, K. M. Merz, D. C. Ferguson, Jr. D. M. Spellmeyer, T. Fox, J. W. Caldwell, et P. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).
- [165] J. Wang, R. M. Wolf, J. Caldwell, P. A. Kollman, et D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- [166] B. R. Brooks, R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan, et M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- [167] A. MacKerell Jr., D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, et M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).
- [168] A. D. MacKerell Jr., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, et M. Karplus, *The Encyclopedia of Computational Chemistry*; (John Wiley & Sons, Chichester, 1998), Vol. 1, p. 271.
- [169] N. Foloppe et A. MacKerell, *J. Comput. Chem.* **21**, (2000).
- [170] W. Xie, J. Pu, A. MacKerell Jr, et J. Gao, *J. Chem. Theory Comput* **3**, 1878 (2007).
- [171] P. Ren et J. Ponder, *J. Comput. Chem.* **23**, 1497 (2002).
- [172] P. Ren et J. W. Ponder, *J. Phys. Chem. B* **107**, 5933 (2003).
- [173] J. W. P. Pengyu Ren and, *J. Phys. Chem. B* **108**, 13427 (2004).

- [174] J. Piquemal, L. Perera, G. Cisneros, P. Ren, L. Pedersen, et T. Darden, *J. Chem. Phys.* **125**, 054511 (2006).
- [175] J. Ponder, C. Wu, P. Ren, V. Pande, J. Chodera, M. Schnieders, I. Haque, D. Mobley, D. Lambrecht, R. DiStasio Jr, *et al.*, (2010).
- [176] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, et M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- [177] G. Lamoureux, A. MacKerell Jr, et B. Roux, *J. Chem. Phys.* **119**, 5185 (2003).
- [178] G. Lamoureux, E. Harder, I. Vorobyov, B. Roux, et A. MacKerell, *Chem. Phys. Lett.* **418**, 245 (2006).
- [179] S. Noskov, G. Lamoureux, et B. Roux, *J. Phys. Chem. B* **109**, 6705 (2005).
- [180] V. Anisimov, G. Lamoureux, I. Vorobyov, N. Huang, B. Roux, et A. MacKerell Jr, *J. Chem. Theory Comput* **1**, 153 (2005).
- [181] P. Lopes, G. Lamoureux, et A. Mackerell Jr, *J. Comput. Chem.* **30**, (2009).
- [182] E. Harder, A. MacKerell Jr, et B. Roux, *J. Am. Chem. Soc* **131**, 2760 (2009).
- [183] A. Rappe, C. J. Casewit, K. S. Colwell, G. W. A., et S. W. M., *J. Am. Chem. Soc.* **114**, 10024–35 (1992).
- [184] S. Patel et C. Brooks III, *J. Comput. Chem.* **25**, (2004).
- [185] S. Patel, A. D. J. Mackerell, et B. C. L. III, *J. Comput. Chem.* **25**, 1504 (2004).
- [186] P. Cieplak, F. Dupradeau, Y. Duan, et J. Wang, *Journal of Physics : Condensed Matter* **21**, 333102 (2009).
- [187] P. Lopes, B. Roux, et A. MacKerell, *Theor. Chem. Acc.* **124**, 11 (2009).
- [188] A. Warshel, M. Kato, et A. V. Pislakov, *J. Chem. Theory Comput.* **3**, 2034 (2007).
- [189] A. Flores-Ortega, J. Casanovas, D. Zanuy, R. Nussinov, et C. Aleman, *J. Phys. Chem. B* **111**, 5475 (2007).
- [190] A. Altun, S. Yokoyama, et K. Morokuma, *J. Phys. Chem. B* .
- [191] M. Concistrè, A. Gansmüller, N. McLean, O. Johannessen, I. Montesinos, P. Bovee-Geurts, P. Verdegem, J. Lugtenburg, R. Brown, W. DeGrip, *et al.*, *J. Am. Chem. Soc* **130**, 10490 (2008).

- [192] S. Sekharan et K. Morokuma, *J. Phys. Chem. Lett.* **1**, 668 (2010).
- [193] B. H. Besler, K. M. Merz, et P. A. Kollman, *J. Comp. Chem.* **11**, 431 (1990).
- [194] U. C. Singh et P. A. Kollman, *J. Comput. Chem.* **5**, 129 (1984).
- [195] L. E. Chirlian et M. L. Francl, *J. Comput. Chem.* **8**, 894 (1987).
- [196] C. M. Breneman et K. B. Wiberg, *J. Comput. Chem.* **11**, 361 (1990).
- [197] F.-Y. Dupradeau, C. Cezard, R. Lelong, E. Stanislawiak, J. Pecher, J. C. Delepine, et P. Cieplak, *Nucl. Acids Res. (Database issue)* D360 (2008).
- [198] F. Maseras et K. Morokuma, *J. Comput. Chem.* **16**, 1170 (1995).
- [199] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, et K. Morokuma, *J. Phys. Chem.* **100**, 19357 (1996).
- [200] S. Dapprich, I. Komromi, K. S. Byun, K. Morokuma, et M. J. Frisch, *J. Mol. Struct. (THEOCHEM)* **461-462**, 1 (1999).
- [201] T. Vreven et K. Morokuma, *Theor. Chem. Acc.* **109**, 125 (2003).
- [202] T. Vreven, K. S. Byun, I. Komromi, S. Dapprich, J. A. Montgomery Jr., K. Morokuma, et M. J. Frisch, *J. Chem. Theory Comput.* **2**, 815 (2006).
- [203] N. Reuter, A. Dejaegere, B. Maignet, et M. Karplus, *J. Phys. Chem. A* **104**, 1720 (2000).
- [204] D. Bakowies et W. Thiel, *J. Phys. Chem* **100**, 10580 (1996).
- [205] N. Ferré et M. Olivucci, *J. Mol. Struct. (THEOCHEM)* **632**, 71 (2003).
- [206] I. Antes et W. Thiel, *J. Phys. Chem. A* **103**, 9290 (1999).
- [207] Y. Zhang, T.-S. Lee, et W. Yang, *J. Chem. Phys.* **110**, 46 (1999).
- [208] Y. Zhang, *J. Chem. Phys.* **122**, 024114 (2005).
- [209] Y. Zhang, *Theor. Chem. Acc.* **116**, 43 (2006).
- [210] G. A. DiLabio, M. M. Hurley, et P. A. Christiansen, *J. Chem. Phys.* **116**, 9578 (2002).
- [211] S. Humbel, S. Sieber, et K. Morokuma, *J. Chem. Phys.* **105**, 1959– (1996).
- [212] T. Matsubara, S. Sieber, et K. Morokuma, *Int. J. Quantum Chem.* **60**, 1101 (1996).
- [213] D. Philipp et R. Friesner, *J. Comput. Chem.* **20**, 1468 (1999).
- [214] R. Murphy, D. Philipp, et R. Friesner, *Journal of Computational Chemistry* **21**, 1442 (2000).

- [215] J. Gao, P. Amara, C. Alhambra, et M. J. Field, *J. Phys. Chem. A* **102**, 4714 (1998).
- [216] P. Amara, M. J. Field, C. Alhambra, et J. Gao, *Theor. Chem. Acc.* **104**, 336 (2000).
- [217] J. Pu, J. Gao, et D. T. Truhlar, *J. Phys. Chem. A* **108**, 632 (2004).
- [218] J. Pu, J. Gao, et D. Truhlar, *J. Phys. Chem. A* **108**, 5454 (2004).
- [219] J. Pu, J. Gao, et D. Truhlar, *Chem. Phys. Chem.* **6**, 1853 (2005).
- [220] V. Kairys et J. Jensen, *J. Phys. Chem. A* **104**, 6656 (2000).
- [221] M. S. Gordon, M. Freitag, P. Bandyopadhyay, J. Jensen, V. Kairys, et W. J. Stevens, *J. Phys. Chem. A* **105**, 293 (2001).
- [222] H. Weinstein et R. Pauncz, *Symp. Faraday Soc.* **2**, 23 (1968).
- [223] H. Weinstein, R. Pauncz, et M. Cohen, *Adv Atom Mol Phys* **7**, 97 (1971).
- [224] V. Magnasco et A. Perico, *J. Chem. Phys.* **47**, 971 (1967).
- [225] V. Magnasco et A. Perico, *J. Chem. Phys.* **48**, 800 (1968).
- [226] J. Pipek et P. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).
- [227] S. F. Boys, *Proc. Roy. Soc.* **200**, 542 (1950).
- [228] J. M. Foster et S. F. Boys, *Rev. Mod. Phys.*, **32**, 300 (1960).
- [229] C. Edmiston et K. Ruedenberg, *Rev. Mod. Phys.* **35**, 457 (1963).
- [230] C. Edmiston et K. Ruedenberg, *J. Chem. Phys.* **43**, S97 (1965).
- [231] P.-F. Loos, E. Dumont, A. D. Laurent, et X. Assfeld, *Chem. Phys. Lett.* **475**, 120 (2003).
- [232] J. Gu, Y. Xie, et H. F. Schaefer III, *J. Am. Chem. Soc.* **128**, 1250 (2005).
- [233] J. Gu, J. Wang, et J. Leszczynski, *J. Am. Chem. Soc.* **128**, 9322 (2006).
- [234] J. M. Mullin et M. S. Gordon, *J. Phys. Chem. B* **113**, 8657 (2009).
- [235] J. M. Mullin et M. S. Gordon, *J. Phys. Chem. B* **113**, 1441 (2009).
- [236] J.-R. Lalanne, F. Carmora, et L. Servant, *Spectroscopies optiques d'absorption électronique*. (Masson S. A., Paris, 1998).
- [237] B. Valeur, *Invitation à la fluorescence moléculaire* (De Boeck Université, Bruxelles, 2004).
- [238] A. D. Laurent, V. Wathelet, M. Bouhy, D. Jacquemin, et E. A. Perpète, *Technique de l'Ingénieur*, in press.

- [239] C. Curutchet, G. D. Scholes, B. Mennucci, et R. Cammi, *J. Phys. Chem. B* **111**, 13253 (2007).
- [240] S. Nakamura et M. Irie, *J. Org. Chem.* **53**, 6136 (1988).
- [241] H. Houjou, Y. Inoue, et M. Sakurai, *J. Am. Chem. Soc.* **120**, 4459 (1998).
- [242] H. Houjou, Y. Inoue, et M. Sakurai, *J. Phys. Chem. B* **105**, 867 (2001).
- [243] M. Dupuis, M. Aida, Y. Kawashima, et K. A. Hirao, *J. Chem. Phys.* **117**, 1242 (2002).
- [244] M. Elstner, D. Porezag, D. Jungnickel, J. Elsner, et T. S. S. S. G. Haugk, M. and Frauenheim, *Phys. Rev. B* **58**, 7276 (1998).
- [245] M. Elstner, *J Phys Chem A* **111**, 5614 (2007).
- [246] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, et M. Karplus, *J. Phys. Chem. B* **105**, 569 (2001).
- [247] M. Wanko, M. Hoffmann, J. Frähmcke, T. Frauenheim, et M. Elstner, *J. Phys. Chem. B* **112**, 11468 (2008).
- [248] P. Söderhjelm, C. Husberg, S. A., M. Olivucci, et U. Ryde, *J. Chem. Theory Comput.* **30**, 649 (2009).
- [249] E. Arunkumar, C. Forbes, B. Noll, et B. Smith, *J. Am. Chem. Soc* **127**, 3288 (2005).
- [250] J. Gassensmith, E. Arunkumar, L. Barr, J. Baumes, K. DiVittorio, J. Johnson, B. Noll, et B. Smith, *J. Am. Chem. Soc* **129**, 15054 (2007).
- [251] E. Arunkumar, P. K. Sudeep, P. V. Kamat, B. C. Noll, et B. Smith, *New J. Chem.* **31**, 677 (2007).
- [252] M. Cossi et V. Barone, *J. Chem. Phys.* **115**, 4708 (2001).
- [253] K. Srinivas, C. Prabhakar, C. Devi, K. Yesudas, K. Bhanuprakash, et V. Rao, *J. Phys. Chem. A* **111**, 3378 (2007).
- [254] D. Jacquemin, E. A. Perpète, G. Scalmani, M. J. Frisch, X. Assfeld, I. Ciofini, et C. Adamo, *J. Chem. Phys.* **125**, 164324 (2006).
- [255] D. Jacquemin, V. Wathelet, et E. A. Perpète, *J. Phys. Chem. A* **110**, 9145 (2006).
- [256] E. Dumont, A. D. Laurent, et A. Assfeld, *J. Phys. Chem. Lett.* **1**, 581 (2010).
- [257] W. R. Rocha, K. J. De Almeida, K. Coutinho, et S. Canuto, *Chem. Phys. Lett.* **345**, 171 (2001).

- [258] O. Shimomura, F. Johnson, et Y. Saiga, *J. cell. Physiol.* **59**, 223 (1974).
- [259] D. Parasher, R. O. McCann, et M. J. Cormier, *Biochemical and Biophysical Research Communications* **126**, 1259 (1985).
- [260] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward, et D. Prasher, **263**, 802 (1994).
- [261] Y. R. Tsien, *Annu. Rev. Biochem.* **67**, 509 (1998).
- [262] M. Zimmer, *Chem. Rev.* **102**, 759 (2002).
- [263] G. Patterson, R. Day, et D. Piston, *Fluorescent protein spectra*, 2001.
- [264] B. Pollok et R. Heim, *Trends in Cell Biology* **9**, 57 (1999).
- [265] A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross, et R. Y. Tsien, *Trends in Biochemical Sciences* **58**, 448 (1995).
- [266] G. Nienhaus, *Angew. Chem. Int. Ed. Engl.* **47**, 8992 (2008).
- [267] O. Shimomura, *Angewandte Chemie (International ed. in English)* (2009).
- [268] R. Tsien, *Angew. Chem. Int. Ed. Engl.* **48**, 5612 (2009).
- [269] M. Chalfie, *Proc. Natl. Acad. Sci.* **106**, 10073 (2009).
- [270] O. Shimomura, *Febs Letters* **104**, 220 (1979).
- [271] C. Cody, D. Prasher, W. Westler, F. Prendergast, et W. Ward, *Biochemistry* **32**, 1212 (1993).
- [272] R. Heim, D. Prasher, et R. Tsien, *Proc. Natl. Acad. Sci.* **92**, 12501 (1994).
- [273] N. Lemay, A. Morgan, E. Archer, L. Dickson, C. Megley, et M. Zimmer, *Chem. Phys.* (2008).
- [274] M. Ormö, A. Cubitt, K. Kallio, L. Gross, R. Tsien, et S. Remington, **273**, 1392 (1996).
- [275] S. Remington, **16**, 714 (2006).
- [276] R. Wachter, M. A. Elsliger, K. Kallio, G. T. Hanson, et S. Remington, *Structure with Folding and Design* **6**, 1267 (1998).
- [277] R. Wachter, **40**, 4767 (2007).
- [278] L. P. Zhang, H.-N. Patel, J. W. Lappe, et R. Wachter, *J. Am. Chem. Soc.* **128**, 4766 (2006).
- [279] V. Helms, T. Straatsma, et J. McCammon, *J. Phys. Chem. B* **103**, 3263 (1999).

- [280] I. Demachy, J. Ridard, H. Laguitton-Pasquier, G. Durnerin, E. an dVallverdu, P. Archirel, et B. Lévy, *J. Phys. Chem. B* **109**, 24121 (2005).
- [281] M. Andresen, A. Stiel, S. Trowitzsch, G. Weber, C. Eggeling, M. Wahl, S. Hell, et S. Jakobs, *Proc. Natl. Acad. Sci.* **104**, 13005 (2007).
- [282] R. Ando, H. Mizuno, et A. Miyawaki, *Science* **306**, 1370 (2004).
- [283] K. A. Lukyanov, A. F. Fradkov, N. G. Gurskaya, M. V. Matz, Y. A. Labas, A. P. Savitsky, M. L. Markelov, A. G. Zaraisky, X. Zhao, Y. Fang, W. Tan, et S. A. Lukyanov, *J. Biol. Chem.* **275**, 25879 (2000).
- [284] J. Wiedenmann, S. Ivanchenko, F. Oswald, F. Schmitt, C. Röcker, A. Salih, K. Spindler, et G. Nienhaus, *Proceedings of the National Academy of Sciences* **101**, 15905 (2004).
- [285] V. Adam, M. Lelimosin, S. Boehme, G. Desfonds, K. Nienhaus, M. Field, J. Wiedenmann, S. McSweeney, G. Nienhaus, et D. Bourgeois, *Proc. Natl. Acad. Sci.* .
- [286] R. Heim et R. Y. Tsien, *Curr. Biol.* **6**, 178 (1996).
- [287] J. Huyen Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, K. J.-S., A. Zumbusch, T. A. Holak, L. Moroder, R. Huber, et N. Budisa, *J. Mol. Biol.* **328**, 1071 (2003).
- [288] M. Seifert, D. Ksiazek, M. Azim, P. Smialowski, N. Budisa, et T. Holak, *J. Am. Chem. Soc.* **124**, 7932 (2002).
- [289] M. A. Rizzo, G. H. Springer, B. Granada, et D. W. Piston, *Nat. Biotechnol.* **22**, 445 (2004).
- [290] G. Malo, L. Pouwels, M. Wang, A. Weichsel, W. Montfort, M. Rizzo, D. Piston, et R. Wachter, *Biochemistry* **46**, 9865 (2007).
- [291] G. Vallverdu, I. Demachy, J. Ridard, et B. Lévy, *J. Mol. Struct. (THEOCHEM)* **898**, 73 (2009).
- [292] G. Vallverdu, I. Demachy, F. Mérola, H. Pasquier, J. Ridard, et B. Lévy, *Proteins* **78**, 1040 (2010).
- [293] M. Lelimosin, M. Noirclerc-Savoie, C. Lazareno-Saez, B. Paetzold, S. Le Vot, R. Chazal, P. Macheboeuf, M. Field, D. Bourgeois, et A. Royant, (2009).
- [294] M. Martin, F. Negri, et M. Olivucci, *J. Am. Chem. Soc.* **126**, 5452 (2004).

- [295] E. Epifanovsky, I. Polyakov, B. Grigorenko, A. Nemukhin, et A. I. Krylov, *J. Chem. Theory Comput.* **5**, 1895 (2009).
- [296] I. Polyakov, E. Epifanovsky, B. Grigorenko, A. I. Krylov, et A. Nemukhin, *J. Chem. Theory Comput.* **5**, 1907– (2009).
- [297] D. Case, T. Darden, I. T.E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, K. Merz, D. Pearlman, R. W. M. Crowley, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. Mathews, C. Schafmeister, W. Ross, et P. Kollman, AMBER 9, university of Californiai, San Francisco, 2006.
- [298] F. W. Ponder, *Tinker*, version 4.2 ed. (Washington University, St. louis, MO, 2004).
- [299] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kalé, et K. Schulten, *J. Comput. Chem.* **26**, 1781 (2005).
- [300] H. Li, A. D. Robertson, et J. H. Jensen, *Proteins* **61**, 704 (2005).
- [301] N. Reuter, H. Lin, et W. Thiel, *J. Phys. Chem. B* **106**, 6310 (2002).
- [302] R. Nifosì, P. amat, et V. Tozzini, *J. Comput. Chem.* **28**, 2366 (2007).
- [303] E. Cubero, F. J. Luque, M. Orozco, et J. Gao, *J. Phys. Chem. B* **107**, 1664 (2003).
- [304] E. Dumont, A. D. Laurent, P.-F. Loos, et X. Assfeld, *J. Chem. Theory Comput.* **5**, 1700 (2009), in press.
- [305] E. Dumont, P.-F. Loos, A. D. Laurent, et X. Assfeld, *Int. J. Quantum Chem.* **110**, 513 (2010).
- [306] E. Dumont, P.-F. Loos, et X. Assfeld, *Chem. Phys. Lett.* **458**, 276 (2008).
- [307] E. Dumont, P.-F. Loos, et X. Assfeld, *J. Phys. Chem. B* **112**, 13661 (2008).
- [308] E. Dumont, P.-F. Loos, A. D. Laurent, et X. Assfeld, *J. Chem. Theory Comput.* **4**, 1171 (2009).
- [309] T. Ishida, *J. Phys. Chem. B* **109**, 18558 (2005).
- [310] M. Cascella, A. Magistrato, I. Tavernelli, P. Carloni, et U. Rothlisberger, *Proc. Natl. Acad. Sci.* **103**, 19641 (2006).
- [311] J. Blumberger et M. Sprik, *J. Phys. Chem. B* **109**, 6793 (2005).

- [312] J. Blumberger, I. Tavernelli, M. Klein, et M. Sprik, *J. Chem. Phys.* **124**, 064507 (2006).
- [313] J. M. Nicovich, K. D. Kreutter, C. A. Vandijk, et P. H. Wine, *J. Phys. Chem.* **96**, 2518 (1992).
- [314] J. P. Tam, Y.-A. Lu, et W. Yu, *J. Am. Chem. Soc.* **121**, 4316 (1999).
- [315] C. Kolano, J. Helbing, G. Bucher, W. Sander, et P. Hamm, *J. Phys. Chem. B* **111**, 11297 (2007).
- [316] H. S. M. Lu, M. Volk, Y. Kholodenko, E. Gooding, R. M. Hochstrasser, et W. F. DeGrado, *J. Am. Chem. Soc.* **119**, 7173 (1997).
- [317] L. Pauling, *J. Am. Chem. Soc.* **53**, 1367 (1931).
- [318] P. M. W. Gill et L. Radom, *J. Am. Chem. Soc.* **110**, 4931 (1987).
- [319] P. M. W. Gill et L. Radom, *J. Am. Chem. Soc.* **110**, 4931 (1988).
- [320] T.-L. Tung et J. A. Stone, *Can. J. Chem.* **53**, 3153 (1975).
- [321] C. Houée-Levin et J. Bergès, *Radiation Physics and Chemistry* **77**, 1286 (2008).
- [322] X. Fang, J. Wu, G. Wei, H. P. Schuchmann, et C. von Sonntag, *Int. J. Radiat. Biol.* **68**, (1995).
- [323] C. C. Lawrence, M. Bennati, H. V. Obias, G. Bar, R. G. Griffin, et J. Stubbe, *Proc. Nat. Acad. Sc. USA* **96**, 8979 (1999).
- [324] D. L. Johnson, S. W. Polyak, J. C. Wallace, et L. L. Martin, *Peptide Science* **10**, 495 (2003).
- [325] K. D. Asmus, *Acc. Chem. Res.* **12**, 436 (1979).
- [326] A. Rauk, *J. Am. Chem. Soc.* **106**, 6517 (1984).
- [327] A. Sawicka, J. Berdys-Kochanska, P. Skurski, et J. Simons, *Int. J. Quant. Chem.* **102**, 838 (2005).
- [328] G. Rickard, J. Bergès, C. Houée-Levin, et A. Rauk, *J. Phys. Chem. B* **112**, 5774 (2008).
- [329] J. Bergès, G. Rickards, A. Rauk, et C. Houée-Levin, *Chem. Phys. Lett.* **421**, 63 (2006).
- [330] J. Bergès, G. Rickard, A. Rauk, et C. Houée-Levin, *Chem. Phys. Lett.* **454**, 118 (2008).
- [331] M. Weik, R. B. Ravelli, S. I., S. J. L., P. Gros, et J. Kroon, *Proc. Nat. Acad. Sc. USA* **97**, 623 (2000).

- [332] M. Weik, J. Bergès, M. L. Raves, P. Gros, S. McSweeney, I. Silman, J. L. Sussman, C. Houée-Levin, et R. B. G. Ravelli, *J. Synchrotron Rad.* **9**, 342 (2002).
- [333] M. L. Colgrave et D. J. Craik, *Biochemistry* **43**, 5965 (2004).
- [334] J. C. Rienstra-Kiracofe, G. S. Tschumper, H. F. S. III, S. Nand, et G. B. Ellison, *Chem. Rev.* **102**, 231 (2002).
- [335] A. Sawicka, P. Skurski, R. R. Hudgins, et J. Simons, *J. Phys. Chem. B* **107**, 13505 (2003).
- [336] A. Sawicka, P. Skurski, et J. Simons, *J. Phys. Chem. A* **108**, 4261 (2004).
- [337] U. Goransson et D. J. Craik, *J. Biol. Chem.* **48**, 48188 (2003).
- [338] R. Zubarev, N. Kruger, E. Fridriksson, M. Lewis, D. Horn, B. Carpenter, et F. McLafferty, *J. Am. Chem. Soc.* **121**, 2857 (1999).
- [339] L. Ley et M. Cardona, *Photoemission in solids II* (Springer-Verlag, Berlin, 1979).
- [340] J. Stöhr, *NEXAFS Spectroscopy* (Springer-Verlog, Berlin, 1992).
- [341] W. L. Jolly, K. D. Bomben, et C. J. Eyermann, *Atom. Data Nucl. Data Tables* **31**, 433 (1984).
- [342] L. J. Saethre, N. Berrah, J. D. Bazek, K. J. Borve, T. X. Carroll, V. Kukk, G. L. Gard, R. Winter, et T. D. Thomas, *J. Am. Chem. Soc.* **123**, 10729 (2001).
- [343] J. Boese, A. Osanna, C. Jacobsen, et J. Kirz, *Journal of Electron Spectroscopy and Related Phenomena* **85**, 9 (1997).
- [344] J. Stewart-Ornstein, A. P. Hitchcock, D. H. Cruz, P. Henklein, J. Overhage, K. Hilpert, J. D. Hale, et R. E. W. Hancock, *J Phys Chem B* **111**, 7691 (2007).
- [345] M. L. Gordon, G. cooper, C. Morin, T. Araki, C. C. Turci, K. Kaznacheyev, et A. P. Hitchcock, *J. Phys. Chem. A* **107**, 6144 (2003).
- [346] R. Zubarev, A. Shaporenko, M. Grunze, et M. Zharnikov, *J. Phys. Chem. B* **111**, 9803 (2007).
- [347] V. Feyer, O. Plekan, R. Richter, M. Coreno, H. C. Prince, et V. Carravetta, *J. Phys. Chem. A* **112**, 7806 (2008).
- [348] Z. B. Maksić et R. Vianello, *J. Phys. Chem. A* **106**, 6515 (2002).
- [349] P. Bagus et H. Schaefer III, *J. Chem. Phys.* **55**, 1474 (1971).

- [350] P. Bagus et H. Schaefer III, *J. Chem. Phys.* **56**, 224 (1972).
- [351] W. A. Shapley et D. P. Chong, *Int. J. Quantum Chem.* **81**, 34 (2001).
- [352] D. Chong et Y. Takahata, *Chem. Phys. Lett.* **418**, 286 (2006).
- [353] Y. Takahata, A. Okamoto, et D. Chong, *Int. J. Quantum Chem.* **106**, 2581 (2006).
- [354] J. Herbert et J. Ortiz, *J. Phys. Chem. A* **104**, 11786 (2000).
- [355] N. Ferré et X. Assfeld, *J. Chem. Phys.* **117**, 4119 (2002).
- [356] P. Loos et X. Assfeld, *Int. J. Quantum Chem.* **107**, (2007).
- [357] V. N. Glushkov, *Chem. Phys. Lett.* **287**, 189 (1998).
- [358] V. N. Glushkov, *Chem. Phys. Lett.* **273**, 122 (1997).
- [359] V. N. Glushkov, *Optics and Spectroscopy* **93**, 15 (2002).
- [360] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, 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, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, et J. A. Pople, , Gaussian, Inc., Wallingford, CT, 2004 and 2005.
- [361] V. N. Glushkov et A. Y. Tsaune, *Zh. Vichislit. Matem. Mater. Fiz.* **25**, 298 (1985).
- [362] K. Kaznachev, C. Osanna, A. and Jacobsen, O. Plashkevych, O. Vahtras, O. Agren, H. V. Carravetta, et A. P. Hitchcock, *J. Phys. Chem. A* **106**, 3153 (2002).
- [363] A. R. Slaughter et M. S. Banna, *J. Phys. Chem.* **98**, 2165 (1988).
- [364] O. Plekan, V. Feyer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, et V. Carravetta, *J. Phys. Chem. A* **111**, 10998 (2007).

- [365] C.-H. Hu, M. Z. Shen, et H. F. Schaefer III, *J. Am. Chem. Soc.* **115**, 2923 (1993).
- [366] B. W. Loos, I. M. Sauerwald, A. P. Hitchcock, et S. S. Rothman, *Journal of Microscopy* **204**, 69 (2001).
- [367] D. Nolting, E. Aziz, N. Ottosson, M. Faubel, I. Hertel, et B. Winter, *J. Am. Chem. Soc.* **129**, 14068 (2007).
- [368] Y. Zubavichus, A. Shaporenko, M. Grunze, et M. Zharnikov, *J. Phys. Chem. B* **110**, 3420 (2006).

Résumé

Les méthodes hybrides alliant la mécanique quantique et la mécanique moléculaire (QM/MM) sont des outils adéquats pour traiter des systèmes biologiques. Les phénomènes électroniques souvent étudiés sur des petites molécules ont, dès lors pu être envisagés dans des environnements macromoléculaires. Ce travail explore trois phénomènes électroniques en présence d'un environnement protéique : l'absorption, la capture électronique et les ionisations de cœur. Nous avons employé les derniers développements de la méthode QM/MM *Local Self-Consistent Field* (LSCF) pour traiter la jonction covalente entre la partie QM et la partie MM ainsi que le couplage QM/MM avec le PCM. En premier lieu, nous nous sommes focalisés sur les spectres d'absorption de chromophores présents dans des macromolécules. Nous avons mis au point un couplage entre les méthodes LSCF/MM et PCM pour prendre en compte la polarisation électronique de l'environnement suite à l'absorption d'un photon par le chromophore. Ce modèle, le LSCF/MM:ERS, a été testé et validé sur le spectre d'absorption du complexe de squaraine-tétralactame. Une étude plus poussée a ensuite été réalisée sur une protéine fluorescente en décomposant la longueur d'onde d'absorption maximale en trois contributions physiques. Les effets de la substitution du chromophore ont aussi été évalués. Dans une seconde partie, nous avons étudié la capture électronique par un cyclotide contenant trois ponts disulfures, qui, après irradiation, forment des demi-liaisons caractéristiques (2c-3e). La dernière partie est consacrée à la validation de l'approche de la projection asymptotique. Elle a été réalisée dans le cadre d'une étude sur les ionisations de cœur d'un ensemble de molécules pour, par la suite, étudier les ionisations de cœur de la glycine présentes dans des systèmes de plus en plus complexes jusqu'à la Sérum-Albumine Humaine.