

Thèse de Doctorat

Titouan JAUNET-LAHARY

*Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes
sous le sceau de l'Université Bretagne Loire*

École doctorale : 3MPL : Matière, Molécules, Matériaux en Pays de la Loire

Discipline : Chimie

Spécialité : Chimie Théorique

Unité de recherche : Laboratoire Chimie Et Interdisciplinarité, Synthèse, Analyse,
Modélisation (CEISAM) UMR 6230, CNRS

Soutenu le 2 octobre 2017

Modélisation et simulation de nouveaux inhibiteurs de Rad51 au sein d'une protéine de transport

JURY

Président du jury :	Daniel VERCAUTEREN , Professeur, Université de Namur
Rapporteurs :	Elise DUMONT , Professeur, ENS de Lyon Nicolas FERRÉ , Professeur, Université d'Aix-Marseille
Directeur de Thèse :	Denis JACQUEMIN , Professeur, Université de Nantes
Co-directeurs de Thèse :	Adèle LAURENT , Chargée de Recherche, Université de Nantes Fabrice FLEURY , Professeur, Université de Nantes

Remerciements

Ces trois années de doctorat que je viens de passer sont pour moi l'aboutissement de mes études universitaires, mais aussi celui d'un projet personnel de longue date. Durant toutes ces années, j'ai eu la chance d'avoir rencontré de nombreuses personnes et aujourd'hui je profite de ces pages pour les remercier.

Je tiens, tout d'abord, à remercier Jean-Michel Bouler et Bruno Bujoli, respectivement, actuel et ancien directeur du laboratoire CEISAM, pour m'avoir permis d'effectuer ma thèse de doctorat. De plus, je suis reconnaissant une seconde fois à Bruno pour son aide et ses conseils lors de l'écriture de mon projet JSPS (Japan association for the Promotion of Science).

Je remercie Mme Élise Dumont et Mr Nicolas Ferré en leur qualité de rapporteurs de ce manuscrit d'avoir accepté d'évaluer ce travail. Je remercie aussi très chaleureusement Mr Daniel Vercauteren de m'avoir répondu positivement à ma demande en tant qu'examinateur ainsi que pour les discussions très intéressantes que nous avons pu partager.

Cher Denis, je te remercie infiniment pour toutes ces années au sein de ton équipe de recherche. Merci de m'avoir initié, formé et fait apprécier la chimie théorique (ce qui n'était pas gagné au début), d'avoir accepté de devenir mon directeur de thèse, il y a déjà 3 ans. Tu m'as toujours apporté de très bons conseils scientifiques et professionnels, et c'est en grande partie grâce à toi si j'ai pu en arriver là aujourd'hui. Ta disponibilité et ta grande gentillesse m'ont permis de vivre d'excellents moments lors de cette thèse et encore merci de m'avoir fait découvrir les excellents paysages de Slovaquie (propice au jogging), sa gastronomie et ses excellentes bières locales.

Chère Adèle, je te suis extrêmement reconnaissant de m'avoir permis de travailler sur ce sujet de recherche. Je te remercie pour toute l'aide que tu m'as apportée durant ces 3 années (et notamment sur la fin) et ça a été un immense plaisir de travailler avec toi. Durant cette thèse, tu m'as toujours soutenu et encouragé à aller de l'avant et ceci m'a permis d'accomplir des choses incroyables à mes yeux. Grâce à toi et tes encouragements, j'ai eu l'occasion de voyager de nombreuses fois et j'y en ai finalement pris goût. Je te suis très redevable pour tout ce que tu m'as apporté que ce soit scientifiquement et professionnellement. Mille merci pour tout !

Je remercie Fabrice Fleury d'avoir accepté de co-dirigé ma thèse. Son aide concernant la partie biologique de mon projet a été très précieuse.

Je tiens aussi à remercier Jean-Yves Le Questel pour tous ses nombreux conseils et sa bonne humeur quotidienne ; Jérôme Graton pour sa précieuse collaboration à mes recherches ainsi que son soutien et sa gentillesse ; et Anaïs Goupille pour son aide lors des expériences menées. Je salue aussi Nicolas Galland et Eric Renault pour leur amabilité et bienveillance.

Je voudrais rendre hommage à Aymeric Blondel, qui a été mon sauveur lors de mes problèmes de cluster, scripts et autres énigmes de l'informatique. Sans lui je n'aurais pas pu connaître le plaisir de la programmation, du babysitting et des défaites du FCN à la Beaujoire.

Je remercie chaleureusement l'ensemble des doctorants et post-doc de l'équipe ModES (anciens et actuels) et en particulier mes trois "colloques" de bureau (Zak, Anouar et Simon), avec qui j'ai passé d'excellents moments, discussions et fous rires dans de nombreuses langues. Je salue également Gabriel, Azzam, Siwar, Kathy, Claudiu, Rui, Ciro et Daniel pour tous les excellents moments passés en leur compagnie. Je remercie particulièrement Cloé pour m'avoir supporté et remonté le moral dans les moments difficiles, pour sa bienveillance et son extrême gentillesse ainsi que pour toutes ces critiques linguistiques constructives (à n'en point douter !). Cela a été un extrême plaisir de te retrouver après ton premier passage à Nantes.

Je tiens à remercier les trois stagiaires dont j'ai eu la chance d'encadrer, Jérémy, Charles et Zakaria. Ils ont tous contribué à l'avancement de mes travaux et je leur en serai toujours reconnaissants. En espérant que leurs projets se concrétiseront dans le futur.

Je remercie (une nouvelle fois) Simon Budzak et Miroslav Medved, avec qui j'ai eu la chance de collaborer en France et en Slovaquie. J'espère pouvoir venir vous revoir dans cette si belle ville qu'est Bystricka.

Je suis très reconnaissant envers la JSPS pour m'avoir permis d'effectuer un séjour de recherche au Japon lors du JSPS Summer Program 2016. De plus, Je remercie le Pr Sakaki et le Pr Morokuma d'avoir accepté de m'accueillir deux mois au sein du FIFC (Fukui Institute for Fundamental Chemistry) où j'y ai fais de nombreuses et belles rencontres (Satoshi, Kimichi, Sameera et Pedro). Cette expérience au Japon m'a aussi permis de créer une excellente relation avec la famille Akatsuka, qui m'a accueilli chaleureusement durant mon séjour. Je vous suis très reconnaissant pour l'extrême hospitalité que vous m'avez donnée.

Ces remerciements sont pour moi l'occasion de saluer trois professeurs qui m'ont beaucoup apporté durant mes études. Le premier est Arnaud Campas pour m'avoir fait découvrir la chimie

et transmis sa passion. Le deuxième est Mohammed Boujtita, qui a été le premier à me donner une chance dans le milieu de la recherche. Les nombreuses et précieuses discussions que j'ai pu avoir avec lui ont toujours été d'excellents moments. J'espère pouvoir le revoir autour d'un bon whisky japonais. Le troisième est Xavier Rocquefelte, qui a su me faire apprécier la chimie théorique. Sa fougue et sa joie d'enseigner sont pour moi un modèle de pédagogie. Notre rencontre à Kyoto fut une excellente retrouvaille.

Je salue l'ensemble des doctorants, post-doctorants et membres du personnel que j'ai pu croisé lors de mon passage au laboratoire CEISAM. Je remercie plus particulièrement mes trois collègues, Kevin, Maxime et Renaud. Je remercie Ugo, Florent, Yoann, Yoann, Joanna, Manon, Noémie, Guillaume, Mathieu, Gilles, Katy, Tanguy, Valentin, Boris, Laetitia, Jérémy et les autres pour tous les bons moments passés en leur présence. Je remercie chaleureusement Florian, avec qui nous nous sommes soutenu mutuellement lors de cette fin de thèse. Petite dédicace à mon Serge national(en espérant qu'il se reconnaîtra), où chacun des moments en sa compagnie, notamment lors de nos repas au Patio, furent des moments exquis.

Enfin, je termine par remercier l'ensemble de mon entourage pour le soutien qu'ils m'ont apporté durant de nombreuses années. Merci à mes parents, mes sœurs, mon beau-frère et le reste de ma famille de m'avoir toujours encouragé dans mes choix. Merci à mes amis du volley (Antoine, Shin, Sami, Sarah, Benoit, Maureen, Rémi et Sydney) pour les super moments passés à vos côtés. Merci à mes amis Noirmoutrins de toujours (Rémy, Nans, Guigou, Nawi, Arnaud, Jo, Jéro et Mélissa) pour votre solidarité. Merci à Sonny pour tous ces moments (indélébiles) vécus ensemble. Merci à Yuri pour son soutien quotidien durant les derniers mois de cette thèse.

Table des matières

Table des matières

Introduction	1
I Systèmes biologiques	5
1 La protéine d'Albumine Sérique Humaine	7
1.1 Généralités	7
1.2 Structure de la ASH	8
1.2.1 Structure primaire de la protéine	8
1.2.2 Sites de Sudlow	9
1.2.3 Premières structures cristallographiques de la ASH	10
1.2.4 Structure de la ASH complexée	11
2 Cas d'études portés sur la ASH	15
2.1 Problématique de biologie structurale	15
2.1.1 États de la protonation de la Lys199	15
2.1.2 Importance des ponts disulfures sur la structure de la ASH	16
2.1.3 Influence des acides gras	17
2.2 Complexe ASH-ligand	18
3 Les inhibiteurs de Rad51	19
3.1 Les familles d'inhibiteurs	20
3.1.1 Composés modulant l'expression de Rad51	20
3.1.2 Modulateurs de l'activité de la RH	21
3.2 Les stilbènes disulfoniques	24
3.2.1 Photochimie du stilbène	24
3.2.2 Propriétés physico-chimiques du DIDS	25

II	Propriétés quantiques en solution des SD	27
4	Les méthodes <i>ab initio</i> en chimie quantique	29
4.1	Hartree-Fock	30
4.2	La Théorie de la Fonctionnelle de Densité	31
4.2.1	Fonctionnelle de Thomas-Fermi	31
4.2.2	Théorèmes de Hohenberg et Kohn	32
4.2.3	Équations de Kohn-Sham	32
4.2.4	Fonctionnelle d'échange-corrélation	34
4.3	Simulation des effets de solvants	36
4.3.1	Solvatation explicite	36
4.3.2	Solvatation implicite	36
4.3.3	Solvatation implicite/explicite (IE)	38
4.4	DFT dépendante du temps	38
4.4.1	Théorèmes de Runge-Gross et van Leeuwen	39
4.4.2	Equation KS dépendante du temps et réponse linéaire	39
4.4.3	Propriétés optiques en solution	41
4.4.4	Limites physiques	41
4.4.5	Contributions vibrationnelles des spectres optiques	42
4.5	Bases de fonctions atomiques	43
5	Études des dérivés stilbènes disulfoniques	47
5.1	États de protonation des SD	47
5.2	Propriétés optiques et solvatochromiques du DIDS ²⁻	51
5.2.1	Partie expérimentale	52
5.2.2	Partie théorique	56
5.3	Propriétés optiques de dérivés SD ²⁻	63
5.3.1	Effets auxochromiques en solution aqueuse	63
5.3.2	Effets solvatochromiques du DIDS ²⁻ , du DADS ²⁻ et structures fines	66
5.4	Conclusions	69
III	Modélisation du complexe ASH-DADS²⁻	71
6	Mécanique et dynamique moléculaires	73
6.1	Énergie totale d'un champ de force	73
6.2	Famille de champ de force	75
6.3	Dynamique moléculaire	76
6.3.1	Résolution des équations du mouvement	76
6.3.2	Ensembles canoniques	78

6.4	Solvatation et conditions périodiques	80
6.5	Traitement des interactions à longues distances	81
6.5.1	L'image minimale	81
6.5.2	Troncature des interactions	81
6.6	Calculs MM/GBSA	82
7	Analyses des complexes ASH-DADS²⁻	85
7.1	Généralités	85
7.2	Validation de la méthode de <i>docking</i> moléculaire	86
7.2.1	Sites étudiés	87
7.2.2	Préparation de la protéine	88
7.2.3	Protocole de <i>docking</i>	89
7.3	Validation du protocole de <i>docking</i>	89
7.3.1	Bilirubine (sous-domaine IB)	90
7.3.2	Warfarine (sous-domaine IIA)	91
7.3.3	Ibuprofène (sous-domaines IIAIIB et IIIA)	92
7.3.4	Bilan	95
7.4	Etude des interactions ASH-DADS ²⁻	95
7.4.1	Sous-domaine IB	95
7.4.2	Sous-domaine IIA	100
7.4.3	Sous-domaine IIAIIB	110
7.4.4	Sous-domaine IIIA	113
7.5	Discussions	117
7.6	Conclusions	119
	Conclusion générale et perspectives	121
	Liste des abréviations	125
	Bibliographie	129

Introduction

La prévention et le dépistage du cancer ont, en 2016, été sujets à une proposition de loi de « grande cause nationale » par la députée Mme Véronique BESSE.^[1] Depuis le début du XXIème, la compréhension des mécanismes biologiques qui sont impliqués dans la formation de cellules cancéreuses est devenue un sujet de santé publique. Plusieurs traitements ont été développés pour lutter contre le cancer comme la radiothérapie, chimiothérapie, l'immunothérapie, etc. Malheureusement, le cancer est une maladie très complexe et de plus en plus de cas de résistance de cellules cancéreuses aux traitements sont observés.^[2] Depuis quelques années, une cible thérapeutique a été identifiée, la protéine polymérase Rad51, comme acteur majeur dans le processus de résistance des cellules cancéreuses.^[3] La protéine Rad51 est, plus précisément, impliquée dans la recombinaison homologue des lésions double brin de l'ADN.^[4] Plusieurs études ont fourni la preuve que dans certains cancers (seins, pancréas et poumons) la protéine Rad51 est en surexpression^[5-7] et qu'il y a une relation de cause à effet entre cette surexpression et la résistance des cellules cancéreuses.^[8,9] La modulation de l'activité de la protéine Rad51 au sein des cellules cancéreuses apparaît donc comme une nouvelle voie pour lutter contre la résistance de ces dernières aux traitements thérapeutiques. Plusieurs composés capables d'inhiber l'activité de Rad51 ont été identifiés.^[10-12] En effet, un dérivé de la famille des stilbènes disulfoniques (SD), le DIDS (4,4'-diisothiocyanostilbene-2,2'-disulfonic acid) a montré une action sur l'activité de la protéine Rad51.^[13] Une concentration de 1-2 μM de DIDS réduit de 50% l'activité de Rad51 (6 μM). À partir de ces résultats et au regard du motif chimique des SD, l'équipe du Pr. Fabrice Fleury, du laboratoire UFIP à Nantes, a entrepris depuis quelques années différentes études sur l'inhibition de Rad51 avec des dérivés du DIDS.^[14] La conception d'autres dérivés SD devrait être une voie prometteuse en vue d'une optimisation de l'inhibition sur la protéine Rad51.

Un aspect à prendre en compte lors de la conception de nouveaux inhibiteurs est leur biodisponibilité, car l'accès à leur cible thérapeutique n'est pas forcément garanti au regard des différentes barrières biologiques dans l'organisme. Pour cela, plusieurs protéines permettent de véhiculer ces inhibiteurs au sein du corps humain et passer ces barrières. Parmi ces protéines, l'albumine sérique humaine (ASH) est connue pour jouer un tel rôle. En effet, cette protéine, qui est la plus abondante du plasma sanguin humain, permet le transport de nombreuses molécules hydrophobes parmi des milieux aqueux. Dans les tissus tumoraux, la ASH peut passer la barrière endothéliale plus facilement car cette dernière est plus perméable que dans le cas de tissus sains. De plus, la concentration de la ASH dans les tissus tumoraux est très élevée car le drainage y est moins ef-

ficace, ce qui favorise l'effet EPR (*Enhance Permability and Retention*).^[15,16] Il a été, en outre, montré que la ASH transporte des molécules possédant une action anti-cancéreuse (cisplatine, camptotécine, etc.), ce qui la positionne comme une macromolécule de choix pour le ciblage de nouveaux médicaments.^[17-19]

L'objectif de notre thèse de doctorat est d'étudier les interactions entre la protéine de transport ASH et les inhibiteurs SD afin de déterminer si ces derniers peuvent être véhiculés vers les tissus tumoraux. Notre manuscrit se décompose en trois grandes parties. La première partie présente les deux acteurs majeurs de notre projet : la protéine de transport ASH et la famille des SD, inhibiteurs potentiels de la protéine Rad51. Le premier chapitre de cette partie se concentre sur la présentation de la protéine ASH d'un point de vue historique, structurale et de son rôle comme protéine de transport dans le corps humain. Le deuxième chapitre expose trois cas d'études où l'utilisation de méthodes théoriques a permis d'apporter des informations complémentaires à l'expérience et a contribué à répondre à la problématique donnée. Quelques exemples de complexes ASH-médicaments y sont aussi évoqués. Le troisième chapitre est consacré à la présentation de la famille des SD d'un point de vue général, ainsi que leur implication dans divers mécanismes biologiques comme l'inhibition de canaux ioniques (Cl^- , HCO_3^- , I^- , NO_3^- , etc.), permettant la régulation du potentiel membranaire.

La seconde partie de notre thèse est dédiée à l'étude des propriétés physico-chimiques des SD en solution. Le quatrième chapitre introduit les méthodes quantiques (*Quantum Mechanics*, QM) que nous avons employées tout au long de cette thèse. Les principes fondamentaux de la méthode Hartree-Fock (HF) et de la Théorie de la Fonctionnelle de Densité (*Density Functional Theory*, DFT), ainsi que sa composante dépendante du temps, la TD-DFT (*Time-Dependent-DFT*), y sont développés. Les différents formalismes permettant de prendre en compte des effets de solvants lors des calculs théoriques sont aussi explicités. Le cinquième chapitre présente une étude combinant des résultats expérimentaux et théoriques sur le comportement en solution d'un ensemble de composés appartenant à la famille des SD. Une étude préliminaire sur l'état de protonation de ce type de molécule est développée, s'appuyant notamment sur la détermination des valeurs de constantes d'acidité théoriques et des mesures expérimentales d'absorption à différentes valeurs de pH. Ensuite, la simulation des propriétés optiques en solution des SD est abordée et une attention toute particulière est portée sur la description des effets solvatochromiques des molécules. La fin du chapitre présente la prise en compte des couplages vibroniques sur la forme des bandes d'absorption.

La dernière partie de notre thèse s'intéresse aux complexes protéine-ligand entre les deux acteurs : la protéine de transport ASH et les composés de la famille SD. Le sixième chapitre fait un rappel de la mécanique moléculaire (MM), sur lesquels repose le principe des champs de force. Une explication des différentes forces composant un système est donnée. La méthode de la dynamique moléculaire (*Molecular Dynamics*, MD) et l'intégration dans le temps des équations du mouvement y sont développées, ainsi que le traitement des interactions à longue distance durant les simulations de MD. L'explication du principe des calculs d'énergie de complexation relative

MM/GBSA (*Molecular Mechanics/Generalised Born Surface Area*) est présentée en fin de chapitre. Le septième chapitre concentre tous nos résultats théoriques concernant la formation du complexe ASH-DADS. Aucune structure cristallographique de ce complexe n'étant à ce jour disponible, nous avons généré, à l'aide de la méthode de *docking* moléculaire, une série de premières structures des complexes ASH-DADS. Le DADS a été placé dans quatre différents sites d'accueil possible de la protéine. À partir des structures générées, des simulations de MD ont ensuite été effectuées : le comportement et les différents modes d'interactions du DADS au sein de la protéine ASH y sont discutés. Enfin, une analyse et une discussion des énergies de complexation relatives par la méthode MM/GBSA concluent le dernier chapitre.

Une conclusion générale de notre travail de thèse est ensuite présentée et des pistes de perspectives potentielles à ce travail sont évoquées, tant expérimentalement que théoriquement.

Première partie

Systemes biologiques

1 La protéine d'Albumine Sérique Humaine

Le système principal étudié durant notre thèse de doctorat est la protéine d'albumine sérique humaine (ASH), protéine de transport la plus abondante dans le sérum sanguin (60%). Un des objectifs de mon travail est l'étude des interactions potentielles entre la ASH et des inhibiteurs anti-cancéreux ; il est donc important de connaître le rôle de cette protéine au sein de l'organisme ainsi que sa structure et les différents modes d'interactions avec d'autres ligands, connus dans la littérature. Ce chapitre comprend un état de l'art, non-exhaustif, des caractéristiques de la ASH. De nombreux ouvrages et revues sont consacrés aux propriétés structurales, biochimiques, génétiques, métaboliques, etc., de la ASH illustrant l'intérêt pour cette protéine.^[20–24]

1.1 Généralités

Historique

Le terme « albumine » est dérivé du terme allemand *albumen*, qualifiant généralement une protéine. *Albumen* provient lui-même du latin *albus*, se traduisant par « blanc », indiquant ainsi la partie blanche d'un œuf une fois ce dernier cuit. La protéine d'albumine fut une des premières protéines du corps humain à être étudiée, et les premiers écrits y faisant référence remontent à l'époque d'Hippocrate.^[25] Au XVI^e siècle, le physicien Paracelsus réussit à précipiter les protéines d'albumine présentes dans l'urine grâce à du vinaigre.^[26] Plus tard, Harvey étudiant la circulation sanguine, découvre que le sérum sanguin contient des protéines et emploie le terme *albumen* pour les qualifier.^[27] À la fin du XIX^e siècle, Kauder entame des expériences afin de connaître la composition de l'*albumen*.^[28] L'ajout de sulfate de magnésium à l'*albumen* précipite les globulines présentes et le filtrat obtenu par dialyse est nommé sérum d'albumine. Gürber réussit, ensuite, à cristalliser l'albumine provenant du cheval. Une première estimation de sa masse moléculaire est de 70 000 Da¹, formulée par Svedberg, par à une technique d'ultracentrifugation.^[29]

Lors de la seconde guerre mondiale, un intérêt particulier fut porté sur la préparation et la purification de l'albumine dans l'objectif d'obtenir une solution alcoolique pouvant servir comme substitut au plasma sanguin. De telles solutions à base d'albumine présentent certains avantages pratiques et sanitaires par rapport à une transfusion de plasma. Ainsi, Cohn *et al.* proposa un

1. Le Dalton (Da) est l'unité de masse pour caractériser la masse des macromolécules. 1 Da. correspond à un douzième de la masse d'un atome de carbone 12.

protocole de purification de l'albumine à partir de plasma bovin applicable en intraveineuse.^[30] Cependant, de nombreux cas de maladies causées par les transfusions, notamment au niveau des reins, furent observés malgré une haute purification de la protéine. Ils constatèrent que l'albumine bovine était différente de la forme humaine, ce qui mena au développement d'un autre protocole à partir de sérum humain afin de fractionner l'albumine dans une solution d'éthanol à froid.

Rôle biologique

Le rôle biologique de la ASH est très important, notamment au niveau de la circulation sanguine. En effet, la ASH est la protéine la plus abondante dans le plasma sanguin et sa concentration, chez un homme sain, est d'environ 40 mg.ml⁻¹ soit une concentration de 0,06 mM. Sa biosynthèse s'effectue au niveau du foie^[31] et le rôle majeur de cette protéine est de transporter les composés endogènes et exogènes dans le corps humain. Elle est reconnue pour avoir une grande capacité à se lier avec différents types de composés comme les acides gras, médicaments, hormones, métaux, etc., faisant d'elle l'une des plus importantes protéines de transport dans le corps humain.^[21,24,32] De plus, elle est un marqueur biologique de certaines maladies comme le cancer, l'ischémie, l'arthrite, le diabète etc.,^[33-36] et possède également une activité anti-oxydante.^[37]

1.2 Structure de la ASH

1.2.1 Structure primaire de la protéine

La structure primaire de la ASH est composée de 585 acides aminés, donnant lieu à une masse d'environ 65 000 Da (assez proche des 70 000 Da comme proposé par Svedberg en 1934). Meloun *et al.* ont proposé la séquence complète de la protéine monomérique en 1975.^[38] Parmi les 35 résidus cystéines que la ASH possède, 17 ponts disulfures sont présents et contribuent à la grande rigidité de la protéine. L'unique cystéine libre est localisée en position 34. De nombreux résidus chargés (Lys, Arg, Glu, Asp) sont présents dans la protéine qui possède une charge nette totale de -15 e⁻, ce qui explique sa solubilité dans l'eau. Comme illustré à la Figure 1, la séquence de la ASH s'organise en trois domaines homologues (I, II et III), chacun divisé en deux sous-domaines (A et B). Cette organisation, proposée par Foster en 1960,^[39] indique que chaque domaine possède la même séquence d'acides aminés, à quelques exceptions près, et que les domaines I, II et III sont respectivement composés des résidus 1 à 195, 196 à 383 et 384 à 585. La ASH est exclusivement composée d'hélices α et chaque domaine en comporte 10.

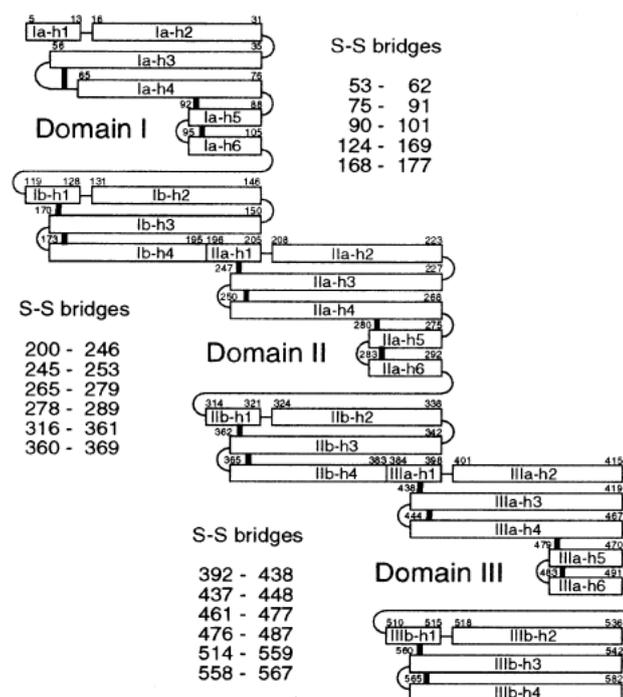


FIGURE 1 – Schéma de l’organisation tridimensionnelle par hélice α et domaines de la ASH. Les ponts disulfures (S-S *bridges*) sont indiqués par le numéro des cystéines impliqués. Image reprise de la référence 40.

1.2.2 Sites de Sudlow

Les travaux de Sudlow en 1975 ont permis d’identifier deux sites principaux d’accueil pour des composés médicamenteux.^[41–43] La caractérisation de ces sites a été effectuée à l’aide de l’utilisation de sondes fluorescentes. En effet, des mesures de compétition d’affinités entre une molécule sonde et une autre molécule permettent d’observer une variation de la fluorescence du système ASH-sonde en présence d’une molécule compétitrice. La variation de fluorescence peut être alors associée à une constante d’association. Trois classes de sondes ont été développées : la première est spécifique au site I, la deuxième au site II et la dernière possède une affinité pour les deux sites. À partir de la mesure de fluorescence, il a été possible de déduire le ou les site(s) hôte(s) de plusieurs composés. Par exemple, la warfarine, la sulfapyrazone, le tolbutamide, le phénylbutazone, l’éthyl-bisoumacétate se localisent au niveau du site I de Sudlow tandis que l’ibuprofène, l’acide flufénamique, l’acide étacrynique, etc., sont localisés dans le site II de Sudlow.

Sudlow note également que certains composés spécifiques au site II, comme par exemple l’ibuprofène, ont un effet inattendu sur la fluorescence de la molécule sonde DNSA (5-diméthylaminonaphtalène-1-sulfonamide) qui est spécifique au site I. La présence de l’ibuprofène dans le site II va induire une augmentation de la fluorescence du DNSA. Ce phénomène met en évidence que l’occupation du site II par une molécule amène une variation de la conformation de la protéine ASH. Ceci est en accord avec les hypothèses de Karush et Koshland, portant sur un changement de conformation de la protéine par l’accueil d’une molécule.^[44,45] Cette hypothèse a été émise dans

le cas de l'albumine sérique bovine (ASB) mais est tout à fait envisageable dans le cas de la ASH.

Deux sites d'accueil de la ASH ont été mis en lumière grâce aux travaux de Sudlow, mais leur localisation relative au sein de la protéine n'a pu être définie. L'identification précise fut possible seulement par l'arrivée des premières structures tridimensionnelles (3D) de la ASH complexée avec des ligands spécifiques à chaque site, comme décrit dans la partie suivante.

1.2.3 Premières structures cristallographiques de la ASH

La première structure cristallographique résolue par la diffraction des rayons X de la protéine ASH libre a été réalisée par Carter et ses collaborateurs à la fin des années quatre-vingt.^[46,47] Ils distinguèrent une structure 3D caractéristique en forme de cœur (« *heart-shaped* »), ainsi que les trois domaines. Ils constatent que les domaines I, II et III peuvent se superposer, confirmant leur homologie séquentielle et structurale. Cependant, la faible résolution, à l'époque, de la structure cristallographique ($\approx 6 \text{ \AA}$) ne permettait pas d'identifier précisément chaque résidu. Lors de la préparation du cristal, un agent mercurieux (HgCl_2) a été ajouté afin de confirmer la présence d'une cystéine libre. Les fonctions thiols sont, en effet, reconnues pour avoir une très grande affinité avec les ions Hg^{2+} , en formant des liaisons S-Hg. Un seul site de fixation de l'atome de mercure fut observé et localisé près de l'amine terminale, ce qui est en accord avec la présence d'une cystéine libre en position 34.

En 1992, les mêmes auteurs obtinrent une structure cristallographique avec une résolution de $2,8 \text{ \AA}$, permettant ainsi d'acquérir des informations plus complètes.^[48] Cependant, il est à noter que les positions des cinq acides aminés terminaux (n°1-3, n°584 et n°585) n'ont pu être résolues correctement et sont donc absentes au sein de la structure proposée. Les auteurs publièrent leur structure dans la PDB (Protein Data Bank) sous le code 1UOR et estimèrent que le pourcentage d'hélice α était de 67 % dans la protéine. Ils déterminèrent que chaque domaine était constitué de 10 hélices (h1-h10), les six premières appartenant au sous-domaine A et les quatre dernières au sous-domaine B (Figure 2). Six ponts disulfures sont présents par domaine (h1-h3, h3-h4, h4-h5, h5-h6, h7-h9 et h9-h10). Le premier domaine fait cependant figure d'exception car il n'en possède que cinq, le pont disulfure h1-h3 étant absent. La connexion entre les sous-domaines A et B s'effectue par le biais d'une chaîne peptidique désordonnée qui s'étend de la Lys106 à la Glu119 pour le domaine I, de la Glu292 à la Val315 pour le domaine II et de la Glu492 à l'Ala511 pour le domaine III. La connexion entre deux domaines différents est caractérisée par l'enchaînement de l'hélice h10 du sous-domaine B avec l'hélice h1 du sous-domaine A, du domaine suivant. Un des apports majeurs des travaux de Carter *et al.* est l'identification des deux sites d'accueil principaux pour les médicaments, évoqués par Sudlow en 1975 (Figure 2). À partir d'une série de différents ligands (aspirine, warfarine, ibuprofène, diazépam, etc.), co-cristallisés avec la ASH, les auteurs localisèrent le site I de Sudlow au niveau du sous-domaine IIA et le site II de Sudlow au niveau du sous-domaine IIIA. Ils déterminèrent que les deux sites sont similaires et forment une poche hydrophobe.

En 1999, Sugio *et al.* obtinrent une structure cristalline de la ASH avec une résolution de 2,5 Å (PDB code : 1AO6).^[40] Les auteurs indiquent que seuls 578 acides aminés (résidus n°5 au n°582) ont pu être identifiés correctement car la densité électronique correspondant aux résidus 1 à 4 et 583 à 585 n'a pu être clairement résolue dû à la grande flexibilité de chaque extrémité de la protéine. La forme globale en cœur fut également observée dans cette structure avec un volume approximatif de 192 000 Å³ (80 × 80 × 30 Å). Sugio *et al.* indiquèrent que les domaines I et II sont perpendiculaires entre eux (« *T-shaped* ») et que le sous-domaine IIA (en vert sur la Figure 2) est localisé près de l'interface des sous-domaines IA (en rouge) et IB (en jaune). Un angle d'environ 45° est mesuré entre le sous-domaine IIB et le domaine III, donnant lieu à une forme de Y entre les domaines II et III (« *Y-shaped* »). Des contacts entre certaines hélices α du sous-domaine IB et celles des sous-domaines IIIA et IIIB sont également mentionnés dans leur travail. Le nombre total d'hélices α est finalement ramené à 28, au lieu des 30 suggérées par Carter et son groupe, car les auteurs considèrent que l'hélice h10 du domaine I (II) et l'hélice h1 du domaine II (III) forment une seule et même hélice.

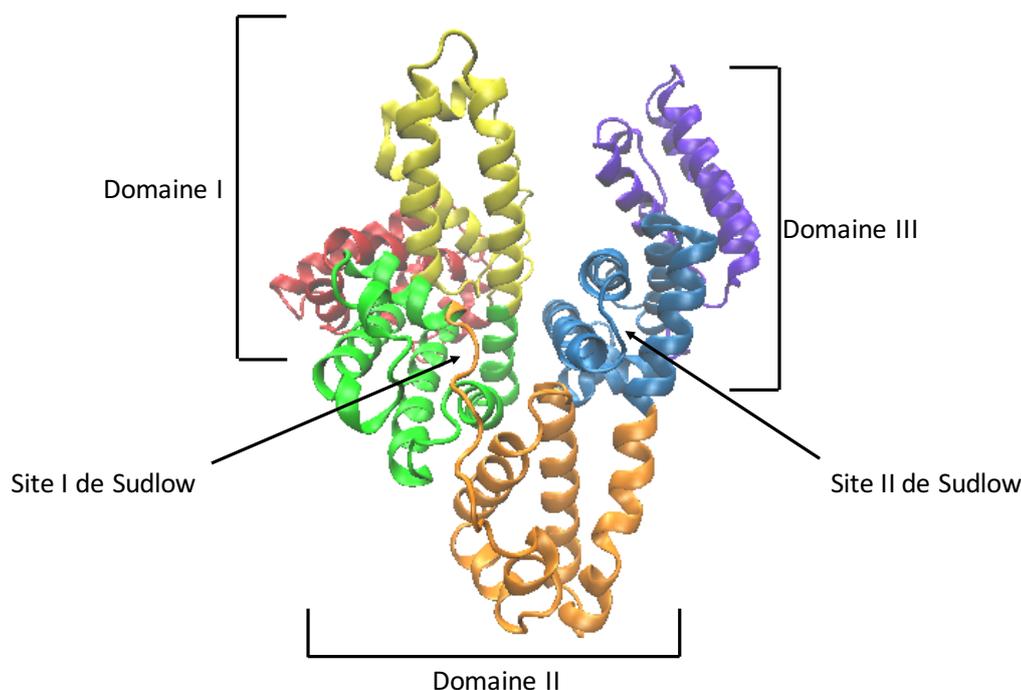


FIGURE 2 – Schéma des domaines (I, II et III) et des sous-domaines (IA (rouge), IB (jaune), IIA (vert), IIB (orange), IIIA (bleu) et IIIB (violet)) de la protéine ASH représentés par des hélices α .

1.2.4 Structure de la ASH complexée

Après la parution des structures cristallographiques de la ASH par Carter *et al.* et Sugio *et al.*, de nombreux groupes entreprirent l'étude de la complexation de cette protéine avec des molécules biologiques, en vue de rationaliser l'incroyable faculté de la ASH à transporter divers composés présents dans le corps humain et ainsi de mieux comprendre leurs processus ADME (Absorption, Distribution, Métabolisme et Excrétion). Le groupe de Curry est un des groupes de recherche les

plus réputés dans le domaine ; il publia de nombreuses structures de la ASH complexée.

Acides gras

En 1998, Curry *et al.* réalisèrent une première étude basée sur la complexation d'acides gras (acide myristique, $C_{14}H_{28}O_2$) avec la ASH. Les structures cristallographiques sont obtenues avec une résolution de 2,5 Å (PDB code : 1BJ5 et 1BKE).^[49] Les sites accueillant les acides gras sont distribués de façon asymétriques au sein de la protéine en dépit de la grande homologie entre les trois domaines. Comme indiqué à la Figure 3, cinq premiers sites de fixation d'acides gras (*Fatty acid*, FA) ont été identifiés : un au niveau du sous-domaine IB (FA1), un autre à l'interface des sous-domaines IA-IB-IIA (FA2) et trois au sein du domaine III (FA3-5). La comparaison des structures de la ASH complexée et ASH non-complexée (1AO6) montre des variations géométriques, notamment pour les domaines I et III. En effet, la superposition des atomes de carbone α de ces deux structures montre respectivement un RMSD (*Root Mean-Square Deviation*) de 8,8, 0,8 et 6,8 Å pour les domaines I, II et III. La présence d'acides gras dans les domaines I et III explique cette forte déviation, notamment au niveau des deux sous-domaines terminaux, IA et IIIB, qui présente respectivement un RMSD de 10,7 et 9,96 Å. L'absence d'acide gras dans le domaine II induit une très faible déviation. Ces résultats montrent l'aptitude de la ASH à se « déformer » en présence de composés extérieurs. Plusieurs structures 3D de complexes entre la ASH et des acides gras de tailles variables ont été obtenues par la suite et ont permis d'identifier deux sites d'accueil additionnels (FA6 et FA7).^[50-52]

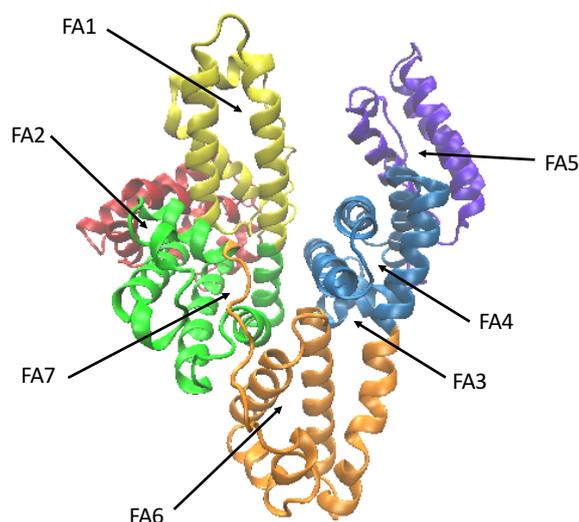


FIGURE 3 – Schéma de la protéine ASH et des 7 sites de fixation des acides gras (FA1-7).

Hème et produits de dégradations

Le transport vers le foie des hèmes et de ses produits de dégradation comme la bilirubine est assuré par la ASH. Les composés sont métabolisés dans le foie et ensuite excrétés du corps par l'urine et les selles. Les structures cristallographiques de ces molécules avec la ASH indiquent

qu'ils se placent au niveau du FA1, situé dans le sous-domaine IB (en jaune sur la Figure 3), en raison de leur grande taille.^[53-55] De plus, le sous-domaine IB présente de nombreux résidus polaires, comme Arg114, Arg117, His146, Arg186 et Lys190, qui stabilisent les fonctions propanoates de l'hème, de la bilirubine et de leurs dérivés.

Composés exogènes et médicaments

Un autre rôle majeur de la ASH est le transport de médicaments au sein de l'organisme. La fixation de ces composés s'effectue essentiellement sur les sites I et II de Sudlow, localisés respectivement dans les sous-domaines IIA et IIIA.^[41,43] Le premier et le deuxième sites de Sudlow peuvent être aussi occupés par des acides gras car ils correspondent respectivement au site FA7 et FA3-FA4. La poche hydrophobe du premier site est composée de trois sous-chambres et de deux groupes de résidus polaires qui sont situés à l'entrée (Lys195, Lys199, Arg218 et Arg222) et à l'intérieur (Ser150, His242 et Arg257) de la poche. Le deuxième site, plus petit que le premier, dispose aussi de résidus polaires à son entrée (Arg410, Tyr411, Lys414 et Ser489). Les deux sites se distinguent clairement par la nature des composés (taille, polarité, forme) qu'ils peuvent chacun accueillir. Le premier site de Sudlow (sous-domaine IIA) peut, dans certains cas, être obstrué par le domaine III car ce dernier est très flexible. Le deuxième site de Sudlow, situé à l'extrémité de la protéine, est plus exposé au solvant, le rendant plus accessible. Il est également possible que des composés soit présents au sein des deux sites de Sudlow, avec des affinités relatives à chaque site différentes.

De nombreuses structures cristallographiques de complexes ASH-médicament sont disponibles au sein de la PDB permettant ainsi de mieux comprendre l'étape de distribution dans le processus pharmacocinétique ADME de ces composés. Des médicaments de type anti-inflammatoires comme la warfarine, l'aspirine, le phénylbutazone, l'azapropazone, etc., vont avoir une plus grande affinité avec le site I de Sudlow^[31,52,56] car ce dernier est assez volumineux pour les accueillir, tandis que d'autres vont se placer au sein du site II de Sudlow comme l'ibuprofène, le propofol, le diazépam, etc.^[31,57] D'autres molécules thérapeutiques peuvent se complexer avec la ASH au sein du site FA1 (IB) comme c'est le cas pour la lidocaïne, le diclofénac et le 3'-azido-3'-déoxythymidiène, un agent anti-rétroviral.^[58-60] Une série de dérivés d'acides aminés avec la ASH fut cristallisé par Ryan *et al.*^[61] Notons que ces dérivés sont ceux utilisés comme sonde fluorescente dans les travaux de Sudlow en 1975. En 2015, Ferraro *et al.* réussirent à obtenir la structure 3D du complexe ASH avec le cisplatine, un agent anti-cancéreux.^[19] Les atomes de platine se lient principalement avec des résidus de type méthionine et histidine, présents à la surface de la protéine.

2 Cas d'études portés sur la ASH

L'avènement des structures cristallographiques de la ASH a permis d'obtenir des informations structurales précieuses, notamment sur la structure secondaire et tertiaire de la protéine. Cependant, les structures cristallographiques donnent peu d'information en ce qui concerne la dynamique de la protéine, les mouvements internes, les mécanismes et les conformations adoptées lors de l'accueil d'un composé. Il est indéniable que des analyses supplémentaires sont nécessaires pour une meilleure compréhension des phénomènes biologiques impliquant la ASH et c'est en cela que les méthodes théoriques peuvent être un atout très puissant. Depuis une trentaine d'année et l'essor de l'informatique, il est devenu possible de modéliser des systèmes biologiques de très grande taille (> 100 000 atomes) dans le but d'obtenir des informations complémentaires à l'expérience.

Ce chapitre consiste à présenter quelques travaux qui ont apporté, à l'aide de la modélisation moléculaire, des réponses à certaines hypothèses concernant la protéine ASH. Dans un premier temps, trois problématiques concernant la ASH sont présentées, illustrant ainsi l'apport de la modélisation moléculaire pour résoudre des problématiques de biologie structurale ; puis dans un second temps, quelques exemples des différents travaux traitant des interactions ASH-médicament sont exposés.

2.1 Problématique de biologie structurale

2.1.1 États de la protonation de la Lys199

L'état de protonation de la Lys199, se trouvant au sein du premier site de Sudlow, est sujet à débat car Walker en 1976 remarqua qu'une réaction d'acétylation a lieu entre l'acide acétylsalicylique (aspirine) et la Lys199 de la ASH.^[62] Ces données suggèrent donc que la Lys199 est sous la forme neutre (NH_2) et possède un caractère nucléophile plus marqué que les autres lysines, se présentant sous leur forme protonée (NH_3^+). La réaction entre la Lys199 et l'aspirine a été étudié, par la suite, par des analyses en RMN ^{13}C et divers travaux expérimentaux centrés sur la réactivité de la Lys199 et son influence ont été menés.^[63-65] La faible valeur de $\text{p}K_a$ de cette lysine ($\text{p}K_a = 7,9$)^[66-68] en comparaison avec les autres lysines protonées ($\text{p}K_a = 11$) est attribuée à une forte interaction qu'elle possède avec l'histidine en position 242.^[69]

En 2001, Diaz *et al.* réalisent les premières simulations de dynamique moléculaire (*Molecular Dynamics*, MD) sur la ASH afin d'étudier l'influence de l'état de protonation de la Lys199 sur la

poche hydrophobe du site I de Sudlow (sous-domaine IIA).^[70]

Cette Lys199 est proche d'une autre Lysine, la Lys195, qui elle est protonée. Trois systèmes fictifs avec des états de protonations différents sont étudiés : le système Lys199-Lys195H⁺, le système Lys199H⁺-Lys195H⁺ et le système Lys199H⁺-Lys195. Ce dernier résulte d'un transfert de proton entre la Lys195 vers la Lys199. Une simulation de MD est réalisée pour chacun des systèmes. Le système Lys199-Lys195H⁺ présente une poche hydrophobe plus large et flexible que les deux autres systèmes où la Lys199 est protonée. La protonation du résidu Lys199 induit un changement de conformation de deux résidus voisins, Phe211 et Trp2124, modifiant ainsi la poche hydrophobe. De plus, une interaction de nature π -cation entre les résidus Tyr150 et Lys199 apparaît quand cette dernière est protonée, ce qui a pour conséquence d'obstruer l'entrée du site. La protonation de Lys199 affecte des interactions entre des résidus à son voisinage, notamment, le pont salin entre la Lys195 et Glu292 qui se trouve affaibli par ce changement conformationnel. Les auteurs indiquent que le système Lys199-Lys195H⁺ induit de meilleures conditions d'accueil pour un composé au sein du premier site de Sudlow et que, néanmoins, un transfert de proton entre les résidus Lys195 et Lys199 est théoriquement possible par l'intermédiaire de molécules d'eau pontant les deux résidus. Dans nos simulations, nous avons pris en compte cette étude et considéré que la Lys199 est neutre et que la Lys195 est protonée.

2.1.2 Importance des ponts disulfures sur la structure de la ASH

La structure tridimensionnelle des protéines est régie par leur structure secondaire (hélice α , feuillet β , liaison hydrogène, etc.) et tertiaire (ponts-salins, liaisons disulfure, etc.), et est très importante car cela caractérise la fonction et les propriétés de cette dernière. Concernant la ASH, sa structure secondaire constituée exclusivement d'hélices α est un élément crucial pour remplir son rôle de protéine vectrice d'une grande diversité de molécules au sein du corps humain. En outre de l'arrangement des résidus en hélice α , les ponts disulfures sont impliqués dans le repliement spatial des protéines et la ASH en possède 17, ce qui est un nombre très élevé par rapport à la plupart des protéines. Il est intéressant de se questionner sur l'importance des ponts disulfures sur les propriétés structurelles et biologiques de la ASH. La réduction des ponts disulfures peut, au sein de certaine protéine, dénaturer totalement la conformation de la protéine et, par conséquent, ses propriétés.^[71-74] Expérimentalement, l'absence de ponts disulfures induit un changement du spectre Raman de la ASH, traduisant des modifications de la structure secondaire de la protéine.^[75] En 2012, les travaux de Paris *et al.* ont porté sur la simulation du comportement de la protéine ASH, avec et sans ponts disulfures grâce à des simulations de MD.^[76] À une température de 300 K, les deux systèmes sont très similaires, amenant les auteurs à suggérer que le changement conformationnel de la ASH est un processus lent. Cependant, à une température de 500 K, les deux systèmes ont un comportement différent lors des simulations, notamment sur le pourcentage d'hélice α qui diffère d'environ 15 %. Cet écart a pour effet de déstabiliser la structure secondaire de la protéine. De plus, la structure tertiaire se voit aussi largement modifiée au vu des grandes distances entre

les cystéines supposées créer un pont disulfure et l'augmentation du volume de la protéine. Ces changements de conformation vont atténuer la capacité de la protéine à accueillir divers composés et, par conséquent, peuvent modifier ses fonctions biologiques. Une autre étude a permis de montrer, à travers des simulations de 70 ns, que certains ponts disulfures jouaient un rôle très important dans la structure de la protéine.^[77] Parmi ces derniers, le pont créé entre la Cys168 et la Cys177 est responsable de la géométrie du premier site de Sudlow.

2.1.3 Influence des acides gras

En 1998, Curry *et al.* comparent la structure adoptée par la ASH en présence et en absence d'acides gras, mettant ainsi en lumière la grande flexibilité des domaines I et III.^[49] En effet, la ASH est capable de transporter entre 0,1 et 2 moles d'acide gras par mole de ASH.^[78] Dès lors, la question liée à la présence d'acides gras devient légitime pour explorer le comportement et l'affinité d'un ligand avec la ASH^[79].

Fujiwara *et al.* ont réalisé des simulations de MD de la ASH en présence et en l'absence d'acides gras et, ils ont analysé les changements conformationnels de ces deux systèmes.^[80] La grande flexibilité des domaines I et III est également observée. La présence d'acides gras induit une fluctuation significative des résidus spécifiques du site I, suggérant que l'affinité de certains composés pour ce site est affectée par la présence au préalable d'acides gras au sein de la ASH. Expérimentalement, ce phénomène a été observé dans le cas de la warfarine.^[81] À partir des trajectoires des simulations de MD, une analyse en composantes principales permet d'identifier trois différents mouvements d'ensemble de la protéine. L'influence des acides gras induit des mouvements différents que ceux de la ASH seule. Plus tard, les affinités de chacun des sept sites accueillant des acides gras de la ASH ont été calculées *via* des calculs MM/PBSA (*Molecular Mechanics/Poisson-Boltzmann Surface Area*).^[82] En comparant les simulations de MD de la ASH complexée avec sept acides gras, sept autres simulations de la ASH complexée à chaque fois avec un seul acide gras dans chacun des sites sont effectuées. Les sites FA2, FA4 et FA5 sont identifiés comme étant ceux possédant les meilleures affinités avec la ASH tandis que les sites FA1 et FA3 sont considérés comme les sites les moins affins.

L'influence de la présence d'acide gras sur l'affinité de la warfarine avec le site I de Sudlow de la ASH a aussi été quantifiée par des calculs MM/PBSA.^[83] Ce qui ressort de cette étude est que le placement d'un seul acide gras dans le deuxième site a un effet stérique qui déstabilise grandement l'affinité de la warfarine mais cette dernière se voit augmenter lorsque trois acides gras sont présents dans les sites FA2, FA4 et FA5. L'occupation de ces trois sites induit une diminution de l'entropie de la protéine, donnant lieu à un effet allostérique favorable entre la ASH et la warfarine. Un second phénomène allostérique est observé lorsque deux acides gras sont localisés dans les sites FA4 et FA5. Une stabilisation enthalpique est obtenue impliquant les résidus Gln196, Lys199 et Arg257.

2.2 Complexe ASH-ligand

La capacité extraordinaire de la ASH à accueillir diverses molécules a été largement étudiée depuis une vingtaine d'année et l'utilisation de méthodes théoriques pour confirmer le mode de liaison du complexe protéine-ligand est devenu de plus en plus courant. Les méthodes de spectroscopie UV-visible et de dichroïsme circulaire sont des techniques expérimentales très utilisées pour identifier le ou les site(s) d'accueil de plusieurs molécules (palmatine, indolone *N*-oxyde, dérivé pyrimidine, etc.) au sein de la ASH.^[84-86] L'utilisation de calculs de *docking* et de simulation de MD s'est révélé être très complémentaire de ces méthodes expérimentales, notamment, dans le cas de la β -sitostérol et de la β -carboline.^[87,88] La localisation de la myricétine au sein du premier site de Sudlow a pu être confirmée par des calculs de *docking* moléculaire, complémentaire à des analyses de spectroscopie infrarouge à transformée de Fourier et des mesures d'absorption et de fluorescence.^[89] Dans un autre registre, des calculs quantiques et de *docking* moléculaires ont été utilisés, en complément de données cinétiques, pour caractériser le caractère catalytique de la Lys199 sur la molécule de benzisoxazole.^[90]

En plus de permettre de compléter des résultats expérimentaux, la modélisation moléculaire a été largement employée pour prédire le mode de liaison de nombreux complexes. En 2001, Colmenarejo *et al.* ont développé plusieurs descripteurs, dont certains sont issus de calculs semi-empiriques, dans un modèle de chémoinformatique *in silico* permettant de prédire l'affinité de molécules avec les deux sites principaux de la ASH.^[91] Un jeu de 95 molécules diverses a été utilisé et les résultats ont souligné que le caractère hydrophobe est un des critères de choix pour une bonne affinité avec la protéine. Quelques années plus tard, Deeb *et al.* ont repris le même jeu de molécules et ont utilisé une approche combinant des calculs de *docking* moléculaire avec des simulations de MD.^[92] Ils mettent en évidence que cette dernière association permet d'améliorer la prédiction des valeurs d'affinités et que la relaxation de la protéine joue un rôle important dans la reconnaissance des molécules par la protéine pour former un complexe.

Parmi les molécules complexées avec la ASH, la warfarine est sans doute l'une des plus connues. Cette molécule est usuellement utilisée comme marqueur du premier site de Sudlow^[41,93] et il n'est donc pas étonnant de la retrouver dans certaines études comme référence.^[94,95] Elle fait partie bien évidemment du jeu de molécules utilisé par Colmenarejo *et al.* et a fait l'objet de nombreux travaux. Parmi ceux-là, l'énantiosélectivité de la ASH pour la warfarine a été analysée par *docking* moléculaire,^[96] ainsi que l'influence d'acides gras sur son mode de liaison et son affinité avec la protéine de transport.^[83] En 2013, Fujiwara *et al.* font un état de l'art de la littérature sur les complexes protéine-ligand impliquant la ASH traitée par modélisation moléculaire et y dénombre plus de 60 cas différents.^[97] Divers médicaments (anti-Parkinson, anti cancéreux, paracétamol, etc.) transportés par la ASH y sont répertoriés.^[98-101] Les auteurs soulignent que l'avènement de nouvelles structures cristallographiques a fortement contribué aux nombreuses études menées sur les ASH-médicaments.

3 Les inhibiteurs de Rad51

Dans le cadre de notre thèse de doctorat, l'un des acteurs principaux de nos travaux est une famille de molécules capable d'inhiber la protéine Rad51. La protéine Rad51 est étudiée par l'équipe de Fabrice Fleury de Nantes (UFIP, Unité Fonctionnalité et Ingénierie des Protéines) avec qui nous collaborons sur ce projet. Elle joue un rôle central dans la recombinaison homologue (RH) de l'ADN, qui est l'un des mécanismes de réparation lors des cassures double brin (CDB). Cette lésion, qui peut être causée par différents facteurs (d'origine exogènes ou endogènes, Figure 4), représente un réel danger pour la préservation de l'intégrité du matériel génétique (chromosomes) de la cellule.^[102]

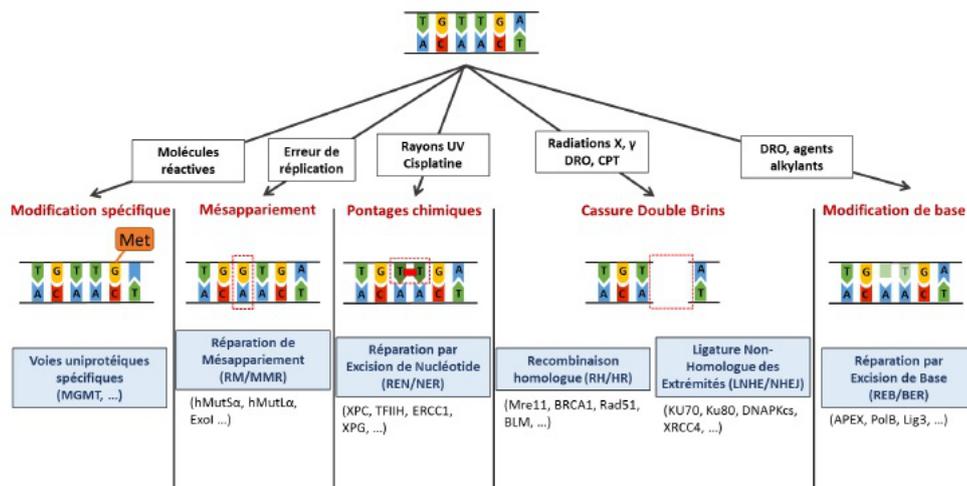


FIGURE 4 – Schéma regroupant l'origine et les différents types d'endommagement que l'ADN peut subir. Figure provenant de la référence 103.

L'apparition de ce type de lésions, au sein de la cellule, provoque une succession d'étapes : de détection de la lésion, de recrutement de/des protéines réparatrices, et de réparation du (ou des) brin(s) d'ADN. La voie de réparation par RH est principalement présente durant les phases S et G2 du cycle cellulaire, phases où l'ADN est complètement décondensé et vulnérable.^[104] Plusieurs protéines successives (Mre11, Rad50, NBS1, Rad52, BRCA2)^[105–107] sont impliquées dans le recrutement de Rad51. Cette dernière se polymérise sur l'ADNsb (ADN simple brin) pour former un nucléofilament. La super-structure ainsi formée a la capacité de rechercher une séquence homologue d'un ADNsb dans un ADNdb (ADN double brin). Dès que cette séquence est localisée, la protéine RAD51 catalyse l'échange du nucléofilament par la séquence homologue de l'ADNdb

résultant en un segment d'ADNsb homologue au site de cassure, jouant ainsi le rôle d'une nouvelle matrice permettant à des polymérases de reconstituer l'ADN de ce dernier. La Figure 5 illustre le mécanisme de réparation de l'ADN. La protéine Rad51 joue un rôle crucial dans le mécanisme de RH et l'action de certaines molécules interagissant avec elle peut mener à la modulation du mécanisme de réparation de l'ADN.

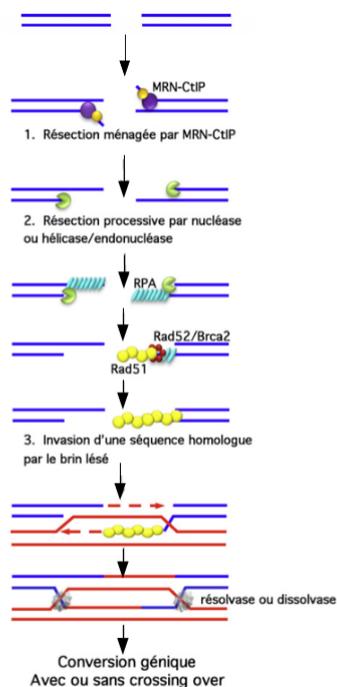


FIGURE 5 – Mécanisme de réparation par recombinaison homologue de l'ADN endommagé. Figure provenant de la référence 102.

3.1 Les familles d'inhibiteurs

La plupart des composés identifiés comme des inhibiteurs de Rad51 a été découvert suite au criblage de bases de données de composés chimiques. Parmi les inhibiteurs ciblant la protéine Rad51, certains d'entre eux modulent l'activité de cette dernière alors que d'autres modulent son expression au sein de la cellule.

3.1.1 Composés modulant l'expression de Rad51

Le méthotrexate, l'acide phénylhydroxamique (ou PCI-24781) et la prodigiosin (Figure 6) font partie des composés connus pour moduler l'expression de Rad51.

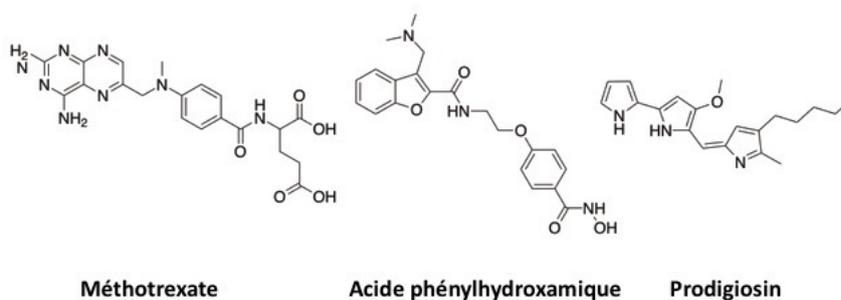


FIGURE 6 – Composés modulant l’expression de Rad51.

Le méthotrexate intervient lors de la synthèse des bases nucléiques durant la phase S du cycle cellulaire et cible les fonctions de la RH en réduisant le taux de réparation des dommages à l’ADN.^[108] L’inhibition de la protéine Rad51 a déjà été observée dans une lignée cellulaire humaine d’ostéosarcome après irradiation.^[109]

Le PCI-24781 appartient aux molécules inhibant l’*histone deacetylase* (HDACs), connues pour leur activité anti-tumorale.^[110] Plus précisément, il agit sur l’HDAC2, qui est impliquée dans la régulation des facteurs de la RH. Selon les travaux d’Adimoolam *et al.*, l’inhibition de l’HDAC par cette molécule réduit l’expression de Rad51 dans les lignées cellulaires dérivant du carcinome colorectal et donc peut réduire la réponse de la RH due à la cassure du double brin d’ADN induite par une thérapie.^[111] PCI-24781 pourrait de manière indirecte activer le clivage de Rad51 et ce, en plus d’induire une diminution de son expression. Le composé peut également limiter la croissance de cellules cancéreuses car elle induit l’expression du gène de la protéine GADD45y, un facteur lié à l’arrêt de la croissance cellulaire.^[112]

La prodigiosin est un pigment tripyrrole rouge issu d’une bactérie ayant été récemment identifiée comme possédant des propriétés immunosuppressives et anti-cancéreuses^[113–115] Bien que la régulation négative de Rad51 ne soit pas encore entièrement comprise, il semblerait que la prodigiosin soit un inhibiteur potentiel de Rad51 qui pourrait être utilisé pour contourner la résistance due aux médicaments (thérapie) agissant sur la RH des cellules cancéreuses.

3.1.2 Modulateurs de l’activité de la RH

Dans ce cas, les composés agissent directement sur les étapes catalytiques de l’activité de recombinase Rad51 et incluent la molécule que nous avons étudiée lors de notre travail de thèse.

Le métatungstate est un polyoxométallate (POM) composé de 12 oxoanions de tungstène (Figure 7) qui se dissocient en monoanions en solution aqueuse alcaline. Ce POM est souvent utilisé comme agent catalytique pour les réactions chimiques des hydrocarbures. Li *et al.* ont démontré que la structure de métatungstate pouvait se lier *in vitro* à un homologue de la protéine Rad51.^[116] Les zones de contact établies entre la protéine et le métatungstate impliquent deux boucles qui interviennent également dans l’interaction Rad51-ADN ; une fois formées, les interactions stabilisent la conformation inactive de Rad51.^[117–119]

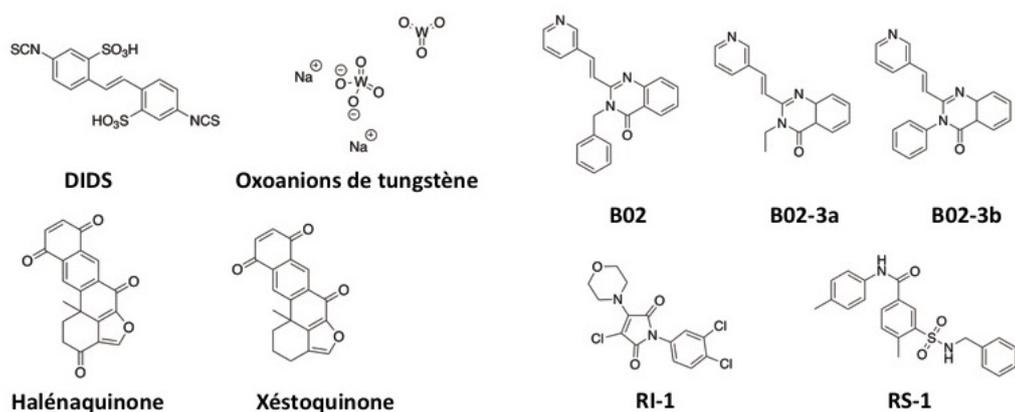


FIGURE 7 – Composés modulant l'activité de RH de Rad51.

Les molécules de xéstoquinone et d'halénaquinone (Figure 7) sont très similaires (un seul atome d'oxygène diffère). L'halénaquinone inhibe la RH au niveau des étapes d'appariement de l'ADN et de la formation de la boucle D, mais aucun effet n'a été observé avec la xéstoquinone.^[120] Lorsque des cellules humaines sont exposées à un rayonnement ionisant et ensuite traitées avec de l'halénaquinone, la formation de foyers de la protéine Rad51 diminue significativement. Ce résultat indique donc que l'halénaquinone déstabilise l'apparition de foyers de Rad51, probablement par inhibition de la formation d'un complexe ternaire.

La molécule B02 (Figure 7) a été identifiée comme un inhibiteur potentiel de Rad51 après un criblage de 200 000 molécules issues de la base de données de la NIH.^[121] Elle présente un IC_{50} de 27,4 μM mais a une spécificité élevée pour HsRad51 (*Homo sapiens* Rad51). En outre, l'étude des dérivés B02 a révélé une inactivation efficace de la protéine Rad51 pour deux dérivés, B02-3a et B02-3b, qui contiennent, respectivement, un groupe éthyle et un groupe *m*-méthylphényle au lieu du groupe benzyle de B02 (Figure 7). La modification de la pyridine de B02 supprime l'inhibition de la boucle D induite par le mécanisme de RH de Rad51, ce qui démontre l'importance de cette fonction chimique. Un travail *in vivo* récent a montré que B02 inhibait les RH induites par des CDB et augmentait la sensibilité cellulaire aux agents de pontage intra-brin comme le cisplatine et la mitomycine C.^[122]

La molécule RI-1 (3-chloro-1-(3,4-dichlorophényl)-4-(4-morpholinyl)-1H-pyrrole-2,5-dione, Figure 7) a été identifiée comme un inhibiteur de HsRad51. Cette molécule s'est distinguée lors d'un criblage de 10 000 molécules issues de la base de données de Chembridge (criblage par polarisation de fluorescence puis par un test d'inhibition de la RH d'une lignée cellulaire d'ostéosarcome humain).^[11] RI-1 est composé d'un groupement chloromaléimide qui favorise une liaison covalente avec le groupe thiol de la Cys319 de la protéine Rad51, par un mécanisme d'addition de Michael. Cette liaison covalente est responsable de l'inhibition de la polymérisation de HsRad51 lors de la formation du nucléofilament. Le site de liaison est situé sur l'interface entre deux monomères de HsRad51 de sorte qu'il inhibe la polymérisation de Rad51 sur ADNs.^[123] L' IC_{50} de RI-1 est de 5 à 30 μM en fonction de la concentration intracellulaire de HsRad51 et est aussi un

inhibiteur potentiel spécifique pour les homologues mammifères de Rad51.

Le RS-1 (3-[(benzylamino)sulfonyl]-4-bromo-N-(4-bromophényl)benzamide, Figure 7) a également été découvert *via* un criblage de molécules issues de la base de données de Chembridge, mais ce composé stimule la liaison de la protéine Rad51 sur l'ADNsb.^[124] Cela augmente donc la stabilité du nucléofilament et, contrairement aux inhibiteurs, cette molécule favorise la RH. L'action stimulante de RS-1 est spécifique à la protéine HsRad51 et peut, dans certains cas, augmenter la résistance des cellules face à des agents anticancéreux. En effet, une concentration de 7,5 μM de RS-1 permet de compenser les dommages du cisplatine.^[124]

Le DIDS (ou 4,4'-Diisothiocyanostilbene-2,2'-disulfonic acid, Figure 7) est une molécule connue depuis le début des années 1970 pour sa capacité à inhiber les canaux ioniques.^[125-137] Par exemple, en tant qu'inhibiteur non-sélectif, le DIDS module les transporteurs de membranes et régule par conséquent les maladies associées.^[138-140] Dans la famille des protéines de transport transmembranaires CIC (*Chloride Carrier/Channel*), le DIDS bloque de manière réversible le canal d'échange d'ions Cl^-/H^+ ^[141] et inhibe de façon irréversible l'antiport $\text{Cl}^-/\text{HCO}_3^-$, les deux canaux ayant un rôle crucial au niveau de la fonction cardiaque (ischémie).^[126,131] La fonction de l'échangeur d'anions 1 (AE1 ou Band 3, protéine membranaire la plus abondante au sein des globules rouges) est également inhibée par le DIDS, cette protéine transporte de petits anions, *e.g.*, Cl^- ou HCO_3^- , mais aussi des substrats, *e.g.*, sulfate, phosphate, phosphoénolpyruvate et pyridoxal 5-phosphate.^[133-137] Plus récemment et c'est ce qui nous intéresse dans la thèse, il a été constaté par nos collaborateurs que le DIDS inhibait directement Rad51.^[113,142-144] En ciblant Rad51, il a été démontré que le DIDS empêchait la formation de foyers Rad51 et favorisait un effet anti-leucémique.^[142,144] Le mode d'action de l'inhibiteur sur Rad51 ainsi que leurs interactions spécifiques affectant le mécanisme de réparation de l'ADN restent inconnus malgré une recherche active dans le domaine.^[10,143,145]

Plusieurs autres dérivés du stilbène disulfonique (voir Figure 8) pour la représentation de certains dérivés), *e.g.*, DNDS (acide 4,4'-dinitrostilbène-2,2'-disulfonique) et SITS (acide 4-acétamido-4'-isothiocyano-stilbène-2,2'-disulfonique) sont également connus pour affecter de manière significative les fonctions des canaux. En bref, le DIDS est une molécule d'intérêt biologique élevé qui est largement utilisée expérimentalement, bien que son mécanisme de liaison et son mode de réactivité exacte restent entièrement flous. Les dérivés d'acide sulfonique du stilbène étudiés dans notre travail de thèse sont présentés à la Figure 8.

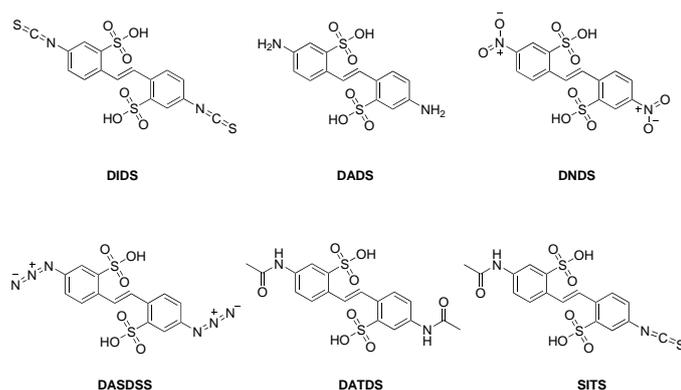


FIGURE 8 – Structures de la série des stilbènes disulfoniques étudiés.

3.2 Les stilbènes disulfoniques

3.2.1 Photochimie du stilbène

D'innombrables études photochimiques ont été menées expérimentalement et théoriquement sur la molécule de stilbène.^[146–153] En effet, cette dernière est une molécule modèle et de choix pour étudier la photoisomérisation *cis* ↔ *trans*.^[154] Sous irradiation lumineuse, le stilbène peut subir deux isomérisations, *cis* à *trans* et *trans* à *cis*, en fonction de la longueur d'onde sélectionnée.

En chimie théorique, le processus de photocyclisation complet ne peut être capturé qu'avec un modèle multidimensionnel,^[155–158] les coordonnées de réaction unidimensionnelle liées à une torsion sont insuffisantes pour capturer entièrement la photodynamique de l'isomère *cis* dans son état excité qui conduit à la formation d'un produit photocyclique.^[154,159,160] Pour cela, des approches théoriques robustes sont habituellement employées pour décrire les états excités du stilbène. L'approche du champ auto-cohérent de l'espace actif complet (*Complete Active Space Self-Consistent Field*, CAS-SCF)^[161] et son pendant employant la théorie des perturbations (*Complete Active Space Perturbation Theory 2*, CASPT2)^[162] sont employées pour modéliser le chemin d'énergie minimale et les surfaces d'énergie potentielle associées.^[151] Les méthodes multi-états CASPT2 (*multistate CASPT2*, MS-CASPT2)^[163] ou encore les approches de type *coupled cluster* sont aussi utilisées pour décrire cette molécule modèle.^[164,165]

En plus des approches multi-références citées ci-dessus, la théorie de la fonctionnelle de densité (*Density Functional Theory*, DFT)^[166,167] et sa composante dépendante du temps (*Time-dependent-DFT*, TD-DFT)^[168,169] ont aussi déjà été employées pour étudier l'état triplet de stilbène^[170] et les surfaces d'énergie potentielles de l'état fondamental et excité.^[153,171] Il est intéressant de noter que pendant plusieurs années, l'ordre énergétique des deux états B_u du *trans*-stilbène était un débat car les prédictions obtenues à l'aide de la TD-DFT^[152,153] ne correspondaient pas aux résultats obtenus avec les méthodes multi-références.^[149,150] En 2009, Angeli *et al.* ont montré que l'ordre obtenu par la TD-DFT était le plus précis.^[172] La TD-DFT a également été couplée à des simulations

de dynamique non-adiabatique pour étudier la photoisomérisation du stilbène.^[173–175] De plus, la *spin-flip* TD-DFT a été utilisée pour mettre en évidence l'intersection conique,^[176] et les couplages vibroniques ont également été évalués pour le *trans*-stilbène.^[177]

3.2.2 Propriétés physico-chimiques du DIDS

Contrairement au stilbène, et à notre connaissances, aucune étude systématique concernant les propriétés du DIDS et de ses dérivés (Figure 8) n'était présente dans la littérature au début de ce travail. Les propriétés des dérivés d'acide sulfonique du stilbène sont très peu connues. Il y a vingt-sept ans, le travail séminal de Jakobsen *et al.* dédié au SITS (Figure 8) alarmait les biologistes de l'utilisation de tels composés en raison de leur rapide et importante décomposition en solution aqueuse.^[178] Ils ont montré que le stockage de tels composés dans l'obscurité n'est pas suffisant pour empêcher la décomposition du SITS et que le milieu de stockage devait également être déshydraté. Il en va de même pour les autres dérivés tels que le DADS, le DNDS et le DIDS.

Plus récemment, une modification du DIDS, non-négligeable, est observée expérimentalement (HPLC et RMN) après que ce dernier soit resté en solution pendant 24 heures.^[179] Matulef *et al.* suggèrent que les deux groupements isothiocyanate (-NCS) s'hydrolysent pour former des amines primaires, conduisant à la formation du dérivés DADS. Cependant, un mélange contenant à la fois le DADS et le DIDS conduit notamment à la formation d'un dimère ayant une liaison de type thiourée, et ce dû à la réaction (rapide) entre les fonctions amino et isothiocyanate (Figure 9).

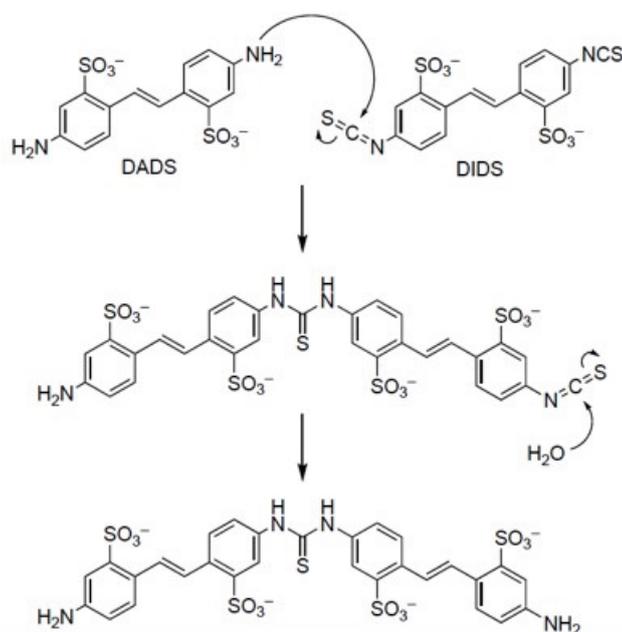


FIGURE 9 – Mécanisme de dimérisation entre une molécule de DIDS et de DADS. Figure provenant de la référence 179.

Ce mécanisme pourrait alors induire une multimérisation des entités DIDS et DADS, menant à la formation de polythiourées de divers poids moléculaires.^[179] Une nouvelle fois, Matulef *et*

al. avertissent clairement que la nature du DIDS solvaté par l'eau change complètement son comportement en un seul jour, ce qui est en accord avec les conclusions des travaux de Jakobsen *et al.* Il est donc surprenant que le comportement et la cinétique du DIDS dans des conditions biologiques ait été totalement négligé à ce jour. En réalité, il semble que le DIDS soit actuellement utilisé sans attention particulière dans de nombreux laboratoires biologiques. Dans cette optique, une étude combinant expérience et théorie, visant à étudier les propriétés structurales et spectrales de DIDS et de ses dérivés en phase condensée est la bienvenue (cfr. Chapitre 5). Ceci est une étape préliminaire pour mieux comprendre les interactions entre la molécule de DIDS et les protéines.

Deuxième partie

Propriétés quantiques en solution des SD

4 Les méthodes *ab initio* en chimie quantique

Ce chapitre (et le Chapitre 6) a été fortement inspiré par les cours suivis durant mon Master et par des ouvrages traitant de la chimie quantique.^[180-185] Il reprend, brièvement, les théories et principes utilisés dans notre thèse de doctorat.

Un des objectifs principaux de la chimie quantique est de résoudre l'équation de Schrödinger, qui, dans sa forme première, se présente avec une dépendance temporelle :

$$H\Psi(r, t) = i\frac{\partial\Psi(r, t)}{\partial t}, \quad (4.1)$$

où le terme H désigne l'opérateur Hamiltonien, $\Psi(r, t)$ est la fonction d'onde du système étudié, dépendante des coordonnées spatiales et temporelles. La résolution de l'équation donne accès à toutes les propriétés du système dépendantes du temps. Cependant, cette équation peut être simplifiée pour étudier les problématiques indépendantes du temps, donnant lieu à une fonction d'onde qui ne dépend que des paramètres spatiaux :

$$H\Psi(r) = E\Psi(r), \quad (4.2)$$

où E est l'énergie à valeur propre de l'opérateur Hamiltonien H dont la fonction d'onde $\Psi(r)$ est le vecteur propre. L'opérateur Hamiltonien décrit l'ensemble des phénomènes qui ont une incidence sur l'énergie et est formé par la somme des contributions cinétiques, T , et potentielles, V , du système étudié :

$$\begin{aligned} H &= T + V, \\ &= -\underbrace{\sum_{A=1}^K \frac{1}{2M_A} \nabla_i^2}_{T_n} - \underbrace{\sum_{i=1}^N \frac{1}{2} \nabla_i^2}_{T_e} - \underbrace{\sum_{A=1}^K \sum_{i=1}^N \frac{Z_A}{r_{iA}}}_{V_{ne}} + \underbrace{\sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}}}_{V_{ee}} + \underbrace{\sum_{A=1}^K \sum_{B>A}^K \frac{Z_A Z_B}{r_{AB}}}_{V_{nn}}, \end{aligned} \quad (4.3)$$

où les termes T_n et T_e sont, respectivement, les énergies cinétiques des noyaux (n) et des électrons (e). V_{ne} décrit l'énergie d'attraction coulombienne entre les noyaux et les électrons, et V_{ee} et V_{nn} sont, respectivement, les énergies de répulsions électroniques et nucléaires. M_A est la masse du noyau A , Z_A et Z_B sont les charges effectives de chaque noyau, r_{AB} est la distance entre les deux noyaux A et B , r_{iA} est la distance entre l'électron i et le noyau A . ∇_i^2 est l'opérateur Laplacien qui s'écrit en coordonnées cartésiennes sous la forme :

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} . \quad (4.4)$$

Malheureusement, l'équation de Schrödinger ne peut être résolue de manière exacte que pour un problème à deux corps, ce qui se limite aux cas de l'hydrogène et des hydrogénoïdes. Dans le cas de systèmes possédant plusieurs atomes, comme c'est le cas pour les composés de cette thèse, l'introduction d'approximations s'avère nécessaire. La première d'entre elles est l'approximation de Born-Oppenheimer (BO) qui se base sur la différence significative entre la masse des noyaux et la masse des électrons (un facteur ≈ 1800 pour l'atome d'hydrogène). Suite à cela, les mouvements des électrons peuvent être découplés de ceux des noyaux, permettant de séparer en deux parties l'équation de Schrödinger (électronique et nucléaire) et ainsi simplifier sa résolution. Cette approximation considère que les noyaux sont « fixes » pour la partie électronique et, qu'ainsi, les mouvements des électrons sont étudiés dans une configuration nucléaire donnée.

4.1 Hartree-Fock

La méthode Hartree-Fock (HF) permet de résoudre la partie électronique de l'équation de Schrödinger pour des systèmes poly-électroniques. Cependant, la prise en compte d'approximations supplémentaires est nécessaire. Dans la méthode HF, l'interaction d'un électron avec chacun de ses voisins est remplacée par une interaction entre cet électron et le champ effectif créée par les autres, qui est désignée comme l'approximation du champ moyen. Elle permet de transformer l'opérateur Hamiltonien poly-électronique en un jeu d'Hamiltoniens mono-électroniques. Pour un système à couches fermées, on écrit ainsi :

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^K \frac{Z_A}{r_{iA}} + \underbrace{\sum_{j,j \neq i}^{N/2} (2J_j(r_i) - K_j(r_i))}_{V_{\text{eff}}} . \quad (4.5)$$

Le terme V_{eff} comprend un terme de répulsion électronique coulombienne ($J_j(r_i)$) créé par les autres électrons du système et un terme correctif provenant du caractère antisymétrique de la fonction d'onde ($K_j(r_i)$). Ce dernier prend en compte le principe de Pauli dans les équations HF et est communément appelé terme d'échange. Les formes mathématiques des deux termes sont :

$$J_j \phi_i(r_i) = \int \frac{\phi_j^* \phi_j}{r_{ij}} \phi_i dr , \quad (4.6)$$

$$K_j \phi_i(r_i) = \int \frac{\phi_j^* \phi_i}{r_{ij}} \phi_j dr . \quad (4.7)$$

Au sein du formalisme de la théorie HF, il est possible que deux électrons de spins opposés puissent se trouver simultanément au même endroit de l'espace, ce qui engendre une sur-représentation de situations énergétiquement défavorables. Cette erreur provient du fait que la théorie HF ne prend pas en compte la corrélation de Coulomb ou, autrement dit, la corrélation

entre les électrons de spin opposé. Les énergies obtenues en HF sont donc toujours supérieures à l'énergie exacte, et l'énergie de corrélation (E_{corr}) électronique est définie comme l'écart énergétique entre les deux énergies :

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}} . \quad (4.8)$$

Pour corriger cette erreur, une des approches est la théorie de la fonctionnelle de densité, qui a été employée lors de notre travail de thèse.

4.2 La Théorie de la Fonctionnelle de Densité

La théorie de la fonctionnelle de densité (*Density functional Theory*, DFT) se base sur la densité électronique $\rho(r)$, et non sur la fonction d'onde $\Psi(r)$, pour résoudre l'équation de Schrödinger. Contrairement à la fonction d'onde, la densité électronique est une quantité observable (cristallographie) et est définie comme la probabilité de présence des électrons à un endroit de l'espace :

$$\rho(r) = \Psi^*(r)\Psi(r) = |\Psi(r)|^2 . \quad (4.9)$$

La densité électronique a en sus l'avantage de ne dépendre que des coordonnées spatiales (définies dans un espace à trois dimensions), réduisant ainsi le nombre de variables à traiter en comparaison avec la méthode HF ($3N$ dimensions).

4.2.1 Fonctionnelle de Thomas-Fermi

La mise en place du formalisme complet de la DFT ne fut pas établie immédiatement et la première formulation de l'énergie en fonction de la densité électronique a été émise par Thomas et Fermi (TF) en 1927. ^[186,187] Ils proposent une première fonctionnelle de densité sous la forme :

$$E^{TF}[\rho(r)] = T_e^{TF}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] , \quad (4.10)$$

où le potentiel d'interaction noyaux-électrons, $V_{ne}[\rho(r)]$, est défini par :

$$V_{ne}[\rho(r)] = \sum_{A=1}^M \int \frac{Z_A}{|r - r_A|} \rho(r) dr , \quad (4.11)$$

et le potentiel de répulsion coulombien « classique », $V_{ee}[\rho(r)]$, est exprimé comme ;

$$V_{ee}[\rho(r)] = \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 , \quad (4.12)$$

où l'énergie cinétique, $T_e^{TF}[\rho(r)]$, est définie sur la base du modèle d'un gaz uniforme d'électrons (gue) dont l'expression mathématique est :

$$T_{e,\text{gue}}^{TF}[\rho(r)] = \frac{3}{10} (3\pi^2)^{2/3} \int \rho^{5/3}(r) dr . \quad (4.13)$$

Le formalisme de la fonctionnelle TF de départ ne prend pas en compte les phénomènes d'échange et de corrélation électronique. Trois ans plus tard, Dirac reprendra la fonctionnelle TF et y ajoutera

un terme d'échange, K_D , basé lui aussi sur l'approximation que pour obtenir le modèle de Thomas-Fermi-Dirac (TFD) :^[188]

$$E^{TFD}[\rho(r)] = T_e^{TF}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)] + K_D[\rho(r)] , \quad (4.14)$$

$$K_D[\rho(r)] = -\frac{3}{4} \left(\frac{3}{\pi}\right)^{1/3} \int \rho^{4/3}(r) dr . \quad (4.15)$$

Malgré l'apport des premières fonctionnelles, les formalismes restent peu précis car basés sur l'approximation que, notamment pour la description des liaisons chimiques.

4.2.2 Théorèmes de Hohenberg et Kohn

En 1964, Hohenberg et Kohn proposent de généraliser les fonctionnelles proposées dans les années 20 par Thomas, Fermi et Dirac.^[166] Leur objectif est de mettre en place un formalisme exact d'une théorie basée sur la densité électronique. Ils y parviennent à l'aide de deux théorèmes qui vont permettre le développement de la DFT telle que nous l'employons à l'heure actuelle.

Le premier théorème est un théorème d'existence selon lequel une fonctionnelle universelle de la densité, indépendante d'un champ externe (en pratique, les noyaux), a comme valeur minimale l'énergie exacte de l'état fondamental. La démonstration de ce théorème se fait *reductio ad absurdum*. Les auteurs considèrent que les électrons sont dans une boîte, sous l'influence d'un champ extérieur et démontrent que (i) ce potentiel extérieur est, à une constante près, une fonctionnelle de la densité et (ii) qu'il agit sur l'opérateur Hamiltonien du système à l'état fondamental.

Le second théorème montre que l'énergie est une fonctionnelle de densité et que celle-ci respecte le principe variationnel. L'énergie calculée à l'état fondamental est toujours supérieure ou égale à l'énergie exacte. Ce théorème permet d'affirmer que la densité qui minimise la fonctionnelle universelle est la densité exacte de l'état fondamental.

Ces deux théorèmes sont le socle de la DFT actuelle et indiquent que la complexité des problèmes multi-électroniques est associée à la détermination de cette fonctionnelle universelle. Cependant, Hohenberg et Kohn ne proposent pas de forme analytique pour la fonctionnelle universelle.

4.2.3 Équations de Kohn-Sham

Un an après avoir publié les deux théorèmes de base de la DFT avec Hohenberg, Kohn se penche sur la forme analytique de la fonctionnelle de densité universelle en collaboration avec Sham.^[167] L'idée principale des équations de Kohn-Sham (KS) est de simplifier l'opérateur Hamiltonien d'un système multi-électronique en interaction car certains termes n'ont pas de forme analytique connue, par exemple, le terme d'énergie cinétique des électrons (T_e). Ils proposent, alors, de considérer un système où les électrons n'interagissent pas entre eux, mais dans lequel il possèdera la même densité que le système *réel*. Ceci permet de diviser l'expression de l'énergie en

deux composantes : classique (connue) et non-classique (inconnue) :

$$E[\rho(r)] = \underbrace{T_{e,ni}[\rho(r)] + V_{ne}[\rho(r)] + V_{ee}[\rho(r)]}_{\text{classique}} + \underbrace{\Delta T_e[\rho(r)] + \Delta V_{ee}[\rho(r)]}_{\text{non-classique}} . \quad (4.16)$$

Les composantes classiques sont l'énergie cinétique d'un électron n'interagissant pas avec les autres électrons $T_{e,ni}[\rho(r)]$, l'interaction noyaux-électrons $V_{ne}[\rho(r)]$ et la répulsion électron-électron $V_{ee}[\rho(r)]$. Les composantes non-classiques $\Delta T_e[\rho(r)]$ et $\Delta V_{ee}[\rho(r)]$ sont des termes de corrections, respectivement, à l'énergie cinétique d'un électron en interaction et à la répulsion électron-électron (échange et corrélation électroniques). Les termes non-classiques sont regroupés dans le terme E_{xc} , appelé terme d'échange et de corrélation. Ainsi l'équation (4.16) peut être reformulée en introduisant les orbitales de KS, φ_i , sous la forme :

$$E[\rho(r)] = \sum_{i=1}^N \left(\langle \varphi_i | -\frac{1}{2} \nabla_i^2 | \varphi_i \rangle - \langle \varphi_i | \sum_{A=1}^K \frac{Z_A}{|r_i - r_A|} | \varphi_i \rangle \right) , \quad (4.17)$$

$$+ \sum_{i=1}^N \langle \varphi_i | \frac{1}{2} \int \frac{\rho(r')}{|r_i - r'|} dr' | \varphi_i \rangle + E_{xc}[\rho(r)] ,$$

où N et K sont respectivement le nombre d'électrons et de noyaux au sein du système. Dans cette dernière équation, la partie classique peut être résolue de manière exacte alors que ce n'est pas le cas pour E_{xc} puisque sa forme analytique est inconnue. Par contre, E_{xc} peut être évaluée par le biais d'approximations. Si on parvenait à trouver la forme exacte de E_{xc} , nous serions donc capable de résoudre l'équation de Schrödinger mono-électronique de manière exacte :

$$h_i^{\text{KS}} \varphi_i = \epsilon_i \varphi_i , \quad (4.18)$$

$$\left[-\frac{1}{2} \nabla_i^2 + v_{\text{KS}} \right] \varphi_i = \epsilon_i \varphi_i , \quad (4.19)$$

où v_{KS} peut être défini comme un potentiel KS regroupant tous les termes potentiels de l'Hamiltonien KS :

$$v_{\text{KS}} = - \underbrace{\sum_{A=1}^K \frac{Z_A}{|r_i - r_A|}}_{V_{\text{eff}}} + \underbrace{\int \frac{\rho(r')}{|r_i - r'|} dr'}_{V_{\text{H}}} + \underbrace{\frac{\partial E_{xc}[\rho(r)]}{\partial \rho(r)}}_{V_{xc}} , \quad (4.20)$$

où V_{eff} est le potentiel effectif induit par les noyaux, V_{H} , le potentiel d'Hartree représentant la répulsion classique entre les électrons, et V_{xc} est le potentiel d'échange-corrélation regroupant tous les termes non-classiques de l'équation (4.16). L'énergie E_{xc} est définie par :

$$E_{xc}[\rho(r)] = \int \rho(r) \epsilon_{xc}[\rho(r)] dr , \quad (4.21)$$

où $\epsilon_{xc}[\rho(r)]$ est la fonctionnelle d'échange-corrélation électronique qui est fonction de la densité. Un des défis de la DFT réside, encore à l'heure actuelle, dans la conception d'une forme analytique de la fonctionnelle d'échange-corrélation à partir d'approximations permettant d'approcher la fonctionnelle universelle. La conception et la philosophie de certaines fonctionnelles d'échange-corrélation sont explicitées dans le paragraphe suivant.

4.2.4 Fonctionnelle d'échange-corrélation

Nous avons vu que dans les équations de KS, la seule composante non-exacte est la fonctionnelle d'échange-corrélation (*eXchange-Correlation Functional*, XCF). Traditionnellement, bien que cela ne soit pas une obligation, la XCF sera approximée et est divisée en deux composantes : une partie liée à l'échange $\varepsilon_x[\rho(r)]$ et une autre à la corrélation $\varepsilon_c[\rho(r)]$:

$$\varepsilon_{xc}[\rho(r)] = \varepsilon_x[\rho(r)] + \varepsilon_c[\rho(r)] . \quad (4.22)$$

Local Density Approximation

L'approximation la plus simple pour le calcul de la XCF est de considérer que la densité en tout point est constante, comme dans le modèle gue. Cette approximation est appelée *Local Density Approximation* (LDA). Cette approximation locale se justifie seulement dans le cas où la densité varie très peu en fonction des coordonnées de l'espace (ce qui est rarement le cas en réalité). Pour un gue, la formule de l'énergie de l'échange électronique correspond au terme $K_D[\rho(r)]$ introduit par Dirac dans la fonctionnelle TFD (Eq. (4.14) et (4.15)). La formulation de $K_D[\rho(r)]$ ne fait pas la distinction entre les différents spins des électrons et considère la densité des spin α et β identique. Une amélioration possible de la LDA est d'inclure une dépendance de la densité de spin au sein de l'énergie d'échange ; ces XCF appartiennent à la famille LSDA (*Local spin Density Approximation*), dont une forme possible est :

$$E_x^{\text{LSDA}} = -2^{1/3} \frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int (\rho_\alpha^{4/3} + \rho_\beta^{4/3}) dr . \quad (4.23)$$

Concernant la partie liée à la corrélation dans la XCF, la forme analytique dans le cas d'un gue n'est pas connue mais des valeurs précises sont toutefois disponibles par des calculs de type Monte-Carlo quantique. L'extrapolation de ces données peut être intégrée en tant que paramètres dans la fonctionnelle de corrélation, comme c'est le cas pour les fonctionnelles VWN^[189] et PW.^[190] Les fonctionnelles L(S)DA sont souvent employées dans le cadre de systèmes où la densité varie très peu dans l'espace comme dans le cas d'agrégats de métaux simples. L'efficacité de ce type de fonctionnelle réside dans le fait que l'erreur commise sur la partie échange, qui est surestimée, compense par la sous-estimation de la partie corrélation. Cependant, cette approximation ne permet pas de décrire correctement les zones présentant une variation importante de la densité.

Generalized Gradient approximation et *meta*-GGA

Au sein d'un système réel, la densité est inhomogène et l'approximation L(S)DA montre rapidement ses limites. Une amélioration par rapport à la LDA est de prendre en compte la variation de la densité électronique dans l'espace, *via* le calcul du gradient de la densité. Cette approche est l'approximation du gradient généralisé (*Generalized Gradient Approximation*, GGA). Elle fournit des résultats plus réalistes que la LDA car une GGA est une fonctionnelle de la densité électronique

et de son gradient, c'est-à-dire de sa dérivée première :

$$\epsilon_{x/c}^{\text{GGA}} = \epsilon_{x/c}^{\text{LDA}} + \Delta\epsilon_{x/c} \left[\frac{|\nabla\rho(r)|}{\rho^{4/3}(r)} \right]. \quad (4.24)$$

L'incorporation du gradient dans la GGA permet de décrire l'inhomogénéité de $\rho(r)$, amenant à une amélioration par rapport aux LDA. Une fonctionnelle GGA décrit cependant le système de façon mathématiquement locale car elle dépend uniquement de la densité et de son gradient en un point donné, donnant lieu à une description imparfaite des propriétés chimiques. Parmi les GGA les plus utilisées, nous pouvons citer les XCF B88,^[191] PBE,^[192,193] B97^[194] et mPW^[195] pour la partie échange, et LYP,^[196,197] P86^[198] et PW91^[199-201] pour la partie corrélation. Mentionnons qu'il est possible de combiner deux fonctionnelles entre elles comme par exemple la fonctionnelle BLYP, qui présente une forme d'échange B88 et une forme de corrélation LYP. Il existe une amélioration des fonctionnelles GGA dite *meta*-GGA qui consiste à intégrer une contribution supplémentaire, par exemple, la dérivée seconde de l'énergie (Laplacien). Ceci permet une meilleure description de petits systèmes mais l'utilisation du Laplacien de la densité peut amener à des divergences et instabilités numériques. Les fonctionnelles PKZB,^[202] TPSS^[203] et VSXC^[204] font partie de la famille des XCF *meta*-GGA.

Fonctionnelles hybrides et hybrides à longues portées

Une autre famille de XCF, plus raffinées, est l'ensemble des fonctionnelles hybrides. Ce groupe inclut un terme d'échange exact (HF) dans la partie d'échange tandis que la corrélation est souvent incorporée à l'aide d'une fonctionnelle GGA. Par exemple, les XCF hybrides, proposées par Becke, combinent un pourcentage d'échange exact HF avec un pourcentage d'échange GGA :

$$\epsilon_{xc}^{\text{HybrideBecke}} = \alpha\epsilon_x^{\text{HF}} + (1 - \alpha)\epsilon_x^{\text{GGA}} + \epsilon_c^{\text{GGA}} \quad 0 \leq \alpha \leq 1. \quad (4.25)$$

Le paramètre qui définit le mélange d'échange varie fortement d'une fonctionnelle à l'autre, mais on peut considérer qu'une valeur de $\alpha = 0,25$ est souvent pertinente pour les molécules organiques « traditionnelles ». Les fonctionnelles hybrides donnent de très bons résultats pour les énergies et les structures, et elles sont largement employées pour l'étude des propriétés chimiques et optiques de systèmes moléculaires. La fonctionnelle B3LYP est la fonctionnelle la plus utilisée (environ 80% des études DFT) en raison de sa grande efficacité pour des composés organiques « classiques ».^[189,196,205,206] Nous pouvons aussi citer les fonctionnelles BMK,^[207] PBE0^[208] et M06^[209] qui sont elles aussi très utilisées. Durant cette thèse, nous avons utilisé principalement la fonctionnelle M06-2X^[209] qui est une XCF hybride possédant 54% d'échange HF. Ce choix est justifié car cette fonctionnelle donne un très bon compromis entre les structures géométriques et les énergies de transition électroniques obtenus.

Un sous groupe de ces fonctionnelles hybrides sont les hybrides à longue portée (*Range Separated Hybrid*, RSH). Le principe est d'augmenter la fraction d'échange HF avec la distance inter-électronique. CAM-B3LYP^[210] et ω B97-XD^[211] font partie de cette famille. L'idée repose

sur le fait que la corrélation électronique ne concerne que des interactions à courte distance, tandis qu'à longues distances l'échange est le terme prédominant et HF en donne une description exacte. Ces fonctionnelles sont essentiellement utilisées pour l'étude de systèmes moléculaires présentant une délocalisation électronique importante, par exemple, le transfert de charge. Elles simulent correctement l'interaction entre un électron délocalisé spatialement et son trou.

4.3 Simulation des effets de solvants

Traditionnellement, les calculs HF, post-HF et DFT sont réalisés en phase isolée (gaz), ce qui est une situation assez rare expérimentalement. En effet, généralement les propriétés sont étudiées en phase condensée. Une comparaison directe entre expérience et théorie est alors délicate et depuis les années 80, plusieurs groupes se sont attelés à décrire le solvant au sein des calculs en mécanique quantique. Il existe plusieurs façons de traiter les effets du solvant sur les propriétés d'un soluté.

4.3.1 Solvation explicite

L'idée la plus naturelle pour simuler un système en solution est d'ajouter, explicitement, des molécules de solvant autour du soluté étudié créant ainsi un réseau de liaisons dans le cas de solvant protique et d'interactions spécifiques pouvant influencer les propriétés étudiées. Parmi les interactions spécifiques entre le soluté et le solvant, on distingue entre-autres les interactions de van der Waals (vdW), les liaisons hydrogène, les interactions hydrophobes et les transferts de protons et/ou de charges. L'attractivité de ce type d'approche est fortement limitée par le temps de calculs requis lorsque le système contient un très grand nombre d'atomes. La simulation de la totalité du système, lors d'une approche quantique, amène à un problème difficilement traitable car la présence des molécules de solvant augmente la taille du système étudié. Le nombre de molécules de solvant doit être choisi méticuleusement ainsi que leurs positionnements autour du soluté. La solvation explicite n'est donc pas facile à mettre en place lors de calculs en mécanique quantique.

4.3.2 Solvation implicite

L'autre approche permettant d'intégrer le solvant au sein des calculs est la solvation implicite. Les molécules de solvants sont remplacées par un continuum polarisable et uniforme, possédant les mêmes propriétés physico-chimiques macroscopiques (constante diélectrique, indice de réfraction, coefficient d'expansion molaire, etc.) que le solvant *réel*. Le continuum est placé autour d'un soluté placé dans une cavité. Les interactions « soluté-solvant » sont déterminées de manière auto-cohérente, *via* la résolution d'un terme électrostatique. Cependant, les interactions spécifiques tel que les liaisons hydrogènes sont négligées dans ce modèle. L'approche implicite est couramment utilisée pour traiter des systèmes en solution car le temps de calcul est très inférieur à celui nécessaire lors d'une solvation explicite. De plus, elle permet d'obtenir, généralement, un bon accord

avec l'expérience. L'énergie de solvation (ΔG_{solv}) dans un modèle basé sur un continuum polarisable est exprimée comme :

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{disp/rep}} + \Delta G_{\text{elec}} . \quad (4.26)$$

Le premier terme, ΔG_{cav} , correspond à l'énergie de création de la cavité au sein du continuum, déstabilisant ainsi le système. Le second terme, $\Delta G_{\text{disp/rep}}$, correspond aux énergies de dispersion et de répulsion entre le soluté et le solvant. Les deux termes s'expriment par la relation :

$$\Delta G_{\text{cav}} + \Delta G_{\text{disp/rep}} = \sum_{k=1}^N \sigma_k SA_k , \quad (4.27)$$

où SA_k est la surface accessible au solvant des atomes de type k et σ_k est un terme empirique déterminé par des paramètres de solvation atomique expérimentale. La somme de tous les termes SA_k permet de définir la surface totale (du soluté) accessible au solvant (*Solvent Accessible Surface*, SAS). Elle est obtenue à l'aide d'une sphère génératrice qui roule sur la surface du soluté afin de créer une cavité (Figure 10). Cette cavité est ici définie par le rayon de VdW de chaque atome de soluté, et plus précisément par les valeurs de Bondi.^[212] En plus de définir la SAS, la sphère génératrice permet de décrire la surface d'exclusion du solvant (*Solvent Exclusion Surface*, SES).

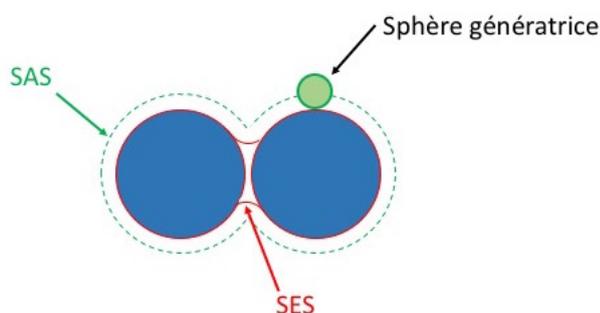


FIGURE 10 – Représentation de la sphère génératrice déterminant la cavité de solvation ainsi que des surfaces SAS et SES.

Le dernier terme, ΔG_{elec} , décrit les interactions électrostatiques entre la surface de la cavité et le continuum. Des charges ponctuelles sont placées à la surface de la cavité et polarisent le soluté. La modification de ce dernier va redistribuer, à son tour, les charges présentes à la surface de la cavité, et ainsi de suite jusqu'à l'obtention d'un équilibre entre les deux éléments. Ce processus auto-cohérent est solutionné *via* à l'équation de Poisson, une équation différentielle du second ordre qui relie le potentiel électrostatique $V(r)$, la distribution des charges à la surface de la cavité ρ_m et la constante diélectrique du milieu $\epsilon_r(r)$:

$$-\nabla[\epsilon_r(r)\nabla V(r)] = 4\pi\rho_m(r) . \quad (4.28)$$

À l'intérieur de la cavité, $\epsilon_r(r) = 1$ tandis qu'à l'extérieur, ϵ_r prend la valeur du milieu. Ces deux conditions mènent à :

$$\begin{cases} -\nabla^2 V(\mathbf{r}) = & 4\pi\rho_m(\mathbf{r}) , \mathbf{r} \in \text{cavité} , \\ -\epsilon\nabla^2 V(\mathbf{r}) = & 0 , \mathbf{r} \in \text{continuum} . \end{cases} \quad (4.29)$$

En dehors de la cavité, $\rho_m(r)$ est nul, expliquant ainsi la seconde égalité ci-dessus. La valeur absolue de la différence de potentiel entre l'intérieur et l'extérieur de la cavité doit être nulle afin de respecter le principe de continuité du potentiel électrostatique. Le terme $V(r)$ est composé de deux termes : le premier correspondant au potentiel de la distribution des charges du soluté à la surface de la cavité et le second au potentiel induit par la polarisation du continuum.

La méthode la plus employée pour la résolution de ce problème électrostatique consiste à créer des quantités de charges apparentes à la surface de la cavité. Cette approche, appelée ASC (*Apparent Surface Charge*), est intégrée dans le formalisme du modèle PCM (*Polarisable Continuum Model*), la méthode de solvation implicite la plus utilisée.^[213] Dans notre travail, nous avons également utilisé le modèle de solvation implicite SMD (*Solvation Model based on solute electron Density*).^[214] Ce modèle, développé par le groupe de Truhlar, est basé sur le même formalisme que le modèle PCM mais utilise un jeu de rayons atomiques paramétrés à partir d'énergies de solvation expérimentales pour créer la cavité et intègre des termes non-électrostatiques différents.

Rappelons que la simulation des effets de solvant par une approche implicite ne permet cependant pas d'inclure des interactions spécifiques entre le soluté et les molécules de solvant. En effet les liaisons hydrogènes, les transferts de protons, les interactions π -stacking, etc., ne sont pas décrites explicitement. Le modèle implicite ne convient donc pas totalement lorsque le solvant forme des interactions spécifiques avec le soluté. Il existe cependant une approche alternative capable d'inclure ces interactions spécifiques explicitement en complément d'une solvation implicite du système avec un coût calculatoire raisonnable comme montré ci-dessous.

4.3.3 Solvation implicite/explicite (IE)

L'approche de solvation implicite/explicite (IE) consiste à ajouter, explicitement, quelques molécules de solvant autour du soluté afin de prendre en compte les interactions soluté-solvant spécifiques, nécessaires lors de certaines études, l'ensemble supramoléculaire étant placé dans une cavité comme dans l'approche implicite. Dans le cas d'une solvation explicite, le nombre et le placement des molécules de solvant explicites autour du soluté sont des données très importantes et sont peu triviales ; cette problématique persiste dans l'approche IE. Des contre-ions peuvent également être ajoutés dans le cas d'espèces chargées en solution. L'approche IE est un bon compromis mais reste plus complexe et coûteuse en temps de calculs que les méthodes implicites. Dans cette thèse, nous avons eu recours à cette approche lors de la simulation des états excités en solution des molécules d'intérêt.

4.4 DFT dépendante du temps

Cette section a pour objectif d'énoncer les principes de la DFT dépendante du temps, qui est la méthode que nous avons utilisée pour l'étude des états excités.

4.4.1 Théorèmes de Runge-Gross et van Leeuwen

La TD-DFT (*Time Dependent-DFT*, TD-DFT) est basée sur deux théorèmes qui permettent de transposer la théorie de la fonctionnelle de densité à une situation non-statique. Le premier théorème, énoncé en 1984 par Runge et Gross,^[168] montre qu'un potentiel externe dépendant du temps $v(r, t)$ produit une fonction d'onde dépendante du temps $\Psi(t)$ à partir d'une fonction d'onde d'un état initial Ψ_0 , générant ensuite une densité dépendante du temps $\rho(r, t)$:

$$v(r, t) \xrightarrow[\Psi_0]{i\partial\Psi(t)/\partial t = \hat{H}(t)\Psi(t)} \Psi(t) \xrightarrow{\langle\Psi(t)|\hat{\rho}|\Psi(t)\rangle} \rho(r, t) . \quad (4.30)$$

À partir de ce constat, le théorème de Runge-Gross démontre qu'en partant du même état initial, Ψ_0 , il existe deux potentiels externes différents, $v(r, t)$ et $v'(r, t)$, donnant lieu, respectivement, à deux densités différentes, $\rho(r, t)$ et $\rho'(r, t)$. L'unique condition pour obtenir deux densités différentes est que les deux potentiels doivent différer à plus d'une constante dépendante du temps près :

$$v(r, t) \neq v'(r, t) + c(t) . \quad (4.31)$$

Le deuxième théorème est une reformulation du premier, proposée par van Leeuwen en 1999.^[215] Il spécifie que pour une densité $\rho(r, t)$ associée à ses interactions données (ω), son champ externe $v(r, t)$ et son état initial (Ψ_0), il existe un autre système avec ses interactions données ω' et un autre potentiel externe ($v'(r, t)$) qui reproduit la même densité $\rho(r, t)$. L'état initial Ψ'_0 de ce système doit cependant être choisi correctement. Ainsi, si $\Psi_0 = \Psi'_0$, le théorème de van Leeuwen devient similaire à celui de Runge-Gross. A l'inverse, si on choisit $\omega' = 0$, alors le deuxième système est sans interaction. Il existe alors un état initial sans interaction Φ_0 soumis à un potentiel externe $v_s(r, t)$ qui permet de reproduire la densité $\rho(r, t)$, ce qui justifie l'approche KS dépendante du temps reposant sur un système sans interaction.

4.4.2 Equation KS dépendante du temps et réponse linéaire

Les équations KS dépendantes du temps peuvent être formulées comme :

$$\left[-\frac{1}{2}\nabla_i^2 + v_{\text{KS}}[\rho](r, t) \right] \varphi_i = i \frac{\partial \varphi_i(r, t)}{\partial t} , \quad (4.32)$$

où la densité électronique du système est définie par la relation :

$$\rho(r, t) = \sum_{i=1}^N |\varphi_i(r, t)|^2 . \quad (4.33)$$

Le système a subi une perturbation par l'intermédiaire d'un champ extérieur dépendant du temps, ce dernier pouvant être assimilé à un champ électrique. Le champ induit une réponse de la densité électronique qui peut être développée en série de Taylor ;

$$\rho(r, t) = \rho_0(r, t) + \rho_1(r, t) + \rho_2(r, t) + \rho_3(r, t) + \dots , \quad (4.34)$$

où $\rho_0(r, t)$ est la densité électronique du système non-perturbé, c'est-à-dire à la densité de l'état fondamental et où $\rho_1(r, t)$, $\rho_2(r, t)$, $\rho_3(r, t)$, etc., sont successivement la réponse linéaire, quadratique, cubique, etc., de la densité électronique à la perturbation appliquée. Ici, seule la réponse linéaire est prise en compte donnant lieu à une nouvelle densité électronique $\rho(r, t)$ composée de la densité de l'état fondamentale, $\rho_{\text{GS}}(r, t)$, et de sa réponse linéaire à la perturbation, $\rho_1(r, t)$:

$$\rho(r, t) = \rho_{\text{GS}}(r, t) + \rho_1(r, t), \quad (4.35)$$

où $\rho_1(r, t)$ peut être exprimée comme :

$$\rho_1(r, t) = \int_{-\infty}^{+\infty} dt' \int d^3r' \chi(r, r', t - t') v_1^{\text{ext}}(r', t') d^3r', \quad (4.36)$$

où v_1^{ext} est le potentiel de la perturbation externe et $\chi(r, r', t - t')$, la réponse au premier ordre de la densité au point r et au temps t de la variation du potentiel externe au point r' et au temps t' . Notons aussi la condition $t < t'$, $\chi(r, r', t - t') = 0$, qui se justifie car la densité n'est pas influencée par une variation du potentiel ultérieure au temps t . L'expression de $\chi(r, r', t - t')$ dans la limite d'un champ extérieur statique v_0^{ext} est :

$$\chi(r, r', t - t') = \left. \frac{\delta\rho(r, t)}{\delta v^{\text{ext}}(r', t')} \right|_{v^{\text{ext}}=v_0^{\text{ext}}}. \quad (4.37)$$

Il est possible de dériver en chaine la dernière équation en introduisant le terme du potentiel effectif du système sans interactions, v_{KS} possédant la même densité électronique du système avec interactions, le développement de l'équation (4.37) donnant ainsi :

$$\chi(r, r', t - t') = \int_{-\infty}^{+\infty} dt'' \int d^3r'' \frac{\delta\rho(r, t)}{\delta v_{\text{KS}}(r'', t'')} \frac{\delta v_{\text{KS}}(r'', t'')}{\delta v^{\text{ext}}(r', t')} d^3r''. \quad (4.38)$$

Le second terme de la double intégrale est la variation du potentiel effectif du système KS sous la perturbation du champ externe qui peut être développée sous la forme :

$$\frac{\delta v_{\text{KS}}(r'', t'')}{\delta v^{\text{ext}}(r', t')} = \frac{\delta V_{\text{eff}}(r'', t'')}{\delta v^{\text{ext}}(r', t')} + \frac{\delta V_{\text{H}}(r'', t'')}{\delta v^{\text{ext}}(r', t')} + \frac{\delta V_{\text{xc}}(r'', t'')}{\delta v^{\text{ext}}(r', t')}. \quad (4.39)$$

Le premier terme de l'équation (4.38) est la réponse linéaire de la densité électronique d'un système KS, similaire à l'équation (4.37), sans interactions :

$$\chi_{\text{KS}}(r, r'', t - t'') = \left. \frac{\delta\rho(r, t)}{\delta v_{\text{KS}}(r'', t'')} \right|_{v_{\text{KS}}=v_{\text{KS}}^{\rho_0}}. \quad (4.40)$$

Cette réponse de la densité s'exprime, dans le domaine des fréquences, comme :

$$\chi_{\text{KS}}(r, r'', \omega) = 2 \lim_{\eta \rightarrow 0^+} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \left\{ \frac{\varphi_j^*(r) \varphi_k(r) \varphi_j(r'') \varphi_k^*(r'')}{\omega - \omega_q + i\eta} - \frac{\varphi_j(r) \varphi_k^*(r) \varphi_j^*(r'') \varphi_k(r'')}{\omega + \omega_q - i\eta} \right\}, \quad (4.41)$$

où les indices j et k représentent respectivement les orbitales KS occupées et inoccupées ($j \neq k$), ω_q est la fréquence correspondante à la transition d'une orbitale KS occupée φ_j vers une autre orbitale KS inoccupée φ_k , c'est-à-dire la différence des énergies propres de ces deux orbitales KS :

$$\omega_q = \varepsilon_k - \varepsilon_j. \quad (4.42)$$

Lorsque la fréquence du potentiel extérieur ω tend vers ω_q , χ_{KS} tend vers l'infini, révélant une résonance de pôle. Ceci indique que la fréquence du champ extérieur correspond à une fréquence de transition électronique possible dans le composé. La valeur des pôles peut être reliée aux intensités obtenue lors d'une absorption optique, c'est-à-dire aux forces d'oscillateur. Casida a mis en place un formalisme basé sur une matrice de simple excitation des pôles afin d'obtenir les valeurs des forces d'oscillateur des transitions électroniques étudiées.^[169]

4.4.3 Propriétés optiques en solution

Au cours de notre travail de thèse, nous avons utilisé trois méthodes pour décrire les propriétés optiques des états excités de composés en solution. Elles sont toutes fondées sur le formalisme PCM, représentant la polarisation du solvant par des charges induites à la surface d'une cavité, entourant le soluté d'intérêt. Néanmoins, ces méthodes diffèrent par leurs descriptions de la polarisation de la cavité à l'état excité. La première est l'approche LR (*Linear Response*),^[216] qui se base sur le moment de transition de densité électronique entre l'état fondamental et l'état excité pour calculer la polarisation de la cavité. La seconde, appelée cLR (*corrected Linear Response*),^[217] est une correction appliquée à l'énergie totale obtenue par le formalisme LR. L'approche cLR prend en compte, au premier ordre, la relaxation de la densité à l'état excité et va ainsi modifier les charges présentes sur la cavité, grâce à un calcul de la matrice densité de cet état. La dernière méthode, SS (*State Specific*),^[218,219] consiste à équilibrer les charges induites avec la densité de charge de la molécule à l'état excité. Ainsi, la polarisation de la cavité est obtenue de façon itérative par un processus auto-cohérent dans l'approche SS. La dernière méthode est plus coûteuse en temps de calcul et montre certaines limites quant à la description de systèmes à fort transfert de charges.

4.4.4 Limites physiques

Lorsqu'un système est sous l'effet d'un champ électrique, la polarisation de ce dernier dépend essentiellement de la fréquence du champ appliqué. Plusieurs phénomènes sont observés lors d'un changement de fréquence. Par exemple, à basse fréquence (micro-ondes), la polarisation des molécules va induire des mouvements de rotations des molécules, et dans le domaine de l'infrarouge, des distorsions moléculaires (vibrations) vont apparaître. À plus haute fréquence, seul le nuage électronique des molécules se déforme sous l'influence du champ électrique, comme c'est le cas dans le domaine ultra-violet. Les trois phénomènes sont illustrés à la Figure 11.

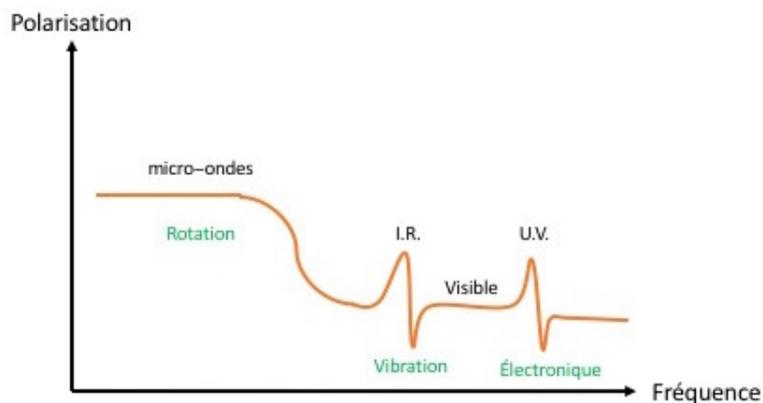


FIGURE 11 – Représentation qualitative de l'évolution de la polarisation en fonction de la fréquence d'un champ électrique appliqué.

Lors d'un phénomène d'absorption dans le domaine UV-Visible, deux limites physiques peuvent être considérées. La première est appelée limite de non-équilibre où seul le nuage électronique du soluté et du solvant répond au champ électrique appliqué. Cette limite s'applique pour les excitations électroniques dites « verticales » où la géométrie de l'état excité est la même qu'à l'état fondamental, respectant ainsi parfaitement le principe de Franck-Condon (FC). La seconde est la limite d'équilibre, où les noyaux et le nuage électronique réagissent au champ électrique. Cette limite s'applique si les molécules de solvant ont le temps de s'adapter à l'état excité du soluté, c'est-à-dire lorsque des phénomènes lents sont étudiés comme, par exemple, la phosphorescence. Les deux limites sont intégrées dans les formalismes de solvation implicite, ce qui permet de traiter la polarisation de la cavité suivant le phénomène étudié.

4.4.5 Contributions vibrationnelles des spectres optiques

Expérimentalement, la forme du spectre d'absorption peut présenter une forme s'écartant de celle d'une gaussienne non-résolue et présenter, par exemple, un second maximum ou un épaulement. Ceci est caractéristique d'un couplage vibronique significatif. Lors d'une excitation, qui est un phénomène rapide, seuls les électrons peuvent se réorganiser alors que les noyaux sont fixes (principe de Franck-Condon). Les transitions sont dites « verticales » car la géométrie initiale du système ne varie pas au cours du processus d'absorption.

L'intensité des transitions vibroniques est directement liée au recouvrement des fonctions d'onde des états vibrationnels de chaque état électronique impliqué. Ce recouvrement entre les fonctions d'onde vibrationnelles est aussi appelé le facteur de Franck-Condon (facteur de FC). Pour illustrer ceci, la Figure 12 représente les courbes d'énergie potentielle de deux états électroniques, E_0 et E_1 , ainsi que leurs fonctions d'ondes vibrationnelles. La transition vibronique $v'' = 0 \rightarrow v' = 0$ a une intensité très faible, voir nulle, car un recouvrement très faible entre les deux fonctions d'ondes vibrationnelles est observé. À l'inverse, la transition $v'' = 0 \rightarrow v' = 2$ possède un important facteur de FC et contribuera donc fortement à l'absorption.

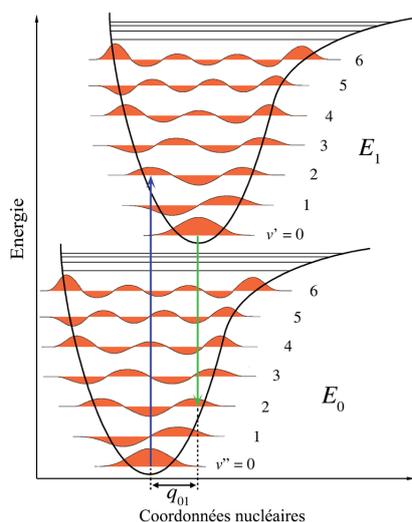


FIGURE 12 – Diagramme énergétique des niveaux vibrationnels des états électroniques E_0 et E_1 . La fonction d'onde vibrationnelle de chaque niveau est représentée en orange. (source : Wikipédia).

4.5 Bases de fonctions atomiques

Pour pouvoir résoudre les équations HF ou KS, il est important de disposer d'une base de fonctions mathématiques, dites atomiques, qui puissent être combinées pour décrire le comportement des électrons. Théoriquement, il faudrait une combinaison d'une infinité de fonctions mathématiques pour décrire parfaitement la structure électronique et obtenir l'énergie exacte d'un système moléculaire. En réalité, les fonctions atomiques d'une molécule sont décrites le plus souvent avec un nombre fini de fonctions de base. C'est au chimiste de faire le choix du nombre de fonctions atomiques au départ du calcul, les orbitales moléculaires étant une Combinaison Linéaire d'Orbitales Atomiques (CLOA), ce qui permet de réduire la résolution des équations HF ou KS à une « optimisation » de coefficients numériques, c'est-à-dire les coefficients CLOA :

$$\varphi_i = \sum_i^n (c_i \chi_i), \quad (4.43)$$

où φ_i est l'orbitale moléculaire, χ_i , le jeu d'orbitales atomiques impliquées dans l'orbitale moléculaire et c_i , leur coefficient CLOA. Les premières bases atomiques proposées étaient basées sur les orbitales dites de Slater (*Slater Type Orbital*, STO) :^[220]

$$\phi_{\zeta,n,l,m}(r, \theta, \varphi) = N Y_{l,m}(\theta, \varphi) r^{n-1} e^{-\zeta r}, \quad (4.44)$$

où N est une constante de normalisation, $Y_{l,m}$, une fonction harmonique sphérique, n , l et m sont, respectivement, les nombres quantiques principal, secondaire et tertiaire, et ζ , une constante liée à la charge effective du noyau. Bien que la dépendance exponentielle de ces orbitales soit physiquement correcte, l'utilisation de ce type de fonctions atomiques pose néanmoins un problème car la résolution analytique d'un système à trois centres ou plus est impossible. L'utilisation de STO est donc principalement réservée aux systèmes diatomiques. De nos jours, on remplace les STO

par des fonctions gaussiennes (*Gaussian Type Orbital*, GTO),^[221] qui sont plus faciles à manipuler mais ne présentent cependant plus une description correcte de la décroissance exponentielle en fonction de la distance :

$$\phi_{\zeta,n,l,m}(r, \theta, \varphi) = NY_{l,m}(\theta, \varphi)r^{2n-2-l}e^{-\zeta r^2}. \quad (4.45)$$

L'inconvénient des GTO est leur description moins précise des phénomènes tant très proches que très éloignés du noyau. L'effet dit de « pointe », très bien décrit par les STO, n'est pas reproduit tandis que les GTO décroissent trop rapidement loin du noyau. Ces deux effets sont la conséquence directe de la dépendance en e^{-r^2} des GTO. Cependant, la combinaison de plusieurs GTO permet de se rapprocher du comportement obtenu avec une STO. Ces combinaisons ont l'avantage que le produit de deux fonctions gaussiennes donne une fonction gaussienne. On appelle CGTO (*Constrained GTO*) la combinaison de plusieurs fonctions gaussiennes dans le but de mieux décrire l'effet de pointe et le comportement à grande distance du noyau. Les CGTO sont caractérisées par une longueur de contraction, qui correspond au nombre de GTO les composant.

Les bases minimales ou simple- ζ n'utilisent qu'une seule GTO pour chaque orbitale occupée dans l'atome isolé. Pour améliorer la description des propriétés des molécules, il existe des bases double- ζ ou triple- ζ qui combinent deux ou trois CGTO par orbitale occupée dans l'atome libre. Une amélioration conceptuelle supplémentaire a été apportée avec le développement des bases *split-valence*. Le principe est fondé sur la différence entre d'une part, les électrons de cœur, qui ne participent pas aux liaisons chimiques et qui sont proches du noyau (fort effet de pointe), et, d'autre part, les électrons de valence qui forment les liaisons et qui sont relativement éloignés des noyaux. Les bases *split-valence* traitent ainsi séparément les électrons de cœur, avec une seule CGTO de longueur de contraction assez grande, et les électrons de valence avec deux ou trois CGTO de plus petites longueurs de contraction. Parmi la multitude de bases de fonctions atomiques existantes,^[222-227] les bases *split-valence* de Pople,^[228-231] sont les plus populaires et sont celles utilisées dans notre travail. Leur nomenclature est la suivante : $c-v_1v_2v_3G$ avec c le nombre de fonctions gaussiennes primitives utilisées par la CGTO représentant les électrons de cœur, et v_1 , v_2 et v_3 sont les longueurs de contractions des CGTO pour décrire les orbitales de valence. Par exemple, la base 6-311G^[231] contient une CGTO combinant 6 gaussiennes pour décrire les orbitales de cœur et 3 CGTO de longueurs de contractions respectives 3, 1 et 1 pour les orbitales de valence. La base 6-311G est donc une base triple- ζ pour les électrons de valence. Pour augmenter la flexibilité de la base, on peut encore y ajouter des fonctions de polarisation^[232] et/ou de diffusion.^[233] Les fonctions de polarisation sont des orbitales dont le moment angulaire est (au minimum) d'une unité supérieure à celui des orbitales occupées dans l'atome libre, par exemple les orbitales p pour l'hydrogène et les orbitales d pour le carbone. Ces fonctions permettent de mieux décrire les liaisons chimiques et la géométrie du système ainsi que la plupart des propriétés. Leur notation se fait entre parenthèses, la première lettre indique quelles fonctions de polarisation sont ajoutées aux atomes « lourds » et, celle après la virgule, concerne les atomes d'hydrogène. Les fonctions diffuses permettent de mieux décrire la délocalisation d'un électron dans le système ;

elles sont donc constituées de gaussiennes très plates (ζ petit dans $e^{-\zeta r^2}$). Ces fonctions ont le même moment angulaire que les orbitales de valence. La présence d'un sigle « + » dans le nom de la base signifie qu'une fonction diffuse est appliquée sur tous les atomes « lourds » et un deuxième « + » signifie que des fonctions de diffusion sont ajoutées aux atomes d'hydrogène. Les fonctions diffuses jouent un rôle très important, voire crucial, dans l'étude de certaines propriétés telles que l'affinité électronique, l'acidité, le transfert de charge, les transitions électroniques, etc.

5 Études des dérivés stilbènes disulfoniques

Après avoir présenté dans le précédent chapitre les outils de chimie quantique que nous avons utilisés dans cette thèse, ce chapitre présente les résultats expérimentaux et théoriques de l'étude que nous avons menée sur une série de dérivés stilbènes disulfoniques (SD) en solution (Figure 13). L'objectif de notre travail est de caractériser les structures géométriques, états de protonations en condition physiologique (pH = 7,4), propriétés optiques et cinétiques des composés en solution.^[79] Ces informations sont cruciales pour la suite de notre thèse et, notamment, pour l'étude des interactions entre les dérivés SD et la protéine ASH, présentés dans le Chapitre 7.

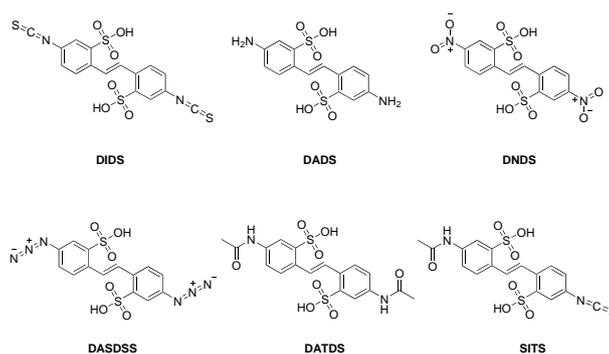


FIGURE 13 – Structures de la série des stilbènes disulfoniques étudiés.

5.1 États de protonation des SD

À notre connaissance et avant ce travail, aucun article présent dans la littérature ne traite de l'état de protonation des groupements sulfoniques présents sur les dérivés SD. En effet, si plusieurs travaux issus de la biologie présentent les SD sous une forme protonée ($-\text{SO}_3\text{H}$),^[234–236] d'autres études les présentent sous la forme déprotonée ($-\text{SO}_3^-$), ce qui, au vu des pK_a typiques des acides sulfoniques, nous semble plus cohérent.^[125,179,237] Il est dangereux de considérer les SD sous une forme incorrecte pour étudier ses propriétés physico-chimiques en solution ou en milieu biologique car ceci peut amener à des conclusions biaisées.

Le processus de déprotonation des composés SD est présenté par l'équation :



où m et n sont, respectivement, la charge électronique totale de la molécule et le nombre de protons présent sur les groupements soufrés. À travers le processus de déprotonation, on peut identifier trois états qui pourraient exister en solution. Le premier état correspond à la forme totalement protonée et neutre de la molécule, $SD(H_2)$ ($n=2$ et $m=0$). Le second état, nommé $SD(H_1)^-$, avec $n=1$ et $m=1$, est celui obtenu après la première déprotonation où une seule fonction sulfonique est présente, la seconde étant sous la forme sulfonate ($-SO_3^-$). Le troisième état est la molécule issue d'une double déprotonation des fonctions sulfoniques en fonctions sulfonates, SD^{2-} ($n=0$ et $m=2$).

A priori, les deux fonctions sulfoniques possèdent des valeurs de pK_a proches. Suite à la première déprotonation, le pK_a de la seconde fonction sulfonique se voit cependant modifié. En effet, la symétrie de la molécule $SD(H_1)^-$ est brisée et un terme d'entropie rotationnelle disparaît, modifiant ainsi la valeur du second pK_a .

Les constantes d'acidité liées aux deux processus de déprotonation, pK_a^1 et pK_a^2 , sont définis comme le logarithme négatif des constantes d'équilibre des deux processus, K_a^1 et K_a^2 :

$$pK_a = -\log(K_a). \quad (5.2)$$

Le pK_a est une propriété reliée à l'énergie libre de solvation (ΔG_{solv}) du système par la relation :

$$pK_a = \frac{\Delta G_{solv}}{2.303RT}, \quad (5.3)$$

où T est la température (en kelvin) et R la constante des gaz parfait ($8,314 \text{ J.mol}^{-1}\text{K}^{-1}$). ΔG_{solv} peut être évaluée à l'aide d'un cycle thermodynamique de type Börn-Haber (Figure 14).^[238,239] ΔG_{solv} est obtenue en calculant l'énergie libre du processus de déprotonation en phase gazeuse, ΔG_{gas} , ainsi que les énergies de solvation de chaque espèce, comme présenté dans les équations suivantes :

$$\Delta G_{gas} = G_{gas}^{SD(H_{n-1})^{(m+1)-}} + G_{gas}^{H^+} - G_{gas}^{SD(H_n)^{m-}}, \quad (5.4)$$

$$\Delta G_{solv} = \Delta G_{aq}^{SD(H_{n-1})^{(m+1)-}} + \Delta G_{aq}^{H^+} - \Delta G_{aq}^{SD(H_n)^{m-}} + \Delta G^{1atm \rightarrow 1M}. \quad (5.5)$$

Les valeurs de $G_{gas}^{H^+}$ et de $\Delta G_{aq}^{H^+}$ sont déterminées expérimentalement, respectivement, $-6,28$ et $-263,98 \text{ kcal.mol}^{-1}$.^[240,241] Dans l'équation (5.5), un terme supplémentaire fait son apparition, $\Delta G^{1atm \rightarrow 1M}$, correspondant à une correction lors du passage d'un milieu gazeux à une pression de 1 atm vers un milieu liquide de concentration 1 M. Cette correction s'élève à $1,89 \text{ kcal.mol}^{-1}$.

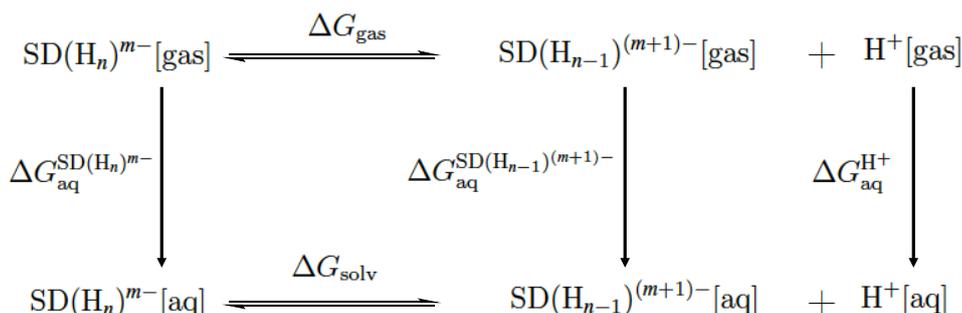


FIGURE 14 – Cycle thermodynamique de *Börn-Haber* adapté au système étudié

La première étape pour établir le cycle de Börn-Haber est l'optimisation géométrique des espèces $SD(H_n)^{m-}$ et $SD(H_{n-1})^{(m+1)-}$ en phase gazeuse. L'optimisation des géométries est effectuée au niveau de théorie M06-2X/6-31G(d). Nous nous assurons d'avoir une structure optimisée correspondant à un minimum local par un calcul des fréquences, au même niveau de théorie. De plus, le calcul de fréquence nous permet d'obtenir les termes $G_{\text{gas}}^{\text{SD}(H_n)^{m-}}$ et $G_{\text{gas}}^{\text{SD}(H_{n-1})^{(m+1)-}}$, ce qui permet le calcul de ΔG_{gas} .

La seconde étape est l'évaluation des termes $\Delta G_{\text{aq}}^{\text{SD}(H_n)^{m-}}$ et $\Delta G_{\text{aq}}^{\text{SD}(H_{n-1})^{(m+1)-}}$. Pour cela, nous avons suivi une méthode qui consiste à effectuer des calculs « simple point » sur la géométrie en phase gazeuse en présence d'un solvant implicite, suivant le modèle de solvation SMD. En effet, l'association de SMD avec les fonctionnelles Minnesota permet de reproduire l'énergie libre de solvation.^[214,242,243] 1

Les résultats théoriques des pK_a sont très dépendants de l'erreur sur le ΔG_{solv} et donc du niveau de théorie utilisé (la base de fonctions atomiques et la fonctionnelle). Une variation de 1,36 kcal.mol⁻¹ au niveau du ΔG_{solv} correspond à un changement d'une unité de pK_a . Dès lors, nous avons également déterminé le pK_a avec une base de fonctions atomiques plus étendue, 6-31+G(2d,p), en vue d'estimer la cohérence de nos résultats (Tableau 1). Quelque soit la base uti-

Tableau 1 – pK_a calculés en solution aqueuse pour les dérivés SD, au niveau de théorie M06-2X/6-31G(d) et M06-2X/6-31+G(2d,p).

Base	Molécule	Isomère	pK_a^1	pK_a^2	Base	Molécule	Isomère	pK_a^1	pK_a^2
6-31G(d)	DIDS	<i>cis</i>	-3.4	1.3	6-31+G(2d,p)	DIDS	<i>cis</i>	-4.5	0
		<i>trans</i>	-4.6	1.6			<i>trans</i>	-5.8	0
	DADS	<i>cis</i>	-0.8	3.2		DADS	<i>cis</i>	-2.7	1.5
		<i>trans</i>	-1.9	4.1			<i>trans</i>	-3.7	1.9
	DASDSS	<i>cis</i>	-2.5	1.9		DASDSS	<i>cis</i>	*	*
		<i>trans</i>	-3.5	2.5			<i>trans</i>	-5.1	0.7
	DNDS	<i>cis</i>	-4.5	0.2		DNDS	<i>cis</i>	-5.9	-1.2
		<i>trans</i>	-5.5	0.6			<i>trans</i>	-7.2	-0.9
	DATDS	<i>cis</i>	-1.6	6.7		DATDS	<i>cis</i>	-3.5	4.5
		<i>trans</i>	-3.3	7.4			<i>trans</i>	-5.2	4.7
	SITS	<i>cis</i>	-0.3	1.7		SITS	<i>cis</i>	-2.5	0,1
		<i>trans</i>	-4.0	4.9			<i>trans</i>	-5.6	2.7

*Optimisations non convergées

lisée, nous constatons que toutes les valeurs de pK_a^1 sont négatives, ce qui est caractéristique des acides forts, confirmant donc la grande acidité des fonctions sulfoniques. Cette première analyse

1. D'après les travaux de Cramer et Truhlar, l'erreur sur le ΔG_{solv} est en moyenne de 0,6 kcal.mol⁻¹ pour les espèces neutres, 3,1 kcal.mol⁻¹ pour les cations et 5,0 kcal.mol⁻¹ pour les anions.

permet d'affirmer qu'en milieu physiologique ($\text{pH} \approx 7,4$), les SD ne se trouvent pas sous leur forme diprotonée. Simultanément, nous avons effectué des mesures expérimentales en collaboration avec le Dr. Jérôme Graton et Anaïs Goupille de l'équipe ModES. Aucun changement n'a été observé en milieu acide ou neutre dans le cas du DIDS (Figure 15), confirmant la présence d'une seule espèce en solution aqueuse dans ces conditions. Cependant, un décalage, qui peut paraître surprenant, est observé pour la solution basique. Cet effet n'est très probablement pas lié aux fonctions sulfoniques. Il est connu que l'attaque nucléophile d'un ion hydroxyl sur une fonction isothiocyanate mène à un dérivé monothiocarbamate, NH-C(=S)-O^- , qui ensuite s'hydrolyse en fonction NH_2 par l'intermédiaire d'une molécule d'eau.^[244] Cette réactivité de la fonction isothiocyanate en milieu basique explique donc le déplacement observé pour la bande d'absorption.

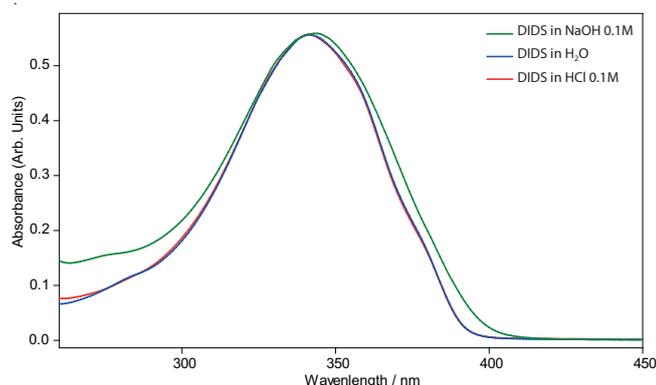


FIGURE 15 – Spectres d'absorption expérimentaux du DIDS en solution aqueuse obtenus dans différents milieux.

Dans le Tableau 1, nous observons que les valeurs calculées pour les $\text{p}K_a^1$ et $\text{p}K_a^2$ sont très différentes. Concernant les isomères *trans*, nous remarquons des variations d'au moins 6 unités de $\text{p}K_a$ entre la première et la seconde déprotonation. Cette différence a pour principale raison le réarrangement géométrique opéré lors du second processus de déprotonation. En effet, dans le cas des états de protonation $\text{SD}(\text{H}_2)_{\text{trans}}$ et $\text{SD}(\text{H}_1^-)_{\text{trans}}$, les deux atomes de soufre sont positionnés du « même côté » de la molécule (Figure 16, gauche et centre). Cette géométrie, surprenante, est obtenue car les protons des acides sulfoniques sont engagés dans des liaisons hydrogènes intramoléculaires, stabilisant le système et compensant la gêne stérique occasionnée par la proximité des deux fonctions soufrées. Lors du passage $\text{SD}(\text{H}_1^-)_{\text{trans}}$ vers $\text{SD}_{\text{trans}}^{2-}$, la disparition de la dernière liaison hydrogène intramoléculaire et la répulsion électrostatique des charges négatives des deux groupements sulfonates engendrent une variation géométrique (et énergétique) très importante. Sous la forme $\text{SD}_{\text{trans}}^{2-}$, les deux fonctions sulfonates sont localisées à l'opposé l'une de l'autre (Figure 16, droite).

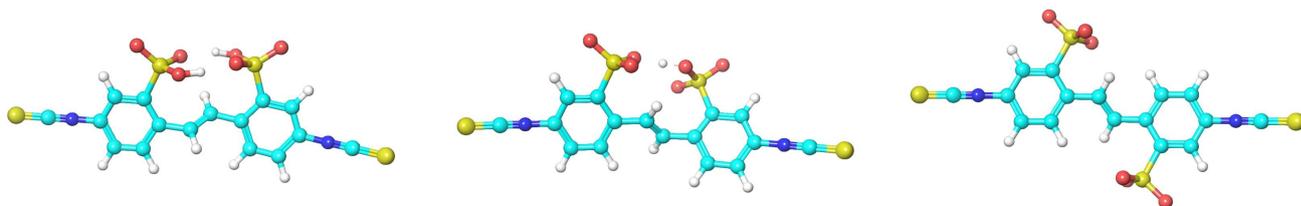


FIGURE 16 – Structure optimisée en phase gazeuse au niveau M06-2X/6-31G(d) des différents états de protonation de l’isomère *trans* du DIDS : DIDS(H₂)_{trans} (gauche), DIDS(H₁)_{trans}⁻ (centre) et DIDS_{trans}²⁻ (droite).

En ce qui concerne les isomères *cis*, des liaisons hydrogènes intramoléculaires sont aussi présentes pour les formes SD(H₂)_{cis} et SD(H₁)_{cis}⁻, comme montré à la Figure 17. En effet, malgré la contrainte stérique très importante occasionnée par la proximité des deux groupements sulfonates au sein de l’isomère *cis*, les liaisons hydrogènes intramoléculaires permettent d’obtenir une stabilisation de l’énergie d’environ une dizaine de kcal.mol⁻¹, par comparaison au même système où les fonctions soufrées seraient opposées spatialement. Comme pour les cas des isomères *trans*, la disparition totale de ces liaisons hydrogènes intramoléculaires est à l’origine de la variation d’environ 4 unités de pK_a entre les valeurs de pK_a¹ et pK_a².

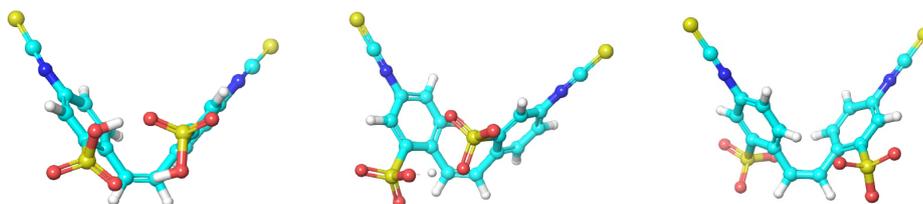


FIGURE 17 – Structure optimisée en phase gazeuse au niveau M06-2X/6-31G(d) des différents états de protonation de l’isomère *cis* du DIDS : DIDS(H₂)_{cis} (gauche), DIDS(H₁)_{cis}⁻ (centre) et DIDS_{cis}²⁻ (droite).

L’analyse des valeurs de pK_a nous permet d’affirmer que seule la forme SD²⁻ est présente en milieu physiologique, quelque soit le système et l’isomère étudié. De plus en phase gaz, les systèmes SD_{trans}²⁻ sont plus stables que les isomères SD_{cis}²⁻ d’environ 4 kcal.mol⁻¹. Dans la suite de notre étude sur les SD²⁻, nous considérons uniquement l’isomère *trans*.

5.2 Propriétés optiques et solvatochromiques du DIDS²⁻

Cette partie s’intéresse aux propriétés optiques des SD²⁻ en solution, obtenues expérimentalement et théoriquement. Dans un premier temps, nous présentons d’abord les expériences menées durant notre thèse puis dans un second temps, nous exposerons les résultats théoriques impliquant des calculs DFT et TD-DFT. Enfin, une dernière partie présente l’apport de la prise en compte des contributions vibroniques pour la simulation des spectres optiques des SD²⁻.

5.2.1 Partie expérimentale

Les spectres d'absorption du DIDS^{2-} ($\approx 0,1 \text{ M}$) ont été, tout d'abord, mesurés en variant la polarité du solvant. Nous avons choisi l'eau, le méthanol (MeOH) et le diméthylsulfoxyde (DMSO). Les spectres ont été enregistrés par moi-même en collaboration avec le Dr Jérôme Graton et Anaïs Goupille (Figure 18). Une attention particulière a été portée à la nature anhydre du DMSO et à la protection des échantillons de la lumière.

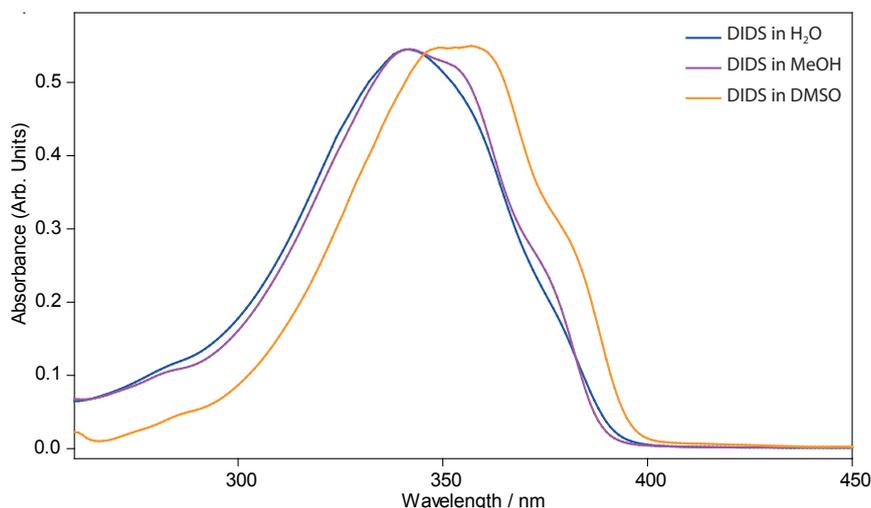


FIGURE 18 – Spectre d'absorption du DIDS^{2-} dans les trois solvants étudiés.

Deux observations principales ressortent de la Figure 18 : premièrement, un décalage des longueurs d'ondes maximales d'absorption (λ_{max}) pour le DMSO ; deuxièmement, une modification de la forme de la bande d'absorption lors du changement de solvant.

Concernant les λ_{max} , nous observons aucun effet solvatochromique entre le spectre obtenu dans l'eau et dans le méthanol ($\lambda_{\text{max}} = 342 \text{ nm}$). Le résultat n'est pas surprenant car ces deux solvants sont tous les deux polaires et protiques. Cependant, au sein d'un solvant polaire et aprotique, comme le DMSO, le λ_{max} est de 347 nm. L'écart solvatochromique de + 5 nm correspond à une énergie d'excitation diminuée de 0,052 eV. En plus du décalage du λ_{max} , nous distinguons un effet plus important des contributions vibrationnelles sur le spectre d'absorption dans le DMSO par rapport aux deux autres solvants, ce qui était attendu car l'emploi de solvants protiques a tendance à « effacer » les contributions vibroniques. Nous avons aussi réalisé des études cinétiques sur la stabilité du DIDS en solution (Figure 19), en enregistrant un spectre toutes les 24 H pendant trois jours. Nous observons que le composé se détériore au bout de 24 heures, quelque soit le solvant. De plus dans l'eau et le MeOH, la diminution de l'intensité du pic principal est corrélée avec l'apparition d'une bande à 280 nm environ. Cette dernière est absente lors de la cinétique dans le DMSO, mais une bande vers les 410 nm apparaît dans le même solvant. La dégradation en fonction du temps est en accord avec la littérature,^[178,179] mais est semble-t-il peu prise en compte dans les études biochimiques.

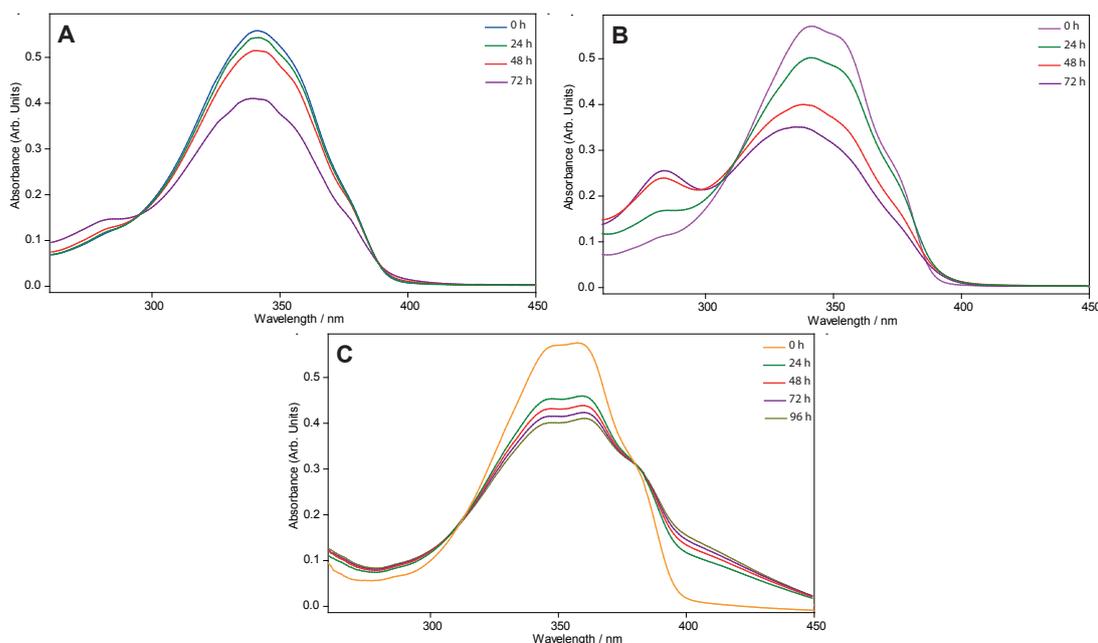


FIGURE 19 – Spectre d’absorption des cinétiques du DIDS^{2-} dans l’eau (A), le méthanol (B) et le DMSO (C).

Le spectre d’absorption du DADS^{2-} (Figure 20) présente une longueur d’onde d’absorption maximale de 339 nm dans l’eau et de 351 nm dans le DMSO, donnant lieu à un déplacement bathochromique de 12 nm (-0,125 eV). Contrairement au DIDS^{2-} , le DADS^{2-} ne montre aucune variation de la forme de la bande d’absorption lorsqu’on change de solvant.

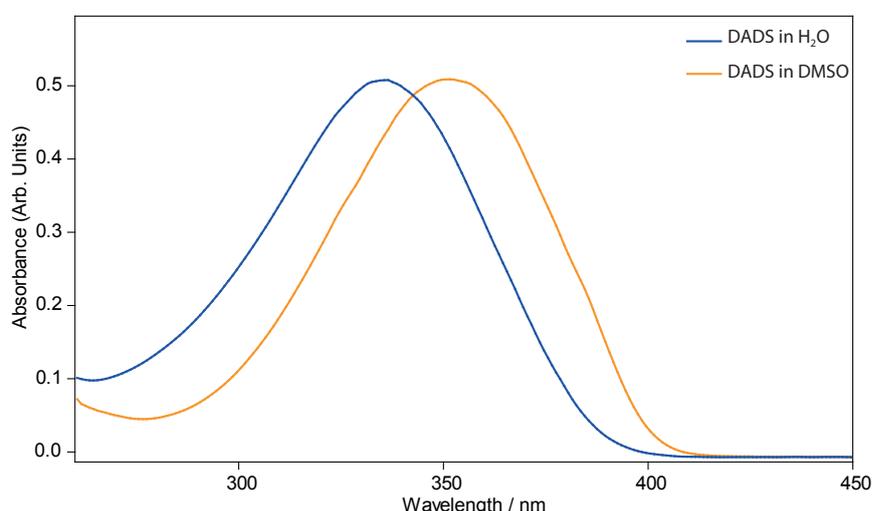


FIGURE 20 – Spectre d’absorption du DADS^{2-} dans l’eau et le DMSO.

Comme pour le DIDS^{2-} , des études cinétiques ont été menées sur la stabilité du DADS^{2-} en solution aqueuse et aucun changement n’est observé durant l’expérience (Figure 21A), permettant ainsi de souligner la stabilité du DADS^{2-} en solution aqueuse. Cependant, il est fréquent de voir un protocole où le DADS^{2-} est solubilisé par l’intermédiaire d’une solution à base de soude caustique avant une seconde dilution dans l’eau. Le passage en milieu basique facilite la solubilisation du

DADS²⁻. La cinétique de cette solution indique très clairement qu'une dégradation du DADS²⁻ a lieu, caractérisé par l'apparition d'un point isobestique autour des 280-290 nm, synonyme de la co-existence de deux espèces (Figure 21B).

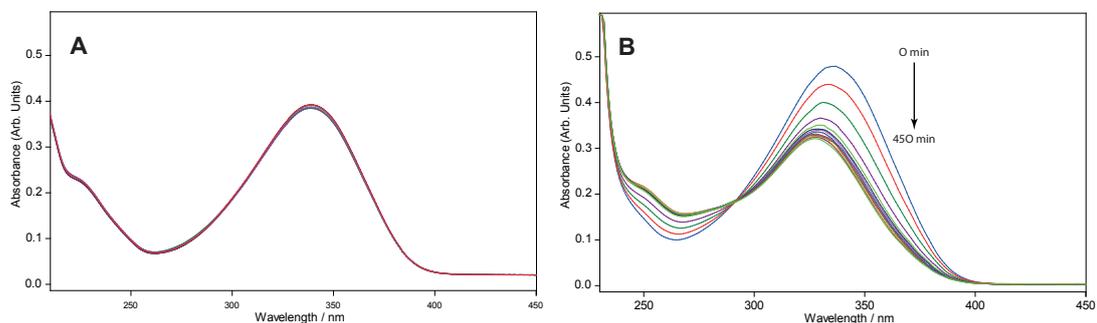


FIGURE 21 – Spectre d'absorption des cinétiques du DADS²⁻ dans l'eau (A) et dans une solution aqueuse, où le DADS²⁻ a été solubilisé en milieu basique puis re-diluer (B).

Des analyses complémentaires en RMN (Résonance Magnétique Nucléaire) du DIDS²⁻ ont été réalisées en collaboration avec Patrick Giraudeau de l'équipe EBSI du laboratoire CEISAM, l'objectif étant d'observer si le produit de dégradation du DIDS²⁻ lors d'une cinétique dans l'eau conduit à la formation de DADS²⁻. Les spectres RMN ¹H du DIDS²⁻ et du DADS²⁻ ont été enregistrés sur un spectromètre Bruker avance HD 700 MHz, équipé d'une sonde cryogénique, dans une solution de D₂O, à une concentration d'environ 1.10⁻³M (Figure 22).

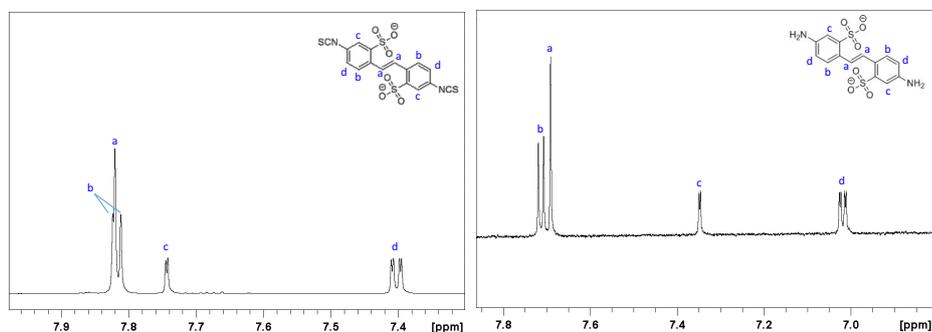


FIGURE 22 – Spectres RMN ¹H du DIDS²⁻ (gauche) et du DADS²⁻ (droite) dans une solution de D₂O.

Nous avons conservé les solutions de DIDS²⁻ et DADS²⁻ dans leur tube RMN, protégés de la lumière, puis nous avons acquis des spectres RMN ¹H une semaine plus tard (Figure 23). Le spectre du DIDS²⁻ après une semaine montre de fortes différences avec celui de la solution fraîche, indiquant un changement de composition significatif au sein du tube RMN. En effet, les déplacements chimiques des pics caractéristiques de ce composé varient beaucoup. De plus, de nombreux pics parasites apparaissent sur le spectre acquis après une semaine, confirmant la présence de nouvelles espèces. Concernant le DADS²⁻, il n'y a aucune variation entre les deux spectres enregistrés à une semaine d'intervalle, confirmant ainsi la stabilité du composé en solution aqueuse.

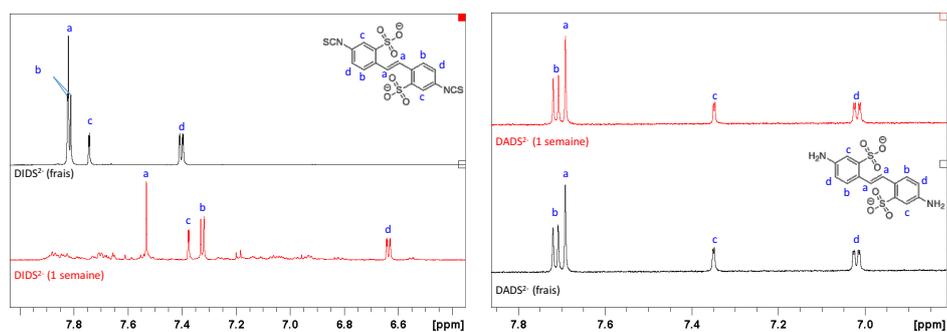


FIGURE 23 – Spectres RMN ^1H (700 MHz) frais et datant d'une semaine du DIDS^{2-} (gauche) et du DADS^{2-} (droite) dans une solution de D_2O .

Une dernière expérience a été menée par l'ajout de DADS^{2-} dans le tube contenant le DIDS^{2-} datant d'une semaine, afin de confirmer ou d'infirmer la présence du DADS^{2-} . Si c'était le cas, l'ajout de DADS^{2-} induirait une augmentation de l'intensité de certains pics déjà présents sur le spectre (Figure 24). Nous constatons que l'ajout de DADS^{2-} n'induit aucune augmentation du pic déjà existant sur le spectre de la solution de DIDS^{2-} datant d'une semaine. Les pics attendus pour le DADS^{2-} apparaissent à la même position que dans la solution de DADS^{2-} seul. Cette observation nous permet de conclure qu'aucune molécule de DADS^{2-} ne s'est formée dans le tube pendant une semaine, ou du moins, que la quantité de DADS^{2-} formée est inférieure à la limite de détection de la technique. Des acquisitions en RMN multidimensionnelle supplémentaires n'ont pas permis d'identifier de façon précise les produits de dégradations du DIDS^{2-} .

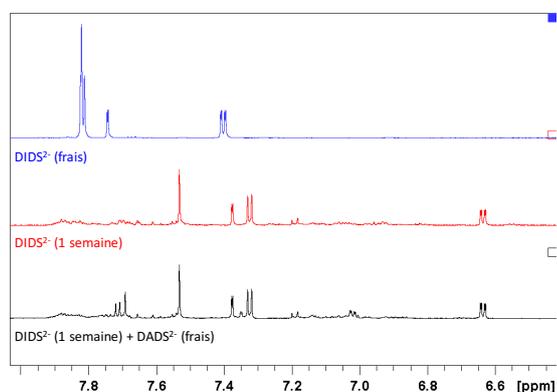


FIGURE 24 – Comparaison des spectres RMN ^1H du DIDS^{2-} frais, du DIDS^{2-} datant d'une semaine et du mélange avec le DADS^{2-} frais dans une solution de D_2O .

L'ensemble des résultats expérimentaux pour le DIDS^{2-} et DADS^{2-} en solution aqueuse est en accord avec les travaux de Jakobsen *et al.* [178] concernant la caractérisation optique en solution aqueuse des deux composés ainsi que de trois autres dérivés appartenant à notre liste de molécules d'intérêt (le DNDS^{2-} , le DATDS^{2-} et le SITS^{2-} , *cfr.* Figure 13, page 47). Les cinétiques ont permis de montrer la dégradation en quelques jours du DIDS^{2-} en solution, la stabilité du DADS^{2-} dans l'eau ainsi que l'effet d'un emploi de la soude caustique sur ce dernier. Les observations montrent l'importance d'une bonne préparation de ces composés en solution. Les résultats en RMN ont pu

déterminer que le DADS²⁻ n'était pas un produit de dégradation directe du DIDS²⁻, du moins dans nos conditions expérimentales.

5.2.2 Partie théorique

Dans cette partie, nous avons d'abord considéré la molécule de DIDS²⁻, qui présente un certain intérêt, notamment, au niveau des effets de solvants et des formes de bandes d'absorption, puis nous avons étendu notre étude aux autres dérivés SD²⁻. Les propriétés optiques du DIDS²⁻ ont été simulées dans les mêmes solvants que ceux cités dans la partie précédente (eau, méthanol et DMSO), ainsi qu'un solvant apolaire et aprotique, le *n*-hexane. Deux modèles de solvation implicite ont été employés, PCM et SMD, ainsi qu'une approche combinée implicite/explicite (IE). La méthodologie adoptée a été la suivante : dans un premier temps, nous avons optimisé la géométrie du DIDS²⁻ en présence du solvant (implicite ou IE) au niveau M06-2X/6-31G(d). Un calcul de fréquences au même niveau de théorie a été effectué pour s'assurer de l'obtention d'un minimum local. Dans un second temps, nous avons calculé les états excités du composé au niveau TD-M06-2X/6-31+G(2d,p), toujours en présence du solvant. Les transitions électroniques ainsi obtenues de manière verticale ont été simulées à l'aide des régimes d'équilibre (eq) et de non-équilibre (neq) ainsi qu'avec trois formalismes différents de la réponse du solvant (LR, cLR et SS). L'analyse des résultats de chaque modèle de solvation implicite, PCM et SMD, seront traités séparément.

Le modèle PCM

Les paramètres géométriques du DIDS²⁻ optimisés en solution avec le modèle implicite PCM sont présentés au Tableau 2. Les propriétés optiques calculées dans les différents solvants sont présentés au Tableau 3. Les trois solvants polaires avec le modèle implicite PCM montrent des caractéristiques géométriques semblables pour les cinq paramètres étudiés, α et τ_{1-4} (Figure 25).

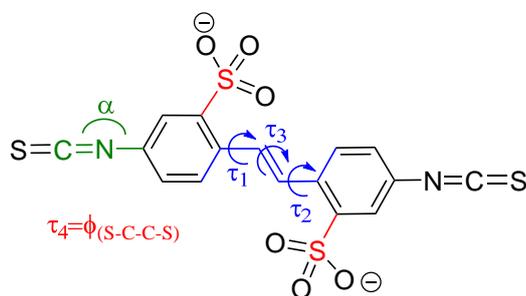


FIGURE 25 – Schéma des différents angles de valence et dièdres étudiés sur le DIDS²⁻.

Pour ce qui concerne la solvation implicite PCM, une variation plus importante des paramètres apparaît dans le cas du *n*-hexane, notamment, au niveau de τ_1 et τ_2 (variation d'environ 2°). Les longueurs d'onde d'absorption du DIDS²⁻ calculées dans les solvants polaires ne montrent

aucun déplacement solvatochromique significatif, quelque soit le formalisme utilisé. Cette observation n'est pas en accord avec l'expérience, où comme mentionné précédemment, un déplacement bathochrome de +5 nm est observé lors du passage de l'eau au DMSO. Concernant le *n*-hexane, un déplacement bathochrome est calculé par rapport aux solvants polaires, excepté lors du calcul LR(eq) où un déplacement hypsochrome de -5 nm apparaît. Ce décalage n'est pas surprenant au vu de la nature apolaire et aprotique du solvant par rapport à l'eau, le méthanol et le DMSO.

Nous nous sommes ensuite tournés vers l'approche de solvation IE afin d'améliorer la description des effets de solvants. Deux contre-ions Na^+ ont d'abord été ajoutés de manière explicite près de chaque groupement $-\text{SO}_3^-$ du DIDS^{2-} . Géométriquement, les seuls changements induits par l'ajout des cations sont la variation des angles τ_1 et τ_2 de $0,8^\circ$, une modification qui est donc négligeable. La signature optique du système $\text{DIDS}^{2-} + 2 \text{Na}^+$ présente peu de différence avec la molécule DIDS^{2-} dans l'eau implicite car la présence des contre-ions n'a aucun effet, ni géométrique, ni électronique sur le DIDS^{2-} . En solution aqueuse, les ions Na^+ sont, en sus, entièrement solvatés, minimisant leur influence sur le soluté.

Tableau 2 – Paramètres géométriques du DIDS^{2-} dans l'eau, le méthanol, le DMSO et le *n*-hexane obtenus au niveau de théorie M06-2X/6-31G(d) avec le modèle de solvation PCM et le groupe de symétrie C_i , sauf pour le système IE(4 DMSO/PCM) (C_1).

Modèle de solvant	Solvant	Position	α ($^\circ$)	τ_1 ($^\circ$)	τ_2 ($^\circ$)	τ_3 ($^\circ$)	τ_4 ($^\circ$)
PCM	eau		179.9	29.6	-29.6	180.0	180.0
	MeOH		179.8	29.4	-29.4	180.0	180.0
	DMSO		180.0	29.5	-29.5	180.0	180.0
	<i>n</i> -hexane		179.6	27.8	-27.8	180.0	180.0
IE(2 Na^+ /PCM)	eau		179.4	30.4	-30.4	180.0	180.0
IE(2 H_2O /PCM)	eau	A-A'	179.4	14.8	-14.8	180.0	180.0
	eau	B-B'	177.3	29.8	-29.8	180.0	180.0
	eau	C-C'	168.9	29.3	-29.3	180.0	180.0
	eau	D-D'	179.7	30.3	-30.3	180.0	180.0
IE(4 H_2O /PCM)	eau	A-A'+B-B'	176.1	15.4	-15.4	180.0	180.0
	eau	B-B'+C-C'	168.5	30.4	-30.4	180.0	180.0
IE(6 H_2O /PCM)	eau	A-A'+B-B'+C-C'	168.5	17.3	-17.3	180.0	180.0
IE(8 H_2O /PCM)	eau	Tous	176.6	16.5	-16.5	180.0	180.0
IE(2 DMSO/PCM)	DMSO		179.9	24.7	-24.7	180.0	180.0
IE(4 DMSO/PCM)	DMSO		179.8	23.8	-25.2	179.4	178.1

Après avoir conclu à l'absence d'effets des contre-ions sur les propriétés optiques du DIDS^{2-} en solution, nous avons ajouté des molécules de solvants (eau et DMSO) de façon explicite au sein du système, que nous avons entouré d'un continuum. Dans un premier temps, nous avons placé 2, 4, 6 et 8 molécules d'eau successivement autour du DIDS^{2-} , à des positions dictées par l'intuition chimique. Pour placer les deux premières molécules explicites, nous avons exploré quatre poses du solvant possibles : A-A', B-B', C-C' et D-D' (Figure 26). Les positions A-A' et B-B' solvatent

Tableau 3 – Longueurs d’onde d’absorption expérimentale et théorique (approximation TD-DFT verticale) du DIDS²⁻ dans l’eau, le méthanol, le DMSO et le *n*-hexane obtenues au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation PCM et le groupe de symétrie C_i, sauf pour le système IE(4 DMSO/PCM) (C₁). Tous les résultats sont exprimés en nm.

Modèle de solvant	Solvant	Position	LR(neq)	LR(eq)	cLR(neq)	cLR(eq)	SS(neq)	SS(eq)
PCM	eau		328	340	321	321	321	322
	MeOH		328	340	321	321	321	321
	DMSO		329	340	321	321	321	322
	<i>n</i> -Hexane		334	334	326	326	323	323
IE(2 Na ⁺ /PCM)	eau		327	339	320	320	318	319
IE(2 H ₂ O/PCM)	eau	A-A'	350	365	350	342	345	346
	eau	B-B'	328	340	321	321	321	322
	eau	C-C'	327	339	321	321	321	322
	eau	D-D'	326	338	320	320	320	321
IE(4 H ₂ O/PCM)	eau	A-A'+B-B'	350	365	342	342	345	346
	eau	B-B'+C-C'	325	337	319	319	319	320
IE(6 H ₂ O/PCM)	eau	A-A'+B-B'+C-C'	347	361	339	339	342	343
IE(8 H ₂ O/PCM)	eau	Tous	349	363	341	341	344	345
IE(2 DMSO/PCM)	DMSO		338	350	330	330	332	333
IE(4 DMSO/PCM) ^a	DMSO		338	350	331	331	332	333
Expérience (λ _{max})	eau				342			
	Méthanol				342			
	DMSO				347			

directement les groupements sulfonates, la position C-C' est proche de la fonction isothiocyanate et la position D-D' fait un pont entre cette dernière fonction et les molécules de solvant en B-B'. Notons qu'aucune molécule d'eau explicite ne fait d'interaction avec l'atome de soufre de la fonction NCS. La position A-A' avec deux molécules d'eau explicites donne lieu à la configuration la plus stable du système. En effet, elle est plus stable de 4 kcal.mol⁻¹ par rapport aux positions B-B' et D-D' et de 10 kcal.mol⁻¹ par rapport à la position C-C'. Comme attendu, l'ajout de molécules d'eau en A-A' mène à des effets géométriques et optiques significatifs, notamment, au niveau des variations de τ₁ et τ₂ (par symétrie). En effet, les deux angles sont modifiés d'une quinzaine de degrés (de 29,6 à 14,8° pour τ₁ et de -29,6 à -14,8° pour τ₂), ce qui mène à une plus grande planarité du squelette carboné de la molécule, améliorant ainsi la délocalisation π du système. Ceci est cohérent avec le déplacement bathochromique observé par rapport aux longueurs d'ondes calculées à l'aide du modèle implicite (Tableau 3). Une différence supérieure à 20 nm est ainsi observée entre les système DIDS²⁻ (PCM-eau) et DIDS²⁻ (IE(2 H₂O-PCM-A-A')), quelque soit le formalisme utilisé. La présence des molécules en B-B' augmente (légèrement) les valeurs de τ₁ et τ₂ (14,8 → 15,4°), sans amener une variation sur les énergies de transitions. La position D-D'

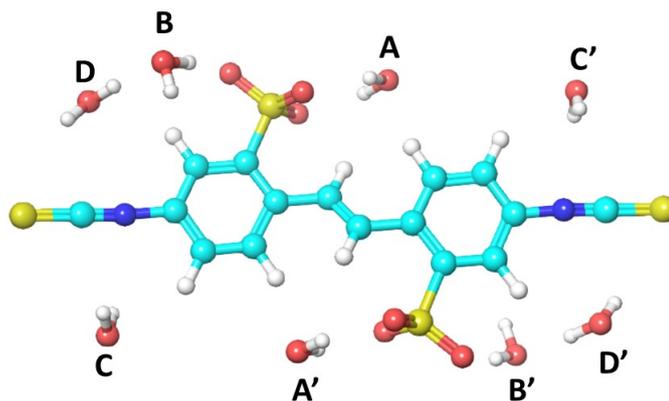


FIGURE 26 – Positions des molécules d'eau explicites autour du DIDS^{2-} après optimisation géométrique.

n'affecte pas les propriétés du DIDS^{2-} , tandis que la position C-C' modifie sensiblement l'angle entre le cycle aromatique et la fonction NCS (angle $\alpha = 168^\circ$), induisant une faible variation sur les propriétés optiques du soluté. Le système comportant quatre molécules d'eau explicites en position A-A' et B-B' possède les mêmes caractéristiques que le système ne contenant que deux molécules en position A-A'. L'influence de la position A-A' est systématiquement prédominante sur le DIDS^{2-} . Le système incluant six molécules d'eau explicites (A-A'+B-B'+C-C') apporte une modification de l'angle α dû à la présence de molécules aux positions C-C'. L'ajout des deux dernières molécules d'eau explicites (en position D-D') a pour effet de contre-balancer l'influence de la position C-C' sur α et modifie très légèrement τ_1 et τ_2 ($0,8^\circ$). Les propriétés optiques sont modifiées seulement de 2 nm par rapport au système avec six molécules explicites. Globalement, nous pouvons affirmer que l'ajout de deux molécules d'eau en position A-A' est suffisant pour modéliser un effet du solvant explicite sur le DIDS^{2-} .

Dans un second temps, nous avons placé deux et quatre molécules de DMSO autour du DIDS^{2-} (Figure 27). Les deux premières molécules sont localisées de chaque côté du DIDS^{2-} , près des groupements sulfonates. Les optimisations de géométrie indiquent que l'addition de deux molécules de DMSO supplémentaires fait apparaître une « cage » de solvant autour du soluté (Figure 27, droite). Les positions des molécules de DMSO hors du plan du DIDS^{2-} sont très différentes que celles adoptées par les molécules d'eau, situées elles dans le plan du DIDS^{2-} .

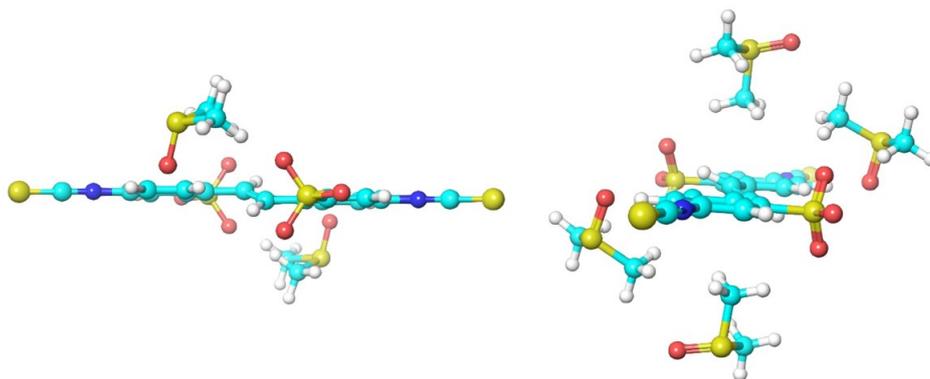


FIGURE 27 – Positions optimisées des molécules de DMSO explicites autour du DIDS^{2-} : deux molécules de DMSO (gauche) et système avec quatre molécules de DMSO (droite).

Les deux premières molécules de DMSO font varier les angles τ_1 et τ_2 d'environ 5° par rapport à la structure obtenue en solvation implicite. Une variation de 1° est observée après l'addition des deux molécules supplémentaires. Il est à noter que le groupe de symétrie C_1 a été utilisé pour le cas du système avec quatre molécules de DMSO car les symétries C_i et C_2 ne permettaient pas d'obtenir une structure optimisée qui soit un minimum énergétique local. La présence des molécules explicites induit un déplacement bathochrome des longueurs d'onde d'absorption calculées d'environ un dizaine de nm en comparaison avec les résultats obtenus avec le continuum seul. Ce décalage bathochromique peut être expliqué par un effet géométrique, électronique ou une combinaison des deux. Nous avons donc extrait la géométrie optimisée du DIDS^{2-} en présence de deux molécules de DMSO explicites. Les propriétés optiques de cette structure calculées en présence d'un continuum pour simuler le solvant indiquent que les mêmes énergies de transitions sont obtenues que dans le cas du système IE(2 DMSO/PCM). Le déplacement bathochrome a donc comme origine la déformation géométrique. Les deux systèmes avec des molécules de DMSO explicites présentent sensiblement les mêmes propriétés optiques et deux molécules de DMSO explicites suffisent donc à décrire les modifications géométriques et optiques sur le DIDS^{2-} .

Notons finalement, que le modèle de solvation implicite PCM ne permet pas de simuler le déplacement solvatochromique observé expérimentalement. En effet, nous n'observons aucun changement des propriétés optiques lors du passage de l'eau au DMSO et en ce qui concerne le modèle IE, un déplacement hypsochromique de 12 à 20 nm (selon le formalisme et les limites utilisés) est observé lors du passage de deux molécules d'eau explicites à deux molécules de DMSO explicites, ce qui est en contradiction avec l'expérience, qui prévoit un déplacement bathochromique de 5 nm. Le modèle de solvation PCM n'est donc pas le plus adéquat pour simuler avec précision les effets solvatochromiques du DIDS^{2-} en solution, que ce soit par la méthode implicite ou combinée IE. À partir de ces résultats, nous avons décidé de tester un second modèle de solvation implicite,

SMD, dont les différences avec le modèle PCM ont été explicitées au chapitre 5.

Le modèle SMD

Les modèles de solvation implicite SMD et combiné IE (avec SMD) ont été utilisés sur les mêmes systèmes que précédemment dans le but de déterminer si le modèle SMD permettait de mieux reproduire les tendances expérimentales. Les paramètres géométriques et les propriétés optiques sont présentés respectivement aux Tableaux 4 et 5.

Tableau 4 – Paramètres géométriques du DIDS²⁻ dans l'eau, le méthanol, le DMSO et le *n*-hexane obtenus au niveau de théorie M06-2X/6-31G(d) avec le modèle de solvation SMD et le groupe de symétrie C_i , sauf pour le *n*-hexane (C_1).

Modèle de solvant	Solvant	Position	α (°)	τ_1 (°)	τ_2 (°)	τ_3 (°)	τ_4 (°)
SMD	eau		178.8	30.9	-30.9	180.0	180.0
	MeOH		179.3	31.6	-31.6	180.0	180.0
	DMSO		179.4	28.3	-28.3	180.0	180.0
	<i>n</i> -hexane		179.8	30.3	-28.4	-177.8	-177.0
IE(2 H ₂ O/SMD)	eau	A-A'	178.0	26.6	-26.6	180.0	180.0
IE(8 H ₂ O/SMD)	eau	Tous	175.4	25.9	-25.9	180.0	180.0
IE(2 DMSO/SMD)	DMSO		179.4	23.2	-23.2	180.0	180.0

Tableau 5 – Longueurs d'onde d'absorption expérimentale et théorique (approximation TD-DFT verticale) du DIDS²⁻ dans l'eau, le méthanol, le DMSO et le *n*-hexane obtenues au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation SMD et le groupe de symétrie C_i , sauf pour le système *n*-hexane (C_1). Tous les résultats sont exprimés en nm.

Modèle de solvant	Solvant	Position	LR(neq)	LR(eq)	cLR(neq)	cLR(eq)	SS(neq)	SS(eq)
SMD	eau		326	338	319	319	321	322
	MeOH		324	336	318	318	319	320
	DMSO		331	342	323	323	321	321
	<i>n</i> -hexane		332	332	325	325	321	321
IE(2 H ₂ O/SMD)	eau	A-A'	334	347	327	327	331	332
IE(8 H ₂ O/SMD)	eau	Tous	335	349	328	329	331	333
IE(2 DMSO/SMD)	DMSO		340	352	332	332	332	333
Expérience (λ_{\max})	eau				342			
	MeOH				342			
	DMSO				347			

Avec le modèle de solvation implicite SMD, des différences géométriques apparaissent entre les différents solvants, contrairement au modèle PCM. Ainsi, les angles τ_1 (et τ_2) varient selon le solvant utilisé (30,9° dans l'eau, 31,6° dans le méthanol, 28,3° dans le DMSO et 30,3° dans

le *n*-hexane). Notons que l'optimisation du DIDS²⁻ dans le *n*-hexane ne permet pas d'obtenir une structure stable de symétrie C_i ou C_2 , expliquant ainsi la différence de valeurs absolues entre τ_1 et τ_2 dans ce milieu. Cette absence de symétrie induit une modification de l'angle τ_4 (177°). En considérant le modèle IE(SMD), on remarque que la présence de molécules d'eau explicites en position A-A' entraîne une variation de 3° de τ_1 et τ_2 . Cette variation est sensiblement moins importante que dans le cas du modèle PCM (14,8°). La présence de 8 molécules d'eau explicites autour du DIDS²⁻ influence l'angle α (175,4°), à cause de la présence de molécules de solvant en positions C-C' et D-D'. Les angles de torsion τ_1 et τ_2 passent de 28,3 à 23,2° lorsque des molécules explicites sont ajoutées.

Les transitions électroniques calculées avec le modèle de solvation SMD montrent un déplacement bathochrome lors du passage de l'eau au DMSO avec les formalismes LR et cLR, quelque soit la limite utilisée (neq ou eq). En effet, des déplacements de +5 et de +4 nm sont calculés en utilisant les méthodes LR(neq) et LR(eq), et un déplacement de +4 nm est constaté pour les méthodes cLR(neq) et cLR(eq). La limite utilisée (neq ou eq) a donc peu d'effet dans le cas présent. Le déplacement solvatochrome n'est pas reproduit en ce qui concerne le formalisme SS, donnant lieu aux mêmes λ_{\max} pour les deux solvants avec la méthode SS(neq) (321 nm) et un déplacement hypsochromique de -1 nm avec la méthode SS(eq). Dans le formalisme IE(SMD), la présence des molécules explicites fait apparaître un déplacement bathochromique de +5 nm avec les formalismes LR et cLR entre l'eau et le DMSO, tandis que le formalisme SS ne donne lieu à aucun déplacement solvatochrome significatif. Il semble évident que le formalisme SS n'est pas adapté pour l'étude des propriétés optiques du DIDS²⁻ et la description des effets de solvant. Les systèmes avec 2 et 8 molécules d'eau présentent des λ_{\max} très similaires, tendance déjà observée avec le modèle PCM.

Le modèle SMD est sans aucun doute le modèle de solvation le mieux adapté à la description des propriétés optiques du DIDS²⁻. Rappelons que plusieurs différences existent entre les deux modèles de solvations. En effet, le modèle SMD prend en compte des termes non-électrostatiques (comme par exemple, les effets de dispersion) et les rayons de cavité des atomes d'oxygène sont différents selon le solvant utilisé (eau ou DMSO), ce qui explique probablement la meilleure description du DIDS²⁻ par ce modèle. De plus, le DIDS²⁻ possède deux groupements chargés et SMD est un modèle qui permet d'obtenir des meilleurs résultats pour les anions que le modèle PCM. La comparaison des modèles implicite et IE montre que les molécules de solvant explicites ne jouent pas un rôle majeur sur les propriétés optiques et qu'une solvation décrite à l'aide d'un continuum adapté est suffisante pour obtenir des résultats satisfaisants. Parmi les 3 formalismes utilisés (LR, cLR, SS) pour simuler les énergies de transitions, les calculs cLR sont ceux qui fournissent l'écart énergétique le plus précis pour un changement de solvant entre l'eau et le DMSO. En effet, les résultats des calculs en cLR neq (eq) donnent une différence d'énergie de -0,047 (-0,045) eV, ce qui est très proche de la valeur expérimentale (-0,052 eV). Le formalisme LR donne aussi des valeurs précises : -0,058 eV pour la limite de neq et -0,043 eV pour celle d'eq. Le formalisme SS donne des valeurs de +0,006 (+0,013) eV pour les limites neq (eq), très éloignées des tendances

expérimentales.

5.3 Propriétés optiques de dérivés SD^{2-}

Dans cette partie, nous étudions les effets auxochromes et solvatochromes des cinq autres dérivés SD^{2-} : $DADS^{2-}$, $DASDSS^{2-}$, $DNDS^{2-}$, $DATDS^{2-}$ et $SITS^{2-}$ (cfr: Figure 13, page 47).

5.3.1 Effets auxochromiques en solution aqueuse

Tableau 6 – Longueurs d’onde d’absorption expérimentale et théorique (approximation TD-DFT verticale) des dérivés du SD^{2-} dans l’eau, le méthanol et le DMSO, obtenues au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation SMD. Tous les résultats sont exprimés en nm.

Molécule	Solvant	LR(neq)	LR(eq)	cLR(neq)	cLR(eq)	SS(neq)	SS(eq)	Exp (nm)
$DADS^{2-}$	eau	312	327	305	307	327	334	339 ^a
	MeOH	310	324	303	305	324	330	
	DMSO	317	331	309	310	324	330	
$DASDSS^{2-}$	eau	317	330	310	322	315	317	
	MeOH	315	327	308	323	356	354	
	DMSO	316	327	308	326	308	309	
$DNDS^{2-}$	eau	334	349	330	338	317	338	354 ^b
	MeOH	332	347	329	337	320	339	
	DMSO	336	347	332	336	316	328	
$DATDS^{2-}$	eau	307	319	301	301	308	310	334 ^b
	MeOH	306	317	299	300	306	307	
	DMSO	313	323	306	306	312	314	
$SITS^{2-}$	eau	318	330	312	313	315	347	336 ^b
	MeOH	316	328	310	311	313	341	
	DMSO	325	337	319	320	322	363	

^aDonnées obtenues lors de notre thèse, ^bdonnées provenant des travaux de Jakobsen *et al.* [178]

La comparaison des résultats théoriques par rapport aux résultats expérimentaux, de la littérature ou obtenus par nous même, est présentée au Tableau 6.

Nous constatons que les λ_{\max} calculés présentent un décalage systématique d’une trentaine de nanomètres vers les plus petites longueurs d’onde par rapport à l’expérience. La surestimation des énergies de transition provient d’une part de l’utilisation de la fonctionnelle M06-2X, [245] et d’autre part, de l’approximation TD-DFT verticale. Nous avons donc testé une autre fonctionnelle, PBE0, pour palier au problème mais cette dernière donne des valeurs plus éloignées comparative-ment aux valeurs expérimentales. L’écart observé avec M06-2X est donc bien la conséquence du pourcentage d’échange exact élevé contenu dans la fonctionnelle hybride (54% comparé au 25%

pour PBE0). À la Figure 28, nous présentons l'erreur associée à la TD-DFT par rapport à l'expérience pour tous les composés SD^{2-} considérés dans l'eau. Le DASDSS $^{2-}$ n'apparaît pas dans cette comparaison car aucune données expérimentales n'est disponible. Bien qu'il soit difficile de dissocier les sources d'erreurs (fonctionnelle, approximation verticale, solvation, etc.), on peut analyser les performances des différents modèles de solvation à l'état excité d'un point de vue purement pragmatique. La méthode LR(eq) est celle qui semble la plus adéquate car elle donne un écart inférieur à 0,2 eV pour l'ensemble des composés (Tableau 7) tandis que la méthode LR(neq) surestime d'avantage les transitions électroniques. Concernant la méthode cLR, nous constatons que la limite utilisée a peu d'influence, à l'exception du DNDS $^{2-}$ où la limite d'équilibre compense un peu la surestimation des calculs due à la fonctionnelle M06-2X. Enfin, les calculs SS donnent des tendances très différentes de celles observées avec les méthodes LR et cLR, notamment dans le cas du SITS $^{2-}$, où une sous estimation de l'énergie de transition électronique est obtenue avec la limite d'équilibre (Figure 28). Cette irrégularité peut être attribuée au phénomène dit de « catastrophe de la polarisation » de la cavité de solvation. En effet, le SITS $^{2-}$ présente un transfert de charge (molécule dissymétrique) à l'état excité conduisant à un traitement imprécis de la polarisation de la cavité de solvation. Au sein du formalisme SS, la cavité de polarisation de l'état fondamental se voit modifier durant les différents cycles itératifs. Une polarisation infinie peut apparaître dans certains cas, notamment, pour les molécules présentant un transfert de charges car l'état fondamental et l'état excité possèdent une polarisabilité très différente.

Tableau 7 – Écart énergétique entre les transitions théoriques et expérimentales pour les dérivés SD^{2-} obtenus au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation SMD, ainsi que l'erreur moyenne (EM) et l'erreur moyenne absolue (EMA) pour chaque méthode de calcul.

Molécule	Solvant	LR(neq)	LR(eq)	cLR(neq)	cLR(eq)	SS(neq)	SS(eq)
DIDS $^{2-}$	eau	0.182	0.044	0.265	0.261	0.233	0.220
DADS $^{2-}$		0.319	0.140	0.409	0.387	0.133	0.055
DNDS $^{2-}$		0.214	0.052	0.251	0.166	0.407	0.166
DATDS $^{2-}$		0.326	0.175	0.409	0.404	0.313	0.286
SITS $^{2-}$		0.213	0.068	0.289	0.278	0.246	-0.118
	EM	0.251	0.096	0.325	0.299	0.267	0.122
	EMA	0.251	0.096	0.325	0.299	0.267	0.169

Nous nous sommes également intéressés aux déplacements auxochromiques pour lesquels le DIDS $^{2-}$ a été choisi comme la molécule de référence. Les énergies de transition calculées avec la méthode LR(eq) sont celles présentant la plus faible erreur moyenne absolue (EMA) pour les effets auxochromes par rapport à l'expérience (EMA = 0,096 eV) et dans la discussion qui suit, nous avons utilisé les valeurs théoriques provenant de cette méthode.

Les valeurs théoriques calculées dans l'eau *via* la méthode LR(eq) peuvent être classées par

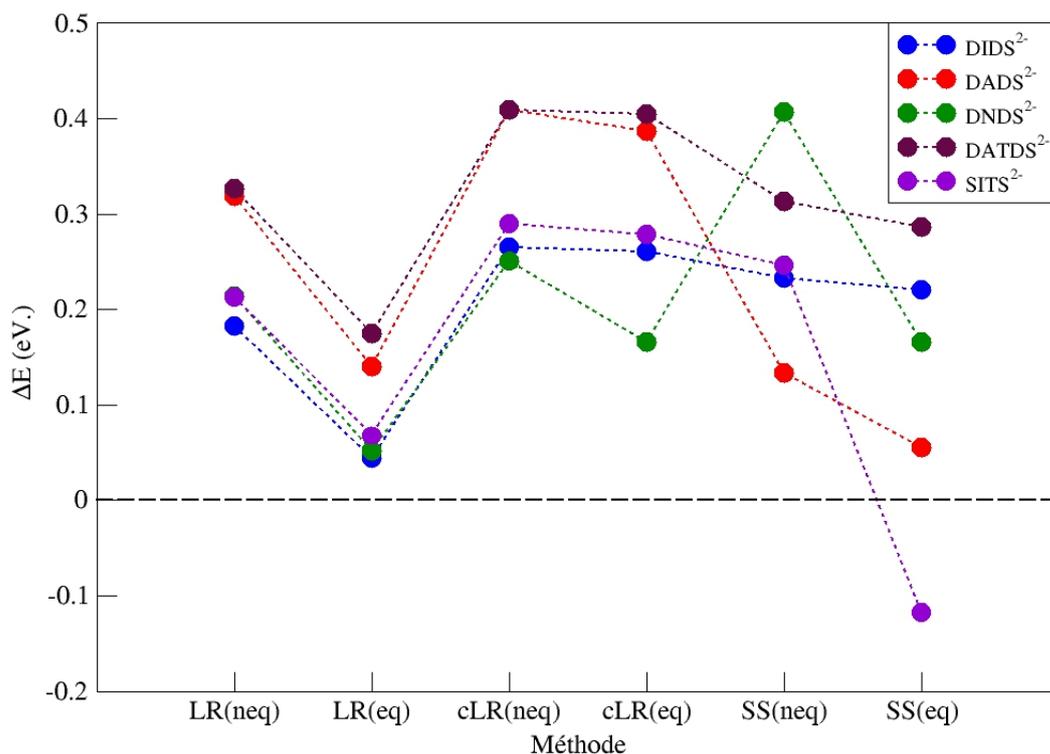


FIGURE 28 – Diagramme représentant l'écart énergétique entre les transitions électroniques calculés par les différents formalismes au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation SMD et l'expérience.

ordre croissant de longueur d'onde : DATDS²⁻ (319 nm) < DADS²⁻ (327 nm) < DASDSS²⁻ (330 nm) < SITS²⁻ (330 nm) < DIDS²⁻ (338 nm) < DNDS²⁻ (349 nm). Le classement est globalement en accord avec les paramètres de Hammett des différentes fonctions chimiques.^[246] Ils sont définis à l'aide de la réactivité d'acides benzoïques et sont très utilisés en chimie organique pour comprendre l'impact de substituants en position *meta* (σ_m) ou *para* (σ_p). Une valeur négative (positive) de σ indique que la fonction chimique a un caractère donneur (attracteur) d'électrons et déstabilise (stabilise) le système lorsque la charge de la base conjuguée de la fonction carboxylique, choisi comme référence par Hammett, apparaît. Dans notre cas, nous pouvons assimiler le groupement $-\text{SO}_3^-$ à la fonction COO^- du benzoate et donc ce sont les paramètres σ_m qui vont nous intéresser. Les groupements NH_2 , N_3 , NCS et NO_2 ont respectivement une valeur de σ_m de -0,16, 0,37, 0,48 et 0,71. Nous constatons qu'il y a une bonne corrélation entre le classement précédent des longueurs d'ondes théoriques et des paramètres de Hammett. Notons cependant, que la théorie ne donne pas le même ordre que l'expérience. En effet, le SITS²⁻ expérimentalement présente un λ_{max} de 336 nm alors que celui du DADS²⁻ est de 339 nm. De plus, la théorie surestime la variation d'énergie entre le DIDS²⁻ et le DATDS²⁻ (DADS²⁻) d'environ 0,13 (0,10) eV,

alors que la différence d'énergie entre le DIDS^{2-} et SITS^{2-} (DNDS^{2-}) est bien reproduite avec 0,02 (0,01) eV de différence avec l'expérience. Nous observons que les systèmes possédant des fonctions chimiques accepteurs d'électrons, tels que le DIDS^{2-} , DNDS^{2-} et SITS^{2-} , ont un écart très faible avec l'expérience tandis que ceux composés de fonctions donneurs d'électrons, tels que le DADS^{2-} et DATDS^{2-} , sont moins bien modélisés.

5.3.2 Effets solvatochromiques du DIDS^{2-} , du DADS^{2-} et structures fines

La Figure 18 (page 52) montre que la forme des bandes d'absorption du DIDS^{2-} est modifiée selon le solvant utilisé. En solution aqueuse, la bande d'absorption présente un seul pic à 342 nm ressemblant à une gaussienne, tandis que dans le DMSO, nous observons trois maxima à 347, 358 et 380 nm, caractéristiques d'une implication significative des vibrations lors du processus d'absorption électronique. Dans le but d'obtenir une description physique complète des effets vibroniques, nous avons utilisé deux paramètres supplémentaires lors de nos calculs : le couplage vibronique et l'énergie de réorganisation du solvant (λ_{reorg}). Cette dernière correspond à la différence d'énergie entre les limites d'équilibre et de non-équilibre. Elle est reliée au phénomène d'élargissement inhomogène des bandes d'absorption.^[247] L'équipe de Barone à Pise a travaillé sur la modélisation de ces effets en les reliant aux paramètres des gaussiennes utilisés pour représenter les bandes d'absorption. L'élargissement des bandes est notamment lié à l'augmentation des interactions électrostatiques entre le soluté et le solvant lors du passage à des solvants plus polaires.

Nous avons donc évalué le spectre d'absorption incluant les contributions vibroniques selon l'approximation de Franck-Condon (FC). Pour cela, les modes normaux de vibrations du système sont calculés dans l'état excité et dans l'état fondamental, permettant par la suite d'obtenir le taux de recouvrement entre les fonctions d'onde vibrationnelle des deux états électroniques étudiés. Le taux de recouvrement de vibrations équivalentes entre les deux états peut être important lorsque les géométries de l'état fondamental et de l'état excité sont très proches. Dans le cas du DIDS^{2-} , ce n'est, cependant, pas le cas car la géométrie de l'état excité est très plane comparée à la géométrie de l'état fondamental, comme illustré à la Figure 29. En effet, les angles τ_1 et τ_2 varient de 30° pour

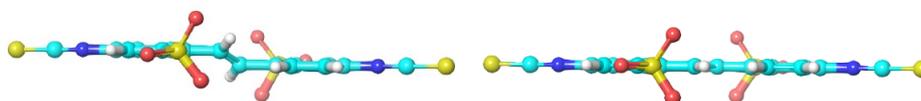


FIGURE 29 – Géométrie optimisée de l'état fondamental (gauche) et de l'état excité (droite) du DIDS^{2-} en solution aqueuse calculée au niveau de théorie SMD-M06-2X/6-31G(d).

atteindre la valeur de 0° lors de l'optimisation de l'état excité dans l'eau et dans le DMSO. Cette différence géométrique conduit à un facteur de recouvrement total très bas tant pour le DIDS^{2-} que pour tous les dérivés SD^{2-} étudiés. Nous atteignons ici les limites de l'approche vibronique utilisée. En effet, dès que les molécules sont flexibles, l'approximation harmonique n'est plus valable et la convergence des calculs vibroniques est difficile. La différence de densité électronique entre

l'état fondamental et le premier état excité S_1 du $DIDS^{2-}$ est présenté à la Figure 30. L'alternance de zones bleues (perte de densité) et de zones rouges (gain de densité) est caractéristique d'une transition $\pi \rightarrow \pi^*$. Cette dernière est délocalisée tout le long du squelette carboné, impliquant aussi la fonction isothiocyanate tandis que les fonctions sulfonates ne participent pas à cette transition électronique.

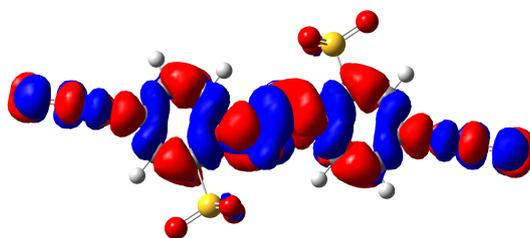


FIGURE 30 – Représentation de la différence de densité électronique entre l'état fondamental et le premier état excité (S_1) du $DIDS^{2-}$ au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation SMD dans l'eau. Les zones en bleu (rouge) correspondent à une diminution (un gain) de densité électronique lors de l'absorption.

Par la suite, nous avons remplacé chaque groupement sulfonate par un atome d'hydrogène pour obtenir le système DID et simulé son spectre vibronique. Si cette approximation est cohérente avec la Figure 30, nous négligeons ainsi la flexibilité de la molécule. En effet, la substitution $-SO_3^- \rightarrow H$ induit une modification géométrique à l'état fondamental (par exemple pour le DID, $\tau_1 = 5^\circ$ et $\tau_2 = -5^\circ$ contre $30,9^\circ$ et $-30,9^\circ$, respectivement pour le $DIDS$). À l'état excité, le DID est plan (τ_1 et $\tau_2 = 0^\circ$) comme le $DIDS^{2-}$. Le calcul du spectre vibronique du DID est possible dans l'approximation car les géométries des deux états sont proches, ce qui permet d'obtenir un bon recouvrement (Facteur FC > 0,9).

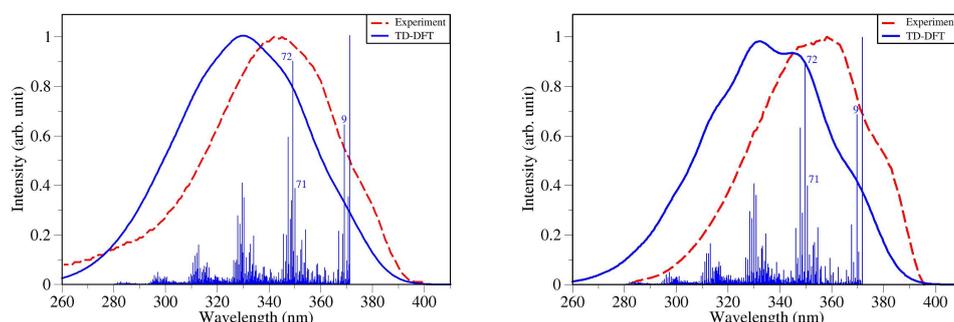


FIGURE 31 – Spectres vibroniques théoriques au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation SMD (bleu) et expérimentaux (rouge) du $DIDS^{2-}$, dans l'eau (gauche) et dans le DMSO (droite).

La Figure 31 montre la comparaison entre le spectre expérimental du $DIDS^{2-}$ et le spectre vibronique calculé du DID, dans l'eau (gauche) et dans le DMSO (droite). Dans les deux cas, le

spectre calculé permet de bien reproduire la forme de bande obtenue expérimentalement, notamment dans le cas du DMSO où les trois maxima sont présents. L'écart sur l'axe des abscisses entre les deux spectres est dû à la fonctionnelle utilisée (comme discuté précédemment). Le bon accord entre le spectre expérimental et théorique indique que les groupements sulfonates ont un effet stérique significatif sur la géométrie de l'état fondamental mais n'ont pas d'influence sur les contributions vibroniques. L'analyse des fréquences de vibration donnant lieu à de grandes contributions vibroniques met en lumière trois d'entre elles qui ont une importance dans les deux solvants. Ces trois modes de vibrations sont des combinaisons d'élongation et de cisaillement dans le plan du squelette carboné. Ces torsions sont aussi présentes dans le cas du stilbène sous sa forme *trans*.^[248] Les largeurs à mi-hauteur des gaussiennes diffèrent en fonction des solvants. Une relation entre la largeur à mi-hauteur des bandes d'absorption et l'énergie de réorganisation du solvant λ , en effet, été établie :^[247]

$$W_{\text{pol}} \approx 4\sqrt{\ln 2 k_B T \lambda_{\text{reorg}}} , \quad (5.6)$$

où k_B est la constante de Boltzmann, T , la température (en kelvin) et $\lambda_{\text{reorg}} \approx E_{\text{neq}}^{\text{S1}} - E_{\text{eq}}^{\text{S1}}$. Le terme W_{pol} est une quantité qui s'apparente à une largeur de bande. Dans le cas du DIDS²⁻ dans l'eau (DMSO), la valeur de W_{pol} calculée est de 0,20 (0,18) eV avec le formalisme LR. En d'autres termes, le W_{pol} dans l'eau est 10% plus grand que celui dans le DMSO, impliquant que la largeur à mi-hauteur du spectre calculé dans l'eau doit être 10% plus large que celle dans le DMSO. Dans notre cas, nous avons utilisé, respectivement, 0,20 et 0,18 eV comme largeur à mi-hauteur pour l'eau et le DMSO. Dans le cas du DADS²⁻, le changement de solvant n'a pas d'incidence sur la forme de la bande d'absorption expérimentale, comme le montre la Figure 20 (page 53). Nous avons néanmoins calculé le spectre vibronique de cette molécule dans les deux solvants (Figure 32). Comme expliqué précédemment, la substitution $-\text{SO}_3^- \rightarrow \text{H}$ a été effectuée pour donner lieu au système DAD.

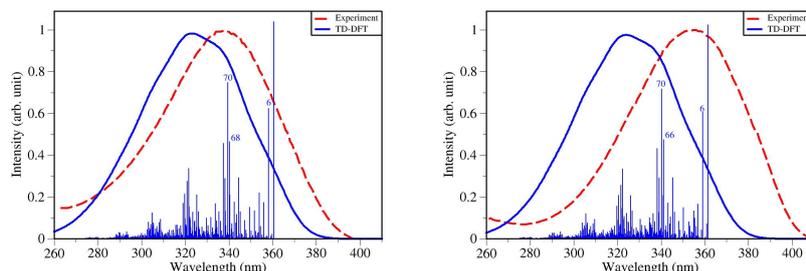


FIGURE 32 – Spectres vibroniques théoriques au niveau de théorie M06-2X/6-31+G(2d,p)//M06-2X/6-31G(d) avec le modèle de solvation SMD (bleu) et expérimentaux (rouge) du DADS²⁻, dans l'eau (gauche) et dans le DMSO (droite).

Les valeurs des deux W_{pol} est de 0,077 eV pour l'eau et de 0,065 eV pour le DMSO. Ces valeurs sont deux fois plus grandes que celles trouvées avec le DIDS²⁻, expliquant ainsi qu'aucune contribution vibrationnelle n'est observée expérimentalement dans le cas du DADS²⁻. Les

modes de vibrations majeurs pour le DADS²⁻ sont également des combinaisons d'élongation et de cisaillement dans le plan de la molécule, comme pour le DIDS²⁻.

5.4 Conclusions

Dans ce chapitre, nous avons étudié les propriétés physico-chimiques de six composés de la famille des SD à l'aide de la DFT et de la TD-DFT. Ces composés sont connus pour être des inhibiteurs de la protéine Rad51. Cette dernière protéine polymérase réparatrice de l'ADN est surexprimée dans les cellules cancéreuses, entraînant la prolifération du cancer.

Dans un premier temps, nous avons déterminé, à l'aide d'un cycle de Börn-Haber, l'état de protonation des molécules SD en solution aqueuse. Les résultats obtenus ont permis d'identifier le caractère di-anionique de ces molécules en milieu physiologique (pH = 7,4). Cette première information structurale est capitale pour notre future investigation avec la protéine de transport ASH. Des spectres d'absorption expérimentaux, effectués par nos soins, ont permis de confirmer la présence d'une seule espèce en solution. Dans un second temps, nous avons réalisé des mesures de spectres d'absorption expérimentaux de deux SD²⁻ (DIDS²⁻ et DADS²⁻) dans différents solvants et nous avons observé un déplacement bathochromique des longueurs d'onde d'absorption maximale lors du passage de l'eau vers le DMSO. Nous avons tenté de reproduire ce déplacement bathochromique en employant la TD-DFT et nous avons constaté que le modèle de solvant implicite SMD est plus adapté que PCM pour décrire ce phénomène. La prise en compte de termes non-électrostatiques et la modification des rayons de cavité des atomes d'oxygène entre les deux solvant (eau et DMSO) semblent être deux facteurs importants pour les propriétés optiques des SD²⁻. Différentes approches permettant de simuler la réponse du solvant à l'état excité ont été testées et la méthode LR(eq) est celle la plus adaptée pour décrire le solvatochromisme et les effets auxochromiques de nos composés. Expérimentalement, nous observons que la forme de la bande d'absorption maximale du DIDS²⁻ présente un double pic dans le DMSO alors que dans l'eau un seul pic est observé. Cette caractéristique spectrale dans le DMSO est due à l'implication de transitions vibroniques lors du processus d'absorption. La prise en compte du recouvrement entre la fonction d'onde de l'état fondamental et de l'état excité a permis d'obtenir cette forme caractéristique de bande.

La partie suivante de cette thèse se concentrera sur la modélisation de complexes ASH-SD²⁻ par des méthodes basées sur la mécanique moléculaire, dans l'objectif de déterminer l'aptitude de la protéine ASH à accueillir et transporter ces inhibiteurs de Rad51 vers les cellules cancéreuses. Le prochain chapitre reprend donc les bases de la mécanique moléculaire et le chapitre suivant les résultats obtenus à l'issue des simulations du complexes ASH-SD²⁻.

Troisième partie

Modélisation du complexe $ASH-DADS^{2-}$

6 Mécanique et dynamique moléculaires

Comme la mécanique quantique, la mécanique moléculaire (MM) repose aussi sur le découplage des mouvements nucléaires et électroniques des atomes d'un système moléculaire (approximation de Born-Oppenheimer) mais seule la partie nucléaire est explicitement traitée. En effet, seule la position des atomes est prise en compte lors du calcul de l'énergie totale du système car la MM ne détermine pas les contributions électroniques. Les liaisons chimiques sont représentées par des ressorts, reliant deux sphères dures simulant les atomes et des paramètres comme la constante de force d'un ressort sont utilisées. Les paramètres et potentiels qui constituent un champ de force sont empiriques ou proviennent de calculs quantiques. L'utilisation de champs de force qui comprennent un ensemble cohérent de paramètres permet l'évaluation de l'énergie de systèmes biologiques complexes tels que les protéines, les membranes et l'ADN.

6.1 Énergie totale d'un champ de force

L'Hamiltonien décrivant l'énergie totale d'un système (supra)moléculaire est semblable à celui utilisé en mécanique quantique. Cependant, seule la partie nucléaire du système est présente :

$$H_{\text{MM}} = \sum_{\text{noyaux}} (T_{\text{MM}} + V_{\text{MM}}), \quad (6.1)$$

où T_{MM} et V_{MM} correspondent respectivement l'énergie cinétique et potentielle de l'ensemble des noyaux du système. Les termes potentiels peuvent être calculés pour une géométrie donnée tandis que les termes cinétiques sont calculés *via* des simulations.

Chaque atome du système moléculaire est identifié par un *type d'atome*, possédant des paramètres et des potentiels qui lui sont propres. Par exemple un carbone sp^3 possède des propriétés différentes par rapport à un carbone sp^2 . Les caractéristiques de chaque *type d'atome* sont utilisées lors du calcul des différents termes composants l'énergie potentielle du système :

$$V_{\text{MM}} = \underbrace{V_{\text{liaisons}} + V_{\text{angles}} + V_{\text{dièdres}} + V_{\text{impropres}}}_{V_{\text{liés}}} + \underbrace{V_{\text{vdW}} + V_{\text{élec}}}_{V_{\text{non-liés}}}. \quad (6.2)$$

Les quatre premiers termes de l'équation (6.2) sont regroupés dans un ensemble appelé $V_{\text{liés}}$, caractérisant les interactions d'un atome avec ses plus proches voisins (jusqu'à 3 liaisons). Les termes V_{liaisons} et V_{angles} correspondent respectivement aux potentiels de liaison et d'angle de valence,

caractérisés le plus souvent par un potentiel harmonique :

$$V_{\text{liaisons}} = \sum_{\text{liaisons}} k_r (r - r_0)^2, \quad (6.3)$$

$$V_{\text{angles}} = \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2, \quad (6.4)$$

où r est la distance entre deux atomes liés, r_0 , la distance d'équilibre entre ces deux atomes, θ , l'angle de valence entre les trois atomes liés, θ_0 , l'angle de valence d'équilibre et k_r (k_θ) est la constante de force du potentiel de liaison (d'angle de valence) correspondant. Dans un champ de force classique, l'utilisation d'un potentiel harmonique implique que les ruptures de liaison ne peuvent être étudiées. Le calcul du terme $V_{\text{dièdres}}$ diffère légèrement des deux précédents termes car celui-ci ne peut être associé à un potentiel harmonique puisque les angles dièdres peuvent présenter plusieurs valeurs extrêmes au niveau de leur potentiel. La formulation du terme $V_{\text{dièdres}}$ se présente donc sous la forme :

$$V_{\text{dièdres}} = \sum_{\text{dièdres}} k_\phi (1 + \cos(n\phi - \gamma)), \quad (6.5)$$

où ϕ est la valeur de l'angle dièdre, γ , l'angle dièdre d'équilibre correspondant à un minimum énergétique, n , la multiplicité de la torsion, c'est-à-dire, le nombre de minima énergétiques que peut atteindre le potentiel $V_{\text{dièdres}}$ et k_ϕ , la constante de raideur de l'angle dièdre. Les angles dièdres impropres sont aussi pris en compte et permettent d'assurer certaines contraintes géométriques, comme par exemple, la géométrie plane des groupements carbonyles :

$$V_{\text{impropres}} = \sum_{\text{impropres}} k_i (d - d_0)^2, \quad (6.6)$$

où d est la valeur de l'angle impropre, d_0 , la valeur d'équilibre et k_i , la constante de force de l'angle impropre.

Les deux derniers termes de l'équation (6.2) sont considérés comme des interactions dites non-liées car elles concernent des atomes pouvant être séparés par plus de 3 liaisons. Le potentiel, V_{vdW} , correspond aux interactions de van der Waals présents au sein du système et est généralement représenté par un potentiel de Lennard-Jones (LJ), dit 12-6 :

$$V_{\text{vdW}} = \sum_i \sum_{j>i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (6.7)$$

où r_{ij} est la distance entre les atomes i et j , ϵ_{ij} la valeur de la profondeur du puits du potentiel de LJ et σ_{ij} est la distance où ce potentiel de LJ est nulle. Lorsque les caractéristiques du potentiel de LJ pour une interaction hétérogène (i et j sont de nature chimique différente) ne sont pas connus, des règles empiriques permettent d'obtenir les valeurs ϵ_{ij} et σ_{ij} :

$$\epsilon_{ij} = (\epsilon_i \epsilon_j)^{\frac{1}{2}}, \quad (6.8)$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}. \quad (6.9)$$

Dans l'équation (6.7), le terme en $(\sigma_{ij}/r_{ij})^6$ représente les interactions de dispersion, qui sont de nature attractive tandis que le terme en $(\sigma_{ij}/r_{ij})^{12}$ décrit des interactions répulsives, qui peuvent être apparentées au principe de Pauli. Enfin, le potentiel, V_{elec} , représente l'interaction entre deux atomes chargés électriquement ; il est calculé à partir de la loi de Coulomb :

$$V_{\text{elec}} = \sum_i \sum_{j>i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \quad (6.10)$$

où q_i et q_j sont respectivement les charges électriques des atomes i et j , et ϵ_0 , la permittivité diélectrique du vide.

6.2 Famille de champ de force

Il existe de nombreux champs de force qui ont été développés, chacun possédant ses propres paramètres et potentiels. Ainsi, certains champs de force sont plus adaptés pour décrire le comportements des protéines, de l'ADN, de fluides, etc. Parmi tous les champs de force, il est commun de les classer en trois familles : les champs de force classiques, les champs de force polarisables et les champs de force réactifs.

La première famille a la caractéristique de présenter une charge atomique constante par *type d'atome*. Le calcul et la répartition de ces charges au sein d'un motif moléculaire est fait de telle manière que la charge globale soit reproduite. Parmi les champs de force classiques, nous pouvons citer AMBER (*Assisted Model Building with Energy Refinement*)^[249,250] et CHARMM (*CHemistry at Harvard Macromolecular Mechanics*),^[251,252] qui sont deux champs de force très populaires. AMBER est paramétré essentiellement à partir de systèmes biologiques (ADN, peptides et protéines) et possède plusieurs versions (ff94, ff99, etc.), dont une permettant d'obtenir les paramètres pour de petites molécules organiques (*Generalized Amber Force Field*, GAFF).^[253] CHARMM est développé, entre autres, par l'équipe de Martin Karplus (prix Nobel 2013) et s'appuie lui aussi sur des propriétés et paramètres provenant de systèmes biologiques. Une déclinaison pour la simulation de molécules organiques existe également (*CHARMM General Force Field*, CGENFF).^[254] Le calcul des énergies et des forces sont très semblables entre AMBER et CHARMM, mais les paramètres et charges présents dans ces deux champs de force sont, eux, bien différents. Lors de nos simulations de MD, nous avons utilisé le champ de force CHARMM36^[255] pour décrire la protéine et le champ de force CGENFF pour la partie ligand.

La deuxième famille correspond aux champs de force polarisables, où la polarisabilité des atomes est implicitement simulée, notamment, aux niveaux des termes non-liés (interactions de van der Waals et électrostatique). Il existe différentes approches permettant d'inclure cette polarisabilité au sein des champs de force. Le champ de force PFF (*Polarizable Force field*) met en place des charges ponctuelles pour introduire la polarisabilité des atomes,^[256] alors que le champ de force AMOEBA (*Atomic Multipole Optimized Energetics for Biomolecular Applications*) utilise des multipôles et des dipôles induits.^[257,258]

La troisième famille est constituée des champs de force réactif, qui ont la caractéristique de prendre en compte la création et la rupture de liaison au sein du système. Le calcul de l'ordre de liaison en fonction de la distance inter-atomique permet à ces champs de force de modéliser la réactivité d'un système. Une des autres différences majeures avec les champs de force classiques est le traitement des termes V_{liaisons} et V_{angles} comme des potentiels de Morse, ce qui est nécessaire pour introduire la notion de dissociation dans ces deux termes. ReaxFF (*Reactive Force Field*)^[259] et RWFF (*Reactive Water Force Field*)^[260] appartiennent à cette famille des champs de force réactifs.

6.3 Dynamique moléculaire

La dynamique moléculaire (*Molecular Dynamics*, MD) permet d'étudier le comportement d'un système évoluant dans le temps. Les premières simulations ont été effectuées en 1957 par Alder et Wainwright dans le cadre de l'étude de collisions de sphères dures au sein d'une boîte.^[261] En 1977, McCammon *et al.* publièrent la première simulation de MD d'une protéine pendant un temps de 8,8 ps.^[262] Par la suite, des molécules de plus en plus grosses ainsi que des complexes supramoléculaires ont pu être simulés, amenant à la mise en place des champs de force et d'ensembles de simulation dans le but de rationaliser des résultats expérimentaux. Les progrès technologiques en informatique ont grandement contribué à l'avancée de la MD au point que de longues simulations de l'ordre de la micro à la milliseconde sont maintenant réalisables.^[263–265]

6.3.1 Résolution des équations du mouvement

Les simulations de MD consistent à calculer l'ensemble des forces s'appliquant au système et donc des mouvements des atomes, de façon itérative pendant un certain temps à une température donnée. Les atomes vont se déplacer en fonction de leurs interactions, en accord avec les équations de mouvement de la mécanique Newtonienne. Les forces F_i qui s'appliquent sur un atome peuvent être définies comme l'opposé du gradient de l'énergie potentielle du système :

$$F_i = -\frac{dV}{dr_i}, \quad (6.11)$$

où V est l'énergie potentielle du système et r_i , le vecteur position dans une coordonnée de l'espace de l'atome i . L'équation de mouvement de Newton est donnée par la relation :

$$\frac{dp_i}{dt} = F_i, \quad (6.12)$$

où p_i est la quantité de mouvement de l'atome i , c'est-à-dire, le produit de sa masse m_i par sa vitesse v_i . La position d'un atome après un délai temporel dans un intervalle fini, suivant une seule dimension, peut être décomposée en série de Taylor comme :

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt} \Delta t + \frac{d^2x(t)}{dt^2} \frac{\Delta t^2}{2} + \dots \quad (6.13)$$

L'équation ci-dessus montre que la position, la vitesse et l'accélération au temps t sont les données nécessaires pour calculer la position à l'intervalle de temps suivant si on se limite aux trois premiers termes. La dérivée seconde de la position en fonction du temps correspond à son accélération et peut être traitée par la deuxième loi de Newton :

$$\frac{d^2x(t)}{dt^2} = \frac{F_x}{m_i}, \quad (6.14)$$

où F_x est la composante en x de la totalité des forces s'exerçant sur l'atome i . La combinaison des équations (6.11) et (6.14) amène à une équation différentielle :

$$\frac{d^2x(t)}{dt^2} m_i = -\frac{dV}{dr_i}. \quad (6.15)$$

La troncature de l'équation (6.13) couramment effectuée à partir de l'ordre trois a l'inconvénient de ne pas respecter la conservation totale de l'énergie, c'est-à-dire, la troisième loi de Newton. Pour mieux respecter ce principe et améliorer l'intégration des équations de mouvements dans le temps, divers algorithmes ont été mis en place. L'un des plus utilisé est l'algorithme de Verlet,^[266] qui établit deux séries de Taylor tronquées à partir de l'ordre quatre, pour les temps $t + \Delta t$ et $t - \Delta t$:

$$x_i(t + \Delta t) = x_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 + \frac{d^3x_i(t)}{dt^3} \frac{\Delta t^3}{6} + \dots, \quad (6.16)$$

$$x_i(t - \Delta t) = x_i(t) - v_i(t)\Delta t + \frac{1}{2}a_i(t)\Delta t^2 - \frac{d^3x_i(t)}{dt^3} \frac{\Delta t^3}{6} + \dots. \quad (6.17)$$

La somme des équations (6.16) et (6.17) permet de déterminer la position au temps $t + \Delta t$ à partir de seulement $x(t)$, $x_i(t - \Delta t)$ et de $a_i(t)$:

$$x_i(t + \Delta t) = 2x_i(t) - x_i(t - \Delta t) + a_i(t)\Delta t^2. \quad (6.18)$$

Cet algorithme présente, cependant, trois inconvénients. Le premier est que la position $x_i(t + \Delta t)$ est obtenue par l'addition d'un terme numériquement petit ($a_i\Delta t^2$), par rapport à la différence de deux grands nombres ($2x_i(t) - x_i(t - \Delta t)$). Le deuxième est qu'il faut connaître la position de l'atome i au temps $t - \Delta t$, qui n'est pas accessible lors des conditions initiales. Pour cela, une étape préliminaire consistant à résoudre l'équation (6.17) afin de déterminer $x_i(t - \Delta t)$ est nécessaire. Le dernier inconvénient est que la vitesse $v_i(t)$ n'est pas présente de façon explicite dans l'équation de propagation des mouvements dans le temps, ce qui peut poser des problèmes lors de simulations où un contrôle de la température est effectué. L'une des approches pour obtenir $v_i(t)$ est de soustraire entre elles les équations (6.16) et (6.17), tronquées au troisième ordre :

$$v_i(t) = \frac{x_i(t + \Delta t) - x_i(t - \Delta t)}{2\Delta t}. \quad (6.19)$$

L'algorithme « saute-mouton » (*leapfrog*) est aussi très populaire car il présente plusieurs avantages par rapport à l'algorithme de Verlet.^[267] Il est basé sur les deux relations suivantes :

$$x_i(t + \Delta t) = x_i(t) + v_i(t + \frac{1}{2}\Delta t)\Delta t, \quad (6.20)$$

$$v_i(t + \frac{1}{2}\Delta t) = v_i(t - \frac{1}{2}\Delta t) + a_i(t)\Delta t. \quad (6.21)$$

Une des particularités de cet algorithme est qu’il utilise la position et la vitesse de l’atome i à des pas de temps différents. Il y a un décalage d’un demi-intervalle de temps entre ces deux quantités. La Figure 33 représente le fonctionnement de cet algorithme.

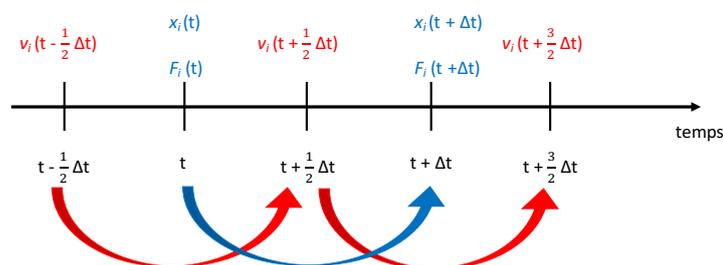


FIGURE 33 – Schéma de l’intégration des équations de mouvements dans le temps par l’algorithme *leapfrog*.

Le premier avantage de l’algorithme *leapfrog* est que la vitesse est présente de façon explicite, ce qui permet un meilleur contrôle des conditions de température durant les simulations de MD. Le deuxième est que l’utilisation d’intervalle temporel demi-entier permet de calculer des positions pour l’atome i plus proche d’une « vraie » trajectoire. En effet, plus l’intervalle temporel se rapproche de zéro, plus la trajectoire calculée sera précise. Cependant, suivant ce formalisme, la position x_i et la vitesse v_i au même temps t ne sont pas disponibles. Il est possible d’obtenir $v_i(t)$ en faisant l’approximation que :

$$v_i(t) = \frac{v_i(t - \frac{\Delta t}{2}) + v_i(t + \frac{\Delta t}{2})}{2} . \quad (6.22)$$

6.3.2 Ensembles canoniques

Lors d’une simulation de MD, il est possible d’appliquer des contraintes physiques au système afin de l’étudier dans des conditions proches de ceux de l’expérience, en régulant des paramètres spécifiques : température, pression, volume, etc.

Ensemble NVE

Dans cet ensemble, le nombre de particules (N) au sein du système et l’énergie totale (E) qui lui est associée sont conservés lors de l’intégration des équations de mouvements dans le temps. Si le système ne change pas de volume (V) dans le temps, alors il se trouve en condition dite micro-canonique où seule la température et la pression fluctueront durant la simulation.

Ensemble NVT

Lorsque des conditions isothermes sont désirées, l’ensemble micro-canonique NVE ne peut pas être utilisé ; il est alors remplacé par l’ensemble canonique NVT. Il est essentiel que la température

soit une grandeur indépendante et, non pas dérivée d'une autre propriété. Hors, la température est reliée à l'énergie cinétique du système par la relation :

$$E_{\text{cinétique}} = \sum_{i=1}^N \frac{|p_i|^2}{2m_i} = \frac{3}{2} N k_B T, \quad (6.23)$$

où N est le nombre de degrés de libertés, k_B , est la constante de Boltzmann et T , la température. Une des manières de maintenir la température à la valeur souhaitée est d'ajuster les vitesses à chaque pas d'intégration par un facteur de correction λ . La différence de température engendrée par la correction des vitesses s'exprime sous la forme :

$$\Delta T = \frac{1}{2} \sum_{i=1}^N \frac{2 m_i (\lambda v_i)^2}{3 N k_B} - \frac{1}{2} \sum_{i=1}^N \frac{2 m_i v_i^2}{3 N k_B}, \quad (6.24)$$

$$\Delta T = (\lambda^2 - 1) T, \quad (6.25)$$

$$\lambda = \sqrt{T'/T}, \quad (6.26)$$

où T' est la température désirée et T , la température système. Cette méthode permet bien au système de se trouver dans des conditions d'un ensemble canonique NVT mais la conservation des quantités de mouvements n'est pas respectée, donnant lieu à une simulation qui n'est plus dans les conditions newtoniennes. Une autre méthode consiste à mettre en place un thermostat externe maintenant une température constante. Parmi les méthodes basées sur cette idée, nous pouvons citer les modèles proposés par Berendsen *et al.*^[268] et celui développé par Nosé-Hoover.^[269–271] Le thermostat que nous avons utilisé lors de notre thèse de doctorat pour les simulations de MD dans un ensemble NVT repose sur les équations de Langevin, une alternative de la deuxième loi de Newton.^[272] L'équation de mouvement des atomes est modifiée par l'ajout de termes de friction et stochastique :

$$\frac{d^2 x(t)}{dt^2} m_i = F_i(t) - \gamma \frac{dx(t)}{dt} + \sqrt{\frac{2\gamma k_b T}{m_i}} R_i(t), \quad (6.27)$$

où γ est la constante de friction et $R(t)$, une force aléatoire appliquée sur l'atome i , simulant des mouvements brownien. Cette équation couple le mouvement de chaque atome à des forces dissipatives et fluctuantes, dépendante de la température, permettant ainsi d'être dans des conditions isothermes.

Ensemble NPT

La pression, dans un ensemble isobare et isotherme NPT, est contrôlée par l'addition d'un degré de liberté au système. Ce dernier correspond au volume de la boîte de simulation qui peut être ajusté afin de reproduire la pression induite par l'effet d'un piston. Ce principe a été proposé par Andersen en 1980.^[273] Cette idée d'effet de piston fut reprise puis améliorée, donnant lieu à l'algorithme du piston de Langevin.^[274] Comme dans le cas de l'ensemble NVT, des termes de friction et stochastique ont été ajoutés dans l'équation de l'effet de mouvement du piston :

$$\frac{d^2 V(t)}{dt^2} = \frac{P(t) - P_{\text{ref}}}{M_p} - \zeta \frac{dV(t)}{dt} + R_p(t), \quad (6.28)$$

où $\frac{d^2V(t)}{dt^2}$ peut être assimilé à l'accélération du volume, M_p est la « masse » du piston, $P(t)$ et P_{ref} sont respectivement la pression du système au temps t et la pression désirée, ζ est une fréquence de collision et $R_p(t)$, un terme stochastique.

6.4 Solvation et conditions périodiques

Pour simuler un système biologique, il est important de recréer les conditions physiologiques, dont notamment, la solvation du milieu. Pour cela, une boîte de molécules explicites de solvant, l'eau principalement, est construite et englobe le système étudié (Figure 34). La taille de la boîte de solvant est variable et doit être adaptée à la taille du système. Des contre-ions (Na^+ , K^+ , Cl^- , etc.) peuvent être ajoutés soit pour neutraliser la charge totale du milieu, soit pour introduire une concentration donnée en espèces conductrices. Il existe différents modèles de champs de force permettant de décrire les molécules d'eau (TIP3P, TIP4P,^[275] TIP5P,^[276] SPC,^[277] etc.), qui se distinguent par leurs différentes approches pour traiter les interactions avec le solvant. Durant notre thèse, nous avons sélectionné le modèle TIP3P, qui est un modèle semi-empirique et qui considère la molécule d'eau comme un système à 3 sites. Il est très utilisé en MD et fournit un bon compromis entre la simulation du comportement des molécules d'eau et le temps de calcul.

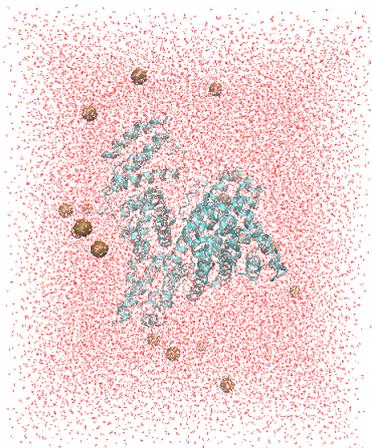


FIGURE 34 – Illustration de la protéine ASH dans une boîte d'eau contenant des contre-ions Na^+ .

Au cours d'une simulation de MD, il est fréquent que le système se déplace au sein de la boîte de solvant et qu'une partie se retrouve à la surface ou en dehors de cette boîte, entourée de vide. Dans ce cas, la totalité du système n'est plus solvatée et le calcul des équations de mouvements est biaisé. Pour palier ces effets de bords, des conditions périodiques aux limites sont mises en places : le système est considéré comme une maille unitaire et une reproduction de celle-ci est effectuée dans les trois dimensions de l'espace. Tous les déplacements moléculaires effectués au sein de la boîte centrale sont imités dans les répliques de celle-ci, ce qui permet d'obtenir un nombre constant d'atomes dans chaque maille unitaire.

6.5 Traitement des interactions à longues distances

La reproduction des conditions périodiques au sein des simulations de MD permet de prendre en compte l'ensemble des interactions avec les autres atomes, présents soit dans la boîte centrale, soit dans une des répliques. Cependant, la mise en place de plusieurs approximations est nécessaire afin de réduire le temps de calcul.

6.5.1 L'image minimale

Le traitement des interactions à longues distances dans une MD est effectué seulement pour les atomes présents dans la boîte centrale (Figure 35). Suivant sa position dans cette dernière, un atome i peut, soit interagir avec un atome j présent dans la boîte centrale, soit avec l'image de l'atome j dans une boîte voisine. L'approximation de l'image minimale stipule que seule l'interaction avec l'atome j le plus proche est prise en compte, ce qui permet d'éviter de considérer deux fois l'interaction avec l'atome j . Par exemple, à la Figure 35, l'interaction entre l'atome i de la boîte centrale et l'atome j_2 est effectué en prenant en compte celui se trouvant dans la boîte de gauche car c'est l'atome j_2 le plus proche spatialement de l'atome i .

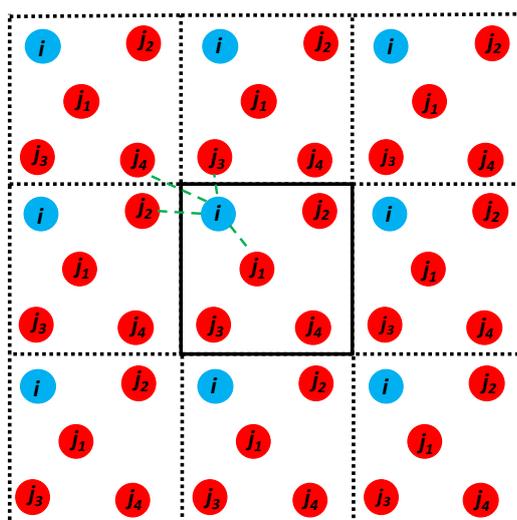


FIGURE 35 – Illustration du principe de l'image minimale. La boîte principale est représentée en trait plein et les interactions de l'atome i avec les atomes j_1 , j_2 , j_3 et j_4 sont représentés en traits discontinus vert.

6.5.2 Troncature des interactions

L'approximation de l'image minimale indique que les interactions de l'atome i avec l'ensemble des autres particules sont considérées. En réalité, pour des systèmes macromoléculaires comme le notre ($> 100\,000$ atomes), la prise en compte de l'ensemble des interactions est très gourmande en temps de calcul et n'a pas forcément un intérêt chimique ou physique. Pour cela, une distance de *cut-off* ($d_{cut-off}$) est appliquée afin de réduire le nombre d'interactions entre atomes à calculer et

les interactions présentes au-delà de cette distance sont ignorées. Le code de calcul que nous avons utilisé pour nos simulations de MD utilise un système de double *cut-off* pour traiter les interactions de van der Waals et électrostatique (Figure 36). L'ensemble des interactions entre les atomes i et j où $r_{ij} < d_{cut-off}$ est traité à chaque pas d'intégration de la simulation de MD à partir de leurs positions instantanées mais si la distance entre les atomes i et j est dans la condition $d_{cut-off} < r_{ij} < d_{pairlist}$, alors cette interaction n'est pas traitée comme dans le cas précédent. En effet, elle est calculée à partir d'une position fixe qui est sauvegardée dans une liste. Cette dernière est actualisée à des intervalles de pas d'intégration définis. Dans nos simulations, nous avons fixé la distance de *cut-off* et de *pairlist* respectivement à 12 et 14 Å, avec un intervalle de 10 pas d'intégration pour la sauvegarde de la liste d'atomes.

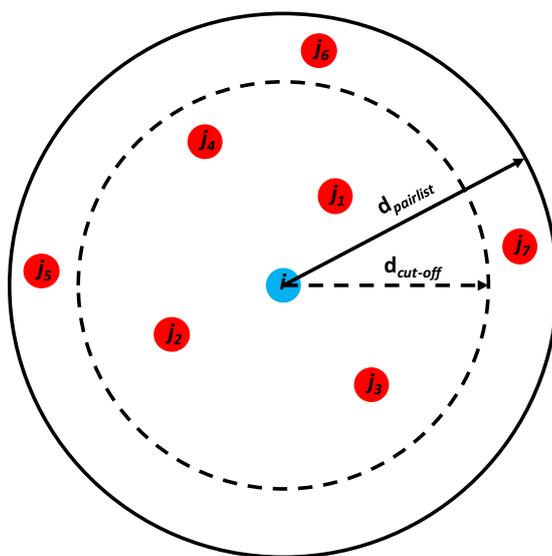


FIGURE 36 – Illustration du principe du double *cut-off*. Les interactions de l'atome i avec les atomes j_1, j_2, j_3 et j_4 sont traitées à chaque pas d'intégration, alors que celles avec les atomes j_5, j_6 et j_7 sont stockées dans une liste.

6.6 Calculs MM/GBSA

La méthode MM/GBSA (*Molecular Mechanics/Generalised Born Surface Area*) permet de calculer une affinité de liaison relative dans le cas de complexe composé de plusieurs unités, comme l'étude des interactions protéine-ligand. L'avantage des calculs MM/GBSA est qu'ils permettent d'évaluer les affinités relatives d'un jeu de molécules au sein d'une protéine avec une certaine précision pour un cout calculatoire très intéressant. En effet, l'énergie libre d'un complexe protéine-ligand (PL) s'écrit comme :

$$\Delta G = G_{PL} - G_P - G_L, \quad (6.29)$$

où G_{PL} est l'énergie libre du complexe PL, G_P , l'énergie libre de la protéine seule (P) et, G_L , l'énergie libre du ligand seul (L). L'énergie libre de chaque système X s'exprime sous la forme :

$$G(X) = E_{MM} + G_{solv} - TS(X), \quad (6.30)$$

où E_{MM} est l'énergie totale issue de la mécanique moléculaire en phase gazeuse, G_{solv} , est une correction de l'énergie libre de solvation, T , la température du système et, $S(X)$, la partie entropique du système.

Le calcul de l'énergie libre de chaque système X doit se faire sur un ensemble de conformations d'équilibre, provenant d'une simulation de MD du complexe PL, comportant des molécules de solvant explicites. Pour améliorer la précision des calculs, il est nécessaire de générer un nombre suffisant de géométries par simulation (> 100). Il existe des protocoles différents : le premier est de générer une seule mais longue simulation de MD du complexe (> 20 ns) ; le deuxième est d'effectuer trois courtes simulation de MD de plusieurs ns. Il a été montré que répliquer les simulations donne de meilleurs résultats, mais avec un coût calculatoire plus grand.^[278] Dans notre thèse, nous avons choisi d'effectuer le premier protocole avec des simulations de plus de 75 ns pour chaque site de la ASH exploré. À partir de la simulation de MD générée, il faut extraire les différentes conformations des trois systèmes (complexe PL, P et L) sans molécules de solvants explicites pour pouvoir calculer les termes G_{PL} , G_P et G_L . L'énergie E_{MM} est composée des mêmes termes que dans l'équation (6.2) (page 73) et est calculée grâce aux paramètres présents dans le champ de force. La correction de l'énergie libre de solvation est calculée à l'aide d'un continuum, expliquant ainsi le retrait des molécules de solvant explicites lors de l'extraction des différentes conformations des trois systèmes. Le terme G_{solv} s'écrit sous la forme :

$$G_{solv} = G_{GB} + G_{SASA} \quad (6.31)$$

où G_{GB} est un terme polaire provenant du modèle de Born généralisé, qui est une approximation du modèle de Poisson-Boltzmann pour la solvation, et, G_{SASA} , est un terme non-polaire représentant la zone de surface accessible au solvant (*Solvent Accessible Surface Area*). La partie entropique dans l'équation (6.30), $S(X)$, est accessible en effectuant une analyse de mode normaux ; par contre ce type de calcul est extrêmement couteux en temps de calculs, amenant généralement à ne pas considérer ce terme. Quelques études ont montré que le terme entropique n'apporte pas forcément une amélioration sur le calcul d'énergie libre et que le négliger donne de bons résultats.^[279–281]

7 Analyses des complexes ASH-DADS²⁻

De nombreuses structures de l'albumine sérique humaine (ASH) cristallisées avec différents ligands sont disponibles au sein de la PDB (Protein Data Bank). Ces structures peuvent servir de point de départ pour l'étude des interactions protéine-ligand ainsi que pour des simulations de MD classiques. Cependant, dans notre cas et, à notre connaissance, aucune structure d'un complexe de la ASH avec un stilbène disulfonique (SD) n'a été cristallisée à ce jour. Sans données expérimentales, nous n'avons aucune information sur la localisation du (ou des) site(s) d'accueil du (des) ligand(s) au sein de la protéine, ni sur les interactions spécifiques entre le(s) SD²⁻ et son environnement. Ceci nous a mené à employer la méthode de *docking* moléculaire pour générer une structure de départ probable du complexe ASH-SD²⁻, et ensuite de réaliser des simulations de MD.

7.1 Généralités

L'approche de *docking*¹ moléculaire permet de positionner un ligand au sein d'une protéine. Il s'agit de générer un ensemble de conformations du ligand par le biais de différents algorithmes. Une première zone (symbolisée par une boîte sur la Figure 37) déterminée par l'utilisateur, est sondée afin d'y placer le ligand. Si l'environnement présent dans la boîte principale ne permet pas de satisfaire les critères choisis alors une boîte secondaire, contenant la première, est employée pour élargir le choix des poses possibles du ligand au sein de la protéine. Les poses du ligand au sein de la protéine sont classées en utilisant des fonctions de scores, basées sur des formules mathématiques établies empiriquement à partir des énergies des interactions du système. Chaque code de *docking* moléculaire possède des fonctions de scores qui lui sont propres. Dans la plupart des cas, la mécanique moléculaire est employée, à travers un champ de force, pour évaluer les différentes composantes énergétiques du complexe et donc les interactions protéine-ligand. Une des utilisations courantes du *docking* moléculaire est le criblage d'une bibliothèque de molécules afin d'identifier les candidats potentiels pour une application biologique ciblée. En effet, la méthode est peu coûteuse en terme de temps de calcul, ce qui permet de traiter un grand nombre de candidats potentiel (jusqu'à des millions).

Trois grands types de *docking* existent : le *docking* rigide, le *docking* semi-flexible et le *do-*

1. Nous emploierons le terme anglais, plutôt que le terme français (« ancrage »)

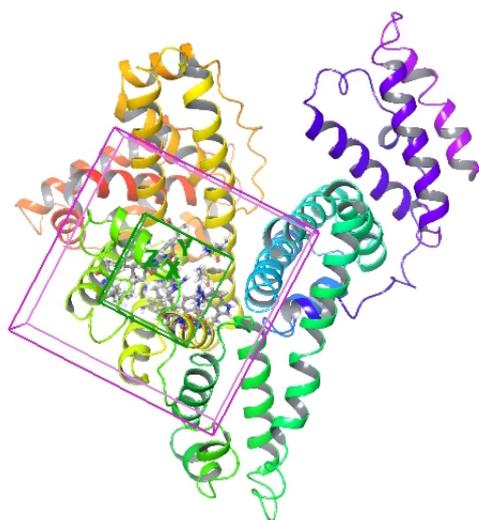


FIGURE 37 – Représentation des zones de la protéine, sondées par la boîte principale (vert) et la boîte secondaire (violet).

cking flexible. Le premier type considère la protéine et le ligand comme rigide, c'est-à-dire, que les deux partenaires ont une géométrie gelée et seules les translations et les rotations du ligand sont autorisées pour trouver la meilleure pose au sein de la protéine. Dans le second type, la protéine conserve toujours une géométrie fixe mais le ligand a plus de degré de liberté, les angles de valence et de torsion sont libres de mouvement, permettant au ligand d'adopter différentes conformations en fonction de son environnement. Enfin, le troisième type correspond à la relaxation totale, à la fois de la protéine et du ligand. Ce dernier s'adapte en fonction de structure initiale de la protéine puis c'est au tour de la protéine de s'ajuster à la nouvelle conformation du ligand pour favoriser les interactions déjà présentes et, en créer potentiellement des nouvelles. Finalement, le ligand se réadapte une nouvelle fois à son nouvel environnement. L'augmentation progressive de la flexibilité du ligand et de la protéine accroît évidemment les temps de calculs ainsi que la diversité des structures du complexe protéine-ligand.

Dans notre thèse de doctorat, nous avons opté pour des calculs de type semi-flexible car ils permettent d'obtenir un large ensemble de conformations des SD^{2-} au sein de la ASH avec un coût calculatoire raisonnable. Rappelons que l'utilisation de cette méthode a pour principal objectif d'obtenir une (ou plusieurs) structure(s) de départ. Le *docking* semi-flexible a été préféré au flexible car le temps de calculs est plus faible et que la relaxation de la protéine autour du ligand est effectuée lors des étapes de minimisation et d'équilibration, avant l'étape de production des simulations de MD.

7.2 Validation de la méthode de *docking* moléculaire

Avant de réaliser le *docking* moléculaire des SD^{2-} au sein de la ASH, nous avons voulu vérifier que le protocole mis en place permette bien d'obtenir des conformations de ligands valides. Pour cela, nous avons utilisé des structures déjà cristallisées de la ASH accueillant de petites molécules

aromatiques, comme les SD^{2-} . À partir d'un ensemble de conformères du ligand original, des calculs de *docking* moléculaire ont été réalisés au sein de son site d'accueil à l'aide du programme Glide, appartenant à la suite de logiciels Maestro version 10,1.^[282-284] La fonction de score *Glide Emodel (GE)* est celle sur qui nous nous sommes appuyés dans certains cas comme critère de sélection supplémentaire. Elle est recommandée lors de l'analyse de différents conformères d'une même molécule. L'objectif de cette partie est de confronter les conformations d'un ligand modèle obtenu par *docking* moléculaire avec sa structure cristallographique.

7.2.1 Sites étudiés

Dans ce travail, nous avons décidé de se concentrer sur quatre sites d'accueil potentiel au sein de la protéine ASH. Les deux premiers correspondent aux deux sites de Sudlow, localisé respectivement dans les sous-domaines IIA et IIIA. De nombreux ligands aromatiques de petites tailles ont été cristallisés avec la ASH. Nous avons choisi de partir de la structure cristallographique du complexe ASH-warfarine (PDB code : 2BXD) pour étudier le sous-domaine IIA et le complexe ASH-ibuprofène pour le sous-domaine IIIA (PDB code : 2BXG).^[31] Le troisième site est le sous-domaine IIAIIB, qui correspond au deuxième site d'accueil de l'ibuprofène au sein de la ASH. La structure cristalline pour ce site est la même que pour l'ibuprofène localisé dans le deuxième site de Sudlow (PDB code : 2BXG). Le quatrième site est le sous-domaine IB, qui est reconnu pour accueillir des molécules de grande taille comme la 4Z,15E-bilirubine-IX α (PDB code : 2VUE).^[55] La Figure 38 (droite) illustre l'ensemble des quatre sites que nous avons sondé. L'idée est de conserver seulement la structure de la protéine où un ligand est déjà présent dans un des sites afin d'étudier ce même site avec nos composés SD^{2-} . En effet, la conformation des résidus présents dans ces sites permet l'accueil de molécules, ce qui n'est pas le cas dans la structure de la ASH seule (PDB code : 1AO6).^[40]

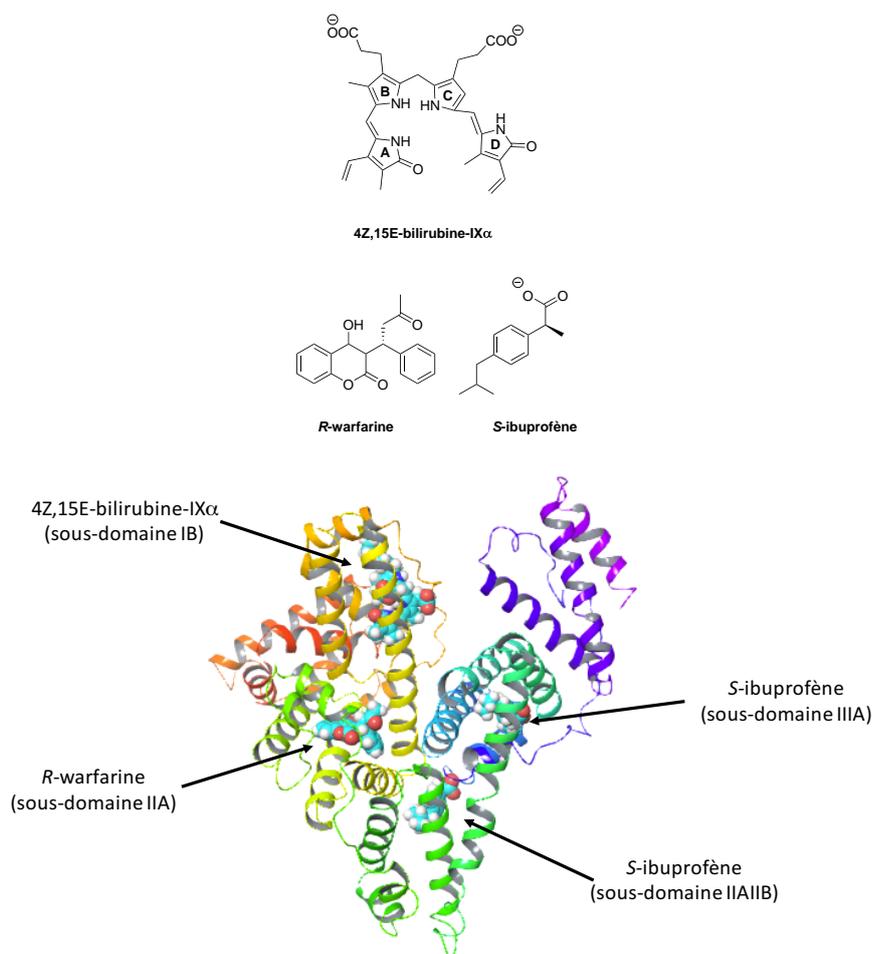


FIGURE 38 – Structure plane de la 4Z,15E-bilirubine-IX α , la warfarine et l’ibuprofène (gauche), et l’illustration de la ASH possédant ses quatre sites occupés (droite).

7.2.2 Préparation de la protéine

La géométrie de départ de la protéine, lors de l’étude d’un des sites, provient de la structure cristallographique qui lui est associée. Par exemple, lors des calculs pour le site IB, la géométrie initiale de la ASH est celle de la structure 2VUE. De même, pour le site IIA (IIAIIB et IIIA), la géométrie de la protéine utilisée est tirée de la structure cristallographique 2BXD (2BXG). Nous nous sommes assurés que les résidus chargés possèdent les mêmes états de protonation dans les trois structures utilisées. La séquence de la protéine ASH que nous avons utilisée est composée de 578 résidus (5-582). La protonation et l’addition des chaînes latérales sont effectuées à l’aide de l’outil *Protein Preparation Wizard* de la suite Maestro version 10.1. Des fonctions COO⁻ et NH₃⁺ ont été ajoutées respectivement sur les parties C et N terminales. 17 ponts disulfures ont été formés et la cystéine 34 a été laissée sous sa forme libre. Toutes les arginines et lysines (aspartates et glutamates) ont été (dé)protonées, en accord avec le logiciel PROPKA 3.0,^[285,286] à l’exception de la Lys199 (Glu244) qui est déprotonée (protonée). Concernant les histidines, seule His9 est protonée tandis que les autres se trouvent sous une forme neutre. Il existe deux tautomères pour les histidines neutres : une configuration en δ ou en ϵ (Figure 39). Nous avons attribué manuellement

la protonation de chaque histidine en fonction de leur environnement, dictée par la possibilité de créer une liaison hydrogène avec un hétéroatome de ses plus proches voisins. Au total, 10 résidus histidine se présentent sous la forme δ et 4 sous la forme ε . La protéine ASH possède une charge totale négative de $-14 e^-$. Lors de l'étape de protonation, une minimisation géométrique est effectuée afin de « favoriser » les réseaux de liaisons hydrogènes au sein de la protéine et de minimiser les « *clash* » stériques, avec un critère de convergence des RMSD des atomes lourds de 0,3 Å.

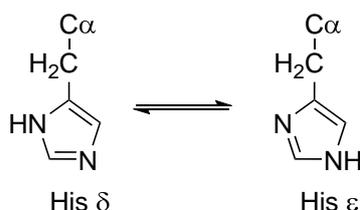


FIGURE 39 – Structure plane simplifiée des deux formes tautomères des résidus histidine neutres : histidine δ (gauche) et histidine ε (droite).

7.2.3 Protocole de *docking*

Après avoir préparé et conservé la géométrie de la protéine ASH provenant des trois structures cristallographique (2VUE pour l'étude du sous-domaine IB, 2BXD pour le sous-domaine IIA, et 2BXG pour les sous-domaines IIAIIB et IIIA), nous avons effectué des calculs de *docking* moléculaire de la bilirubine, de la warfarine et de l'ibuprofène, au sein de leur(s) site(s) respectif(s). À partir de la géométrie issue des données cristallographiques du ligand, nous avons généré, à l'aide du module ConfGen,^[287] un ensemble de conformères utilisés comme géométries de départ lors des calculs. Pour chaque site étudié, l'espace sondé pour l'insertion des ligands a été défini en suivant la procédure standard de Glide : la boîte principale mesure 10 Å³ et la boîte secondaire a une dimension de 30 Å³. Tous les calculs de *docking* moléculaire de type semi-flexible ont été effectués avec le mode XP (Extra Precision) de Glide. Le champ de force utilisé lors du *docking* moléculaire est OPLS-AA.^[288]

7.3 Validation du protocole de *docking*

Lors de la génération des conformères de chaque ligand, une attention particulière a été portée sur leur stéréochimie. En effet, la bilirubine qui a été cristallisée avec la ASH dans la structure 2VUE est la 4Z,15E-bilirubine-IX α ; la warfarine se présente sous sa forme isomérique *R* dans la structure 2BXD et l'ibuprofène sous sa forme *S* dans la structure 2BXG.

7.3.1 Bilirubine (sous-domaine IB)

Un jeu de 27 conformères a été généré à partir de la structure cristallographique de la 4Z,15E-bilirubine-IX α . L'écart énergétique relatif entre toutes ces formes est de 53 kcal.mol⁻¹. Ceci peut être considéré comme un intervalle très grand mais la grande taille de la molécule ainsi que son nombre de degrés de liberté mènent à l'obtention de conformères très variés géométriquement et énergétiquement. Au total, 90 poses de la 4Z,15E-bilirubine ont été produites au sein du sous-domaine IB et peuvent être séparées en 11 sous-groupes de poses, *via* le module *clustering of conformers*. Le critère de cette répartition est basé sur le RMSD (*Root Mean Square Deviation*) relatif entre l'ensemble des poses. La Figure 40 représente la matrice des RMSD entre toutes les poses obtenues lors des calculs de *docking* moléculaire. Comparé à la structure cristallographique de la 4Z,15E-bilirubine-IX α , l'ensemble des poses a une déviation géométrique comprise entre 0,57 et 9 Å. Nous observons qu'un groupe, constitué de 43 poses, se distingue des autres. Il regroupe toutes les poses qui se rapproche le plus de la géométrie de référence (RMSD de 0,57 à 1,7 Å) tandis que les autres groupes sont constitués de conformations de la 4Z,15E-bilirubine-IX α différentes (RMSD > 2 Å). La superposition de la structure cristalline de la 4Z,15E-bilirubine-IX α et de la pose possédant le RMSD le plus bas est présentée sur la Figure 41. L'énergie *GE* de cette dernière est de -117,9 kJ.mol⁻¹ et fait partie des trois poses les plus stables sur les 90 poses.

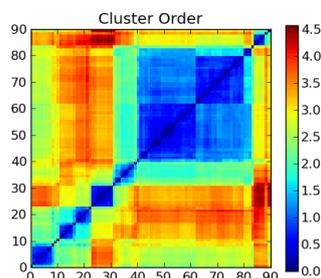


FIGURE 40 – Matrice des RMSD (Å) entre les différentes poses de la 4Z,15E-bilirubine-IX α au sein du sous-domaine IB de la ASH, obtenue par le module *clustering of conformers*, suite au *docking* moléculaire.

L'orientation du pyrrole A diffère légèrement entre la structure cristallographique et la pose sélectionnée, tandis que les autres pyrroles et les deux fonctions propionates se superposent très bien. Les principales interactions avec le sous-domaine IB sont présentes, la fonction propionate du pyrrole B crée deux liaisons hydrogènes avec l'Arg117, une provenant de l'amine en position ϵ (2,73 Å) et l'autre d'une des deux amines terminales (2,84 Å); la deuxième fonction propionate, localisée sur le pyrrole C, crée une liaison hydrogène avec une amine terminale de l'Arg114 (2,71 Å). L'oxygène chargé négativement de cette même fonction établit un pont salin avec l'azote chargé positivement de l'Arg186 (3,86 Å) et la fonction carbonyle du pyrrole D présente une liaison hydrogène avec la Tyr138 (3,00 Å). Ces interactions sont également présentes dans la structure

cristallographique.

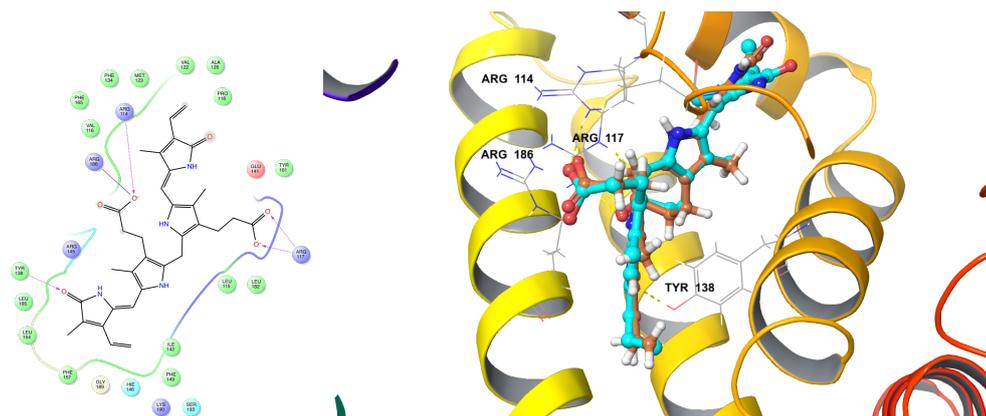


FIGURE 41 – Schématisation en 2D des principales interactions entre la 4Z,15E-bilirubine-IX α et les résidus du sous-domaine IB (gauche) et superposition entre la pose issue du *docking* moléculaire, en orange, et la structure cristalline de la 4Z,15E-bilirubine-IX α , en bleu (droite).

7.3.2 Warfarine (sous-domaine IIA)

Les calculs de la warfarine au sein du premier site de Sudlow ont été effectués sur un ensemble de 24 conformères de départ. 56 poses ont été générées, réparties en 5 groupes distincts. La matrice des RMSD de l'ensemble des poses (Figure 42) fait apparaître très clairement deux *clusters*.

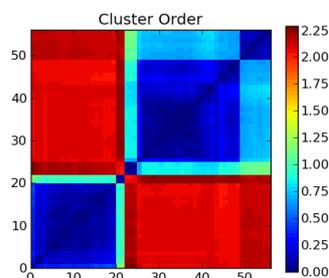


FIGURE 42 – Matrice des RMSD (Å) entre les différentes poses de la warfarine au sein du sous-domaine IIA de la ASH, obtenue par le module *clustering of conformers*, suite au *docking* moléculaire.

La premier groupe, composé d'une vingtaine de poses, rassemble les conformations les plus proches de la structure cristallographique de l'isomère *R* de la warfarine, tandis que la seconde famille, composé du reste des poses, présente des conformations possédant une orientation du motif coumarine différente au sein de ce site d'accueil, induisant logiquement des valeurs de RMSD élevées. Au sein de la première famille, l'écart avec la géométrie de référence est compris entre 0,65 et 0,85 Å. La majorité des poses présentent les mêmes interactions que celles identifiées dans

la structure cristallographique. La pose possédant le RMSD le plus bas (0,65 Å) à une énergie GE de $-68,1 \text{ kJ.mol}^{-1}$ et fait partie des 10 structures les plus stables. D'un côté de la molécule, une liaison hydrogène est créée entre le motif coumarine et l'His242 (3,55 Å) et de l'autre côté, deux liaisons hydrogène sont présentes (2,95 et 3,15 Å) avec Arg222 (Figure 43). La conformation trouvée par les calculs de *docking* moléculaire montre un très bon accord avec la structure cristalline de la warfarine.

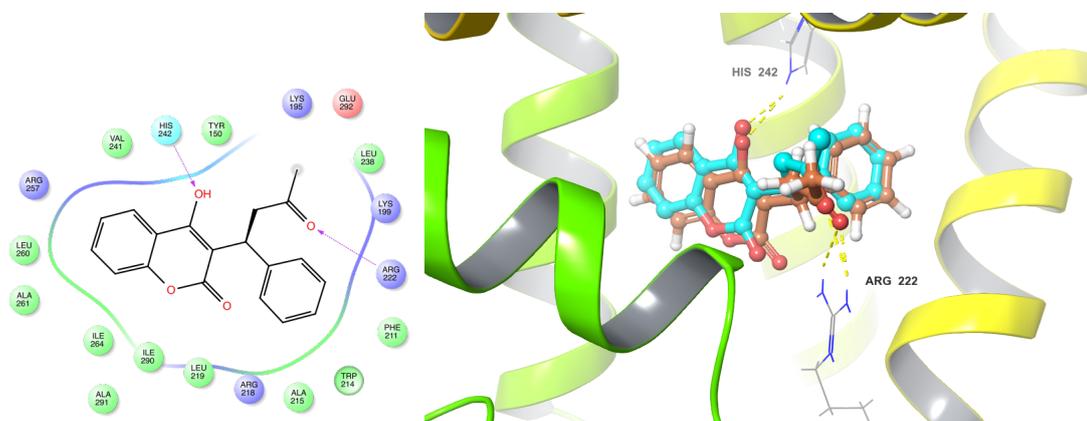


FIGURE 43 – Schématisation en 2D des principales interactions entre la warfarine et les résidus du sous-domaine IIA (gauche) et superposition entre la pose issue du *docking* moléculaire, en orange, et la structure cristalline de la warfarine, en bleu (droite).

7.3.3 Ibuprofène (sous-domaines IIAIIB et IIIA)

Un ensemble de 6 conformères de départ de la molécule d'ibuprofène a été utilisé pour le *docking* moléculaire au sein des sous-domaine IIAIIB et IIIA.

Sous-domaine IIAIIB

Le sous-domaine IIAIIB est le second site d'accueil pour une molécule d'ibuprofène. Celui-ci n'est pas considéré dans la littérature comme un site majeur d'accueil car sa position au sein de la protéine est très externe et exposée au solvant. 15 poses de l'ibuprofène ont été trouvées par les calculs. La matrice de RMSD ne présente pas de famille de conformation majeure, contrairement aux exemples précédents (Figure 44).

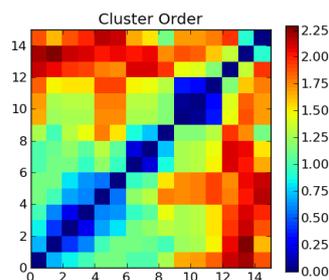


FIGURE 44 – Matrice des RMSD (Å) entre les différentes poses de l’ibuprofène au sein du sous-domaine IIAIB de la ASH, obtenue par le module *clustering of conformers*, suite au *docking* moléculaire.

9 poses sont très proches de la structure cristallographique (RMSD < 2 Å). La meilleur pose énergétique (-51,2 kJ.mol⁻¹) est aussi celle présentant le RMSD le plus faible avec la structure de référence (0,70 Å). Leur superposition est présentée en Figure 45. La fonction carboxylate de l’ibuprofène est le siège de toutes les interactions avec le sous-domaine IIAIB. L’oxygène chargé négativement forme un pont salin très fort avec la Lys351 (2,70 Å). De plus, le second oxygène de la fonction carboxylate est localisé de façon à pouvoir créer trois liaisons hydrogène différentes : une avec la fonction OH de la Ser480 (2,90 Å), une avec la fonction NH de la liaison peptidique de la Leu481 et une autre avec la fonction NH de la liaison peptidique de la Val482.

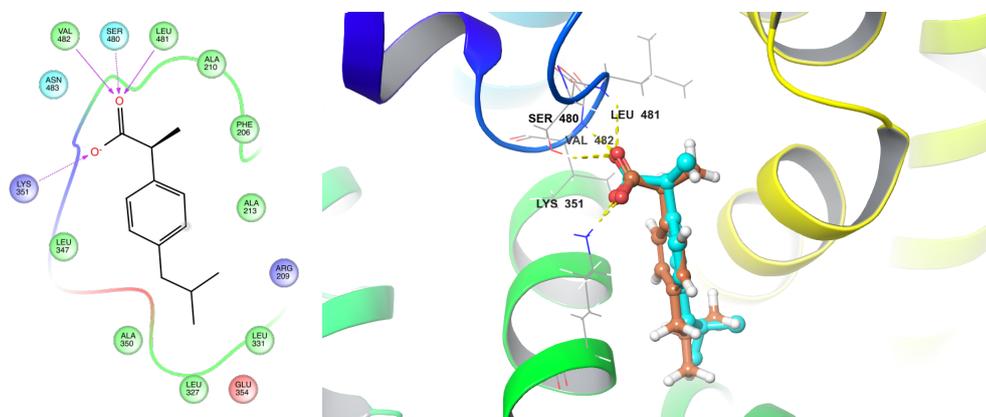


FIGURE 45 – Schématisation en 2D des principales interactions entre l’ibuprofène et les résidus du sous-domaine IIAIB (gauche) et superposition entre la pose issue du *docking* moléculaire, en orange, et la structure cristalline de l’ibuprofène, en bleu (droite).

Sous-domaine IIIA

Le sous-domaine IIIA est le site principal d’accueil pour la molécule d’ibuprofène. 13 conformations de départ ont été générées par les calculs de *docking* moléculaire. La superposition de toutes les poses montre clairement qu’elles se ressemblent fortement et que seule la disposition

du squelette carboné au sein de la poche hydrophobe les différencie (Figure 46, droite). De plus, elles ont toutes en commun les mêmes interactions avec les résidus polaires présents à l'entrée de la poche. La pose qui se rapproche le plus de la structure cristallographique de l'ibuprofène à un RMSD de 0,95 Å. De plus, elle est la structure la plus stable avec une énergie GE de $-54,8 \text{ kJ.mol}^{-1}$

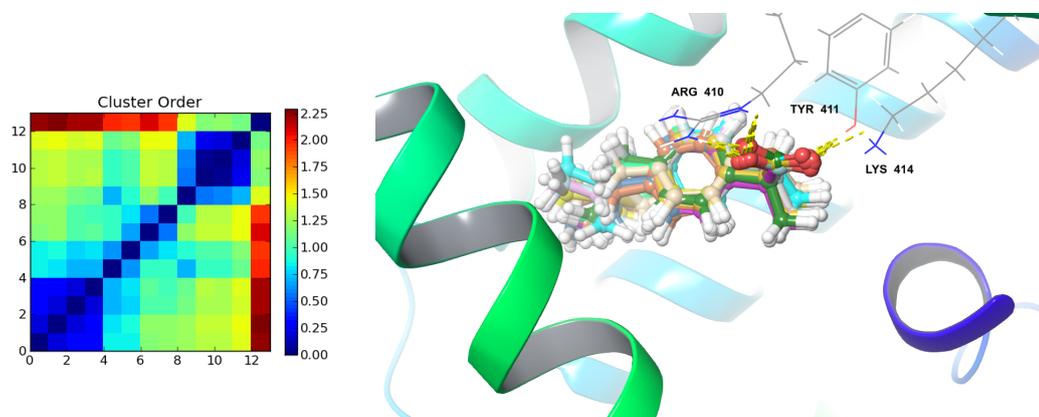


FIGURE 46 – Matrice des RMSD (Å) entre les différentes poses de l'ibuprofène au sein du sous-domaine IIIA de la ASH, obtenue par le module *clustering of conformers* (gauche) et superposition des 13 poses issues des calculs de *docking* moléculaire (droite), suite au *docking* moléculaire.

Comme dans le cas du sous-domaine IIAIIB, la fonction carboxylate crée des interactions avec les résidus polaires situés à l'entrée de la poche hydrophobe (Figure 47). Un pont salin est formé entre l'oxygène chargé négativement de l'ibuprofène et l'Arg410 (2,57 Å). Une liaison hydrogène est présente entre le second oxygène de la fonction carboxylate et la Tyr411 (3,03 Å). Il est à noter que la conformation issue des calculs n'interagit pas avec la Lys414 alors que la structure cristallographique est dans des bonnes conditions géométriques pour présenter une liaison hydrogène supplémentaire.

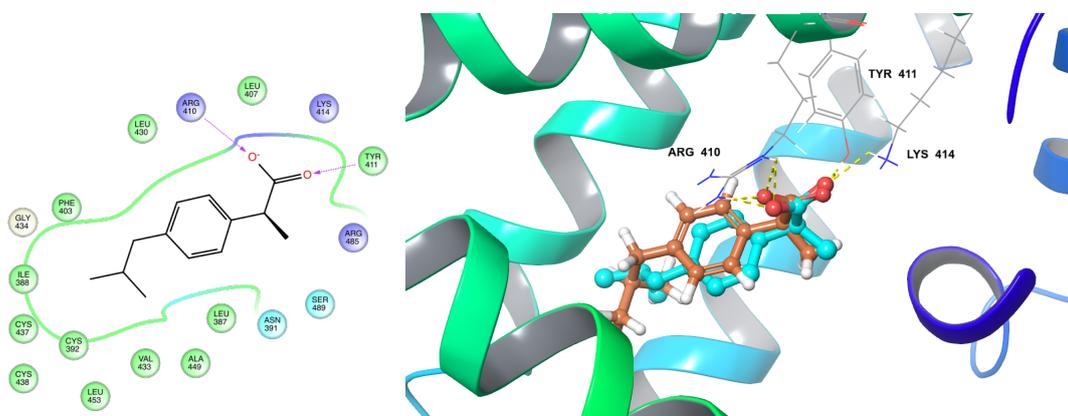


FIGURE 47 – Schématisation en 2D des principales interactions entre l'ibuprofène et les résidus du sous-domaine IIIA (gauche) et superposition entre la pose issue du *docking* moléculaire, en orange, et la structure cristalline de l'ibuprofène, en bleu (droite).

7.3.4 Bilan

À partir de quatre structures cristallographiques de complexe ASH-ligand, nous avons été capable de générer, avec les critères par défaut de Glide, différentes conformations de ligand qui sont très proches de celles observées expérimentalement. En effet, dans les quatre sites d'accueil que nous avons ciblés pour notre étude, nous avons obtenu des ensembles de poses du ligand possédant des RMSD faibles par rapport à la géométrie de référence. Celles se rapprochant le plus de la structure cristallographique du ligand présentent des RMSD inférieurs à 1 Å et font parties des poses ayant les meilleurs GE , ce qui est très concluant et font apparaître les mêmes interactions avec la protéine ASH. Une inspection visuelle est néanmoins nécessaire pour sélectionner la meilleure pose, en plus des critères géométriques (RMSD) et énergétiques (GE). Cette première analyse de *docking* moléculaire permet donc de valider notre protocole de calculs et d'avoir une certaine confiance vis-à-vis des résultats obtenus pour générer des structures ASH-ligand satisfaisantes. Dans la section suivante, nous avons utilisé le même protocole de *docking* moléculaire pour l'étude des interactions des différents SD^{2-} avec la protéine ASH.

7.4 Etude des interactions ASH-DADS²⁻

Nous allons décrire le comportement du DADS²⁻ au sein de la protéine ASH pour chacun des sites étudiés par *docking* moléculaire *via* un jeu de 25 conformères du DADS²⁻, générés à l'aide du module Confgen à partir de la structure optimisée en QM (M06-2X/6-31G(d)). Ensuite, des simulations de MD sont effectuées avec un protocole en trois étapes : la minimisation du système, l'équilibration et la production des trajectoires de MD. La première étape consiste à minimiser l'énergie potentielle du système entier avec 5 000 étapes de minimisation. La seconde étape est l'équilibration du système avec un ensemble NVT, à une température de 300 K pendant 100 ps. Nous nous sommes assuré qu'aucune bulle de vide n'apparaît pendant cette période. La deuxième partie de l'équilibration est réalisée avec un ensemble NPT, toujours à une température de 300 K, avec une pression de 1 atm (1,01325 bar) pendant 200 ps. La dernière étape est la production de la MD où le système évolue pendant 75 ns à une température de 300 K et à une pression de 1 atm (NPT). Toutes les simulations de MD ont été effectuées avec le code NAMD.^[289]

7.4.1 Sous-domaine IB

Docking moléculaire

Le calcul de *docking* moléculaire donne lieu à 53 poses différentes dans ce sous-domaine, se regroupant en 3 familles distinctes, respectivement composés de 27, 8 et 18 poses (Figure 48, gauche). Cependant, la superposition de toutes les poses fait apparaître seulement deux types de conformations majoritaire du DADS²⁻ au sein du sous-domaine IB (Figure 48, droite).

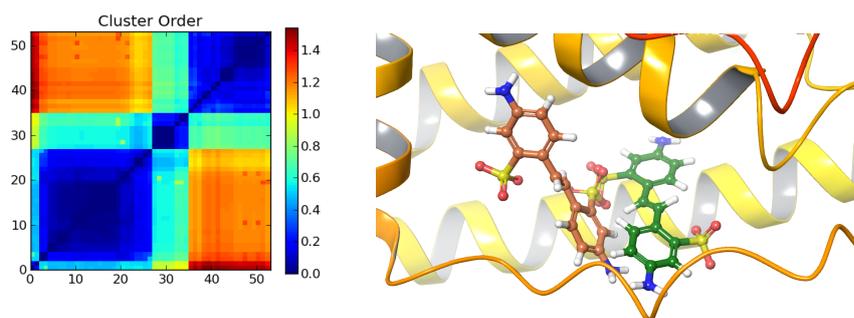


FIGURE 48 – Matrice des RMSD (Å) entre les différentes poses du DADS²⁻ au sein du sous-domaine IB de la ASH (gauche) et représentation des deux structures représentatives issues du *docking* moléculaire (droite).

La conformation en vert, à la Figure 48 (droite), est celle qui fut choisie comme la meilleure pose car elle représente la majorité des poses obtenues (35) et possède le plus d'interactions spécifiques avec la protéine. La pose que nous avons sélectionnée interagit avec quatre résidus différents (Figure 49). Une des fonctions amines terminales du DADS²⁻ crée une liaison hydrogène (2,96 Å) avec la fonction carbonyle de la Leu115. La fonction sulfonate située sur le même cycle aromatique est stabilisée par un pont salin avec la Lys190 (3,50 Å) et une liaison hydrogène avec l'His146 (2,66 Å). La deuxième fonction sulfonate accepte une liaison hydrogène avec l'Arg117 (2,48 Å). C'est cette conformation du DADS²⁻ qui sera utilisée comme point de départ pour les simulations de MD.

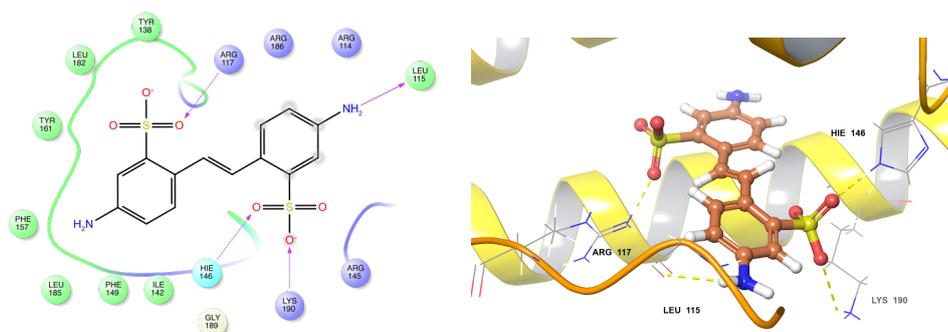


FIGURE 49 – Schématisation en 2D (gauche) et représentation en 3D (droite) des principales interactions entre la pose sélectionnée du DADS²⁻ et les résidus du sous-domaine IB, suite au *docking* moléculaire.

Simulation de dynamique moléculaire

Pour décrire le comportement du DADS²⁻ et les interactions présents avec la protéine, nous avons mis en place une nomenclature au niveau du ligand. En effet, nous distinguons séparément les deux moitiés symétriques de la molécule. La première (deuxième) moitié est composée de la fonction amine terminale porté par l'atome N₃₂ (N₃₅), du cycle aromatique A (B) et du groupement

sulfonate porté par l'atome S_{23} (S_{27}). De plus, deux angles dièdres de torsion, τ_1 et τ_2 , sont définis. τ_1 (τ_2) représente l'angle de rotation de la simple liaison situé entre la double liaison stilbène et le cycle aromatique de la première (deuxième) moitié de la molécule. La nomenclature générale du $DADS^{2-}$ est illustrée à la Figure 50.

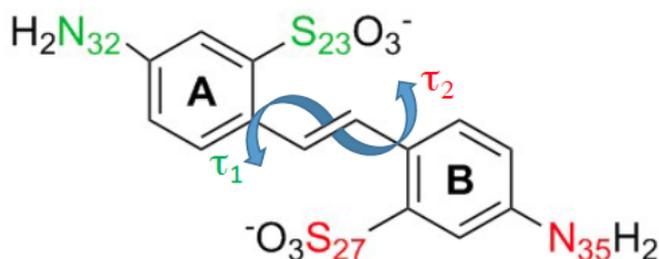


FIGURE 50 – Nomenclature de la structure du $DADS^{2-}$.

Le RMSD du squelette carboné de la protéine complexée présente quelques fluctuations comparé à celui de la ASH seule (Figure 51). En effet, le RMSD des carbones α de la ASH oscille autour de 2 à 3 Å, tandis que la présence du $DADS^{2-}$ dans le site IB augmente la déformation de la protéine (3-4 Å). Cette constatation n'est pas très surprenante car le domaine I est très flexible et la présence de ligand (ou d'acide gras) en son sein à un impact important sur la structure de la ASH, comme l'avait montré l'équipe de Curry *et al.* dans la référence 49 (*cf*: Chapitre 1).

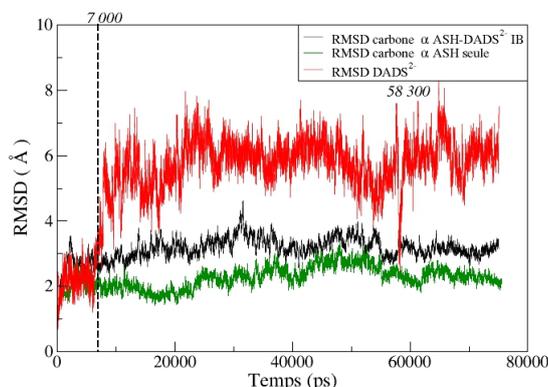


FIGURE 51 – RMSD des carbones α de la ASH complexée (noir), seule (vert) et du $DADS^{2-}$ (rouge) lors de la simulation de dynamique moléculaire au sein du site IB de la ASH.

Concernant le comportement du ligand, nous constatons un changement majeur après 7 ns de simulation (variation d'environ 4 Å sur le RMSD) et il semble évident que la conformation et/ou les interactions du $DADS^{2-}$ sont modifiées au sein du site IB de la ASH. Une autre fluctuation, plus courte et ponctuelle, est identifiée à 58,3 ns.

Pendant la période 0-7 ns, le $DADS^{2-}$ est en interaction avec 3 résidus polaires. Le premier résidu est la Tyr138, qui se trouve dans une orientation favorable pour former une liaison hydrogène avec l'amine terminale portée par l'atome N_{32} du $DADS^{2-}$ (Figure 52, gauche). Le second résidu

est la Lys190 ; elle se situe près du groupement sulfonate qui comprend l'atome S₂₇ et établit un pont salin avec ses oxygènes (Figure 52, droite).

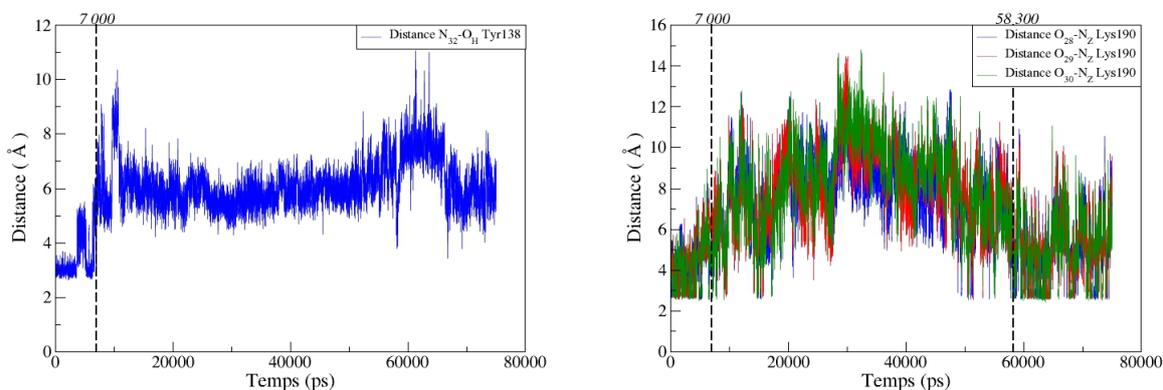


FIGURE 52 – Distances entre l'amine terminale N₃₅ du DADS²⁻ et la fonction OH de la Tyr138 (gauche), et entre les oxygènes du groupement sulfonate S₂₇ et l'ammonium terminal de la Lys190 (droite) lors de la simulation de dynamique moléculaire au sein du site IB de la ASH.

L'Arg186 est le troisième résidu en interaction avec le DADS²⁻ ; il est localisé à proximité de l'autre fonction sulfonate (atome S₂₃). Deux fonctions amine de ce résidu (une amine terminale et l'amine secondaire en position ε) forment un pont salin ainsi qu'une liaison hydrogène supplémentaire avec les trois oxygènes (Figure 53, gauche). À partir de 7 ns, l'Arg186 change de conformation et rompt son pont salin créé avec le DADS²⁻. Le déplacement de cette arginine permet au ligand de migrer plus profondément au sein du site. Les liaisons hydrogènes entre la Tyr138 et la Lys190 sont rompues et la nouvelle conformation adoptée par le DADS²⁻ permet à l'Arg117 de créer un pont salin avec le groupement sulfonate de l'atome S₂₃ (Figure 53, droite).

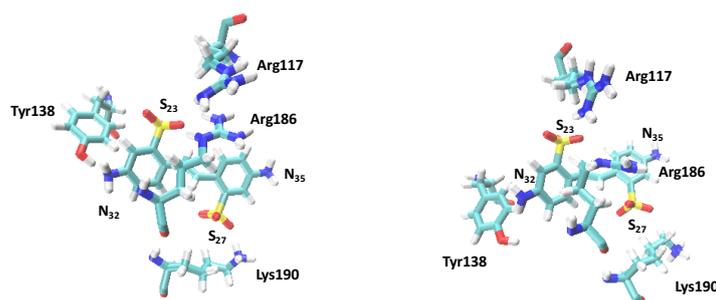


FIGURE 53 – Position du DADS²⁻ à 5 ns (gauche) et 7 ns (droite) lors de la simulation de dynamique moléculaire au sein du site IB de la ASH.

Vers 10 ns, deux arginines supplémentaires du site, l'Arg114 et l'Arg145, se déplacent pour entourer le DADS²⁻. La disposition des quatre arginines forme une cage autour du ligand et le stabilise grandement à travers des ponts salins et interactions π-cation. En effet, l'Arg145 forme un pont salin avec le sulfonate de l'atome S₂₇ et, à partir de 20 ns, sa disposition est favorable à

une interaction de type π -cation avec le cycle aromatique A (Figure 54, gauche). L'Arg114 ponte les deux groupements sulfonates *via* son amine en ϵ et une de ses amines terminales. Pendant une dizaine de nanoseconde, ce résidu est dans de bonnes conditions pour une interaction π -cation avec le cycle aromatique B (Figure 54, droite). Notons tout de même que les types d'interactions π (π -cation et π - π) sont mal décrites en mécanique moléculaire, notamment, avec un champ de force classique. La Figure 54 établit les intervalles de temps où les conditions géométriques pour ce type d'interactions sont remplies par les arginines autour du ligand.

La Figure 55 illustre la stabilisation du DADS^{2-} par la cage d'arginines au cours du temps. Durant la période de 10-55 ns, les interactions précédemment citées sont présentes et maintiennent parfaitement le ligand au sein du site IB de ASH.

À 50 ns, l'Arg114 s'éloigne du ligand, ce qui initie un mouvement de ce dernier et change, de nouveau, de conformation ainsi que de mode d'interactions avec l'environnement. En effet, l'Arg186 recrée un pont salin avec le groupement sulfonate du S_{23} comme au début de la simulation (0-7 ns). L'Arg 145 (117) est positionnée pour une interaction π -cation avec le cycle aromatique A (B). Le déplacement du DADS^{2-} permet à la Lys190 de stabiliser le groupement sulfonate du S_{27} *via* un pont salin (Figure 56, gauche). À 65 ns, le DADS^{2-} adopte une nouvelle conformation où le pont salin avec l'Arg186 est une fois encore rompu mais cela permet à l'Arg114 de revenir près du sulfonate de l'atome S_{23} du ligand (Figure 56, droite).

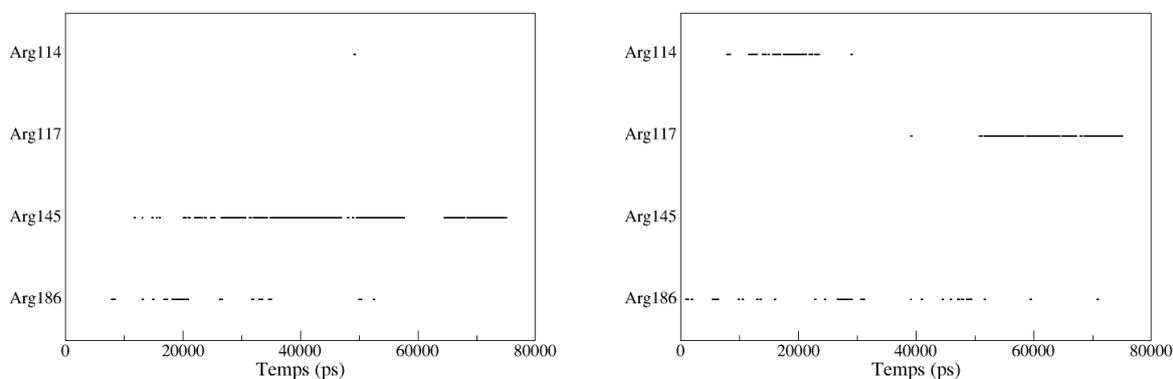


FIGURE 54 – Diagramme des interactions π -cation des arginines 114, 117, 145 et 186 avec les cycles aromatiques A (gauche) et B (droite) du DADS^{2-} lors de la simulation de dynamique moléculaire au sein du site IB de la ASH.

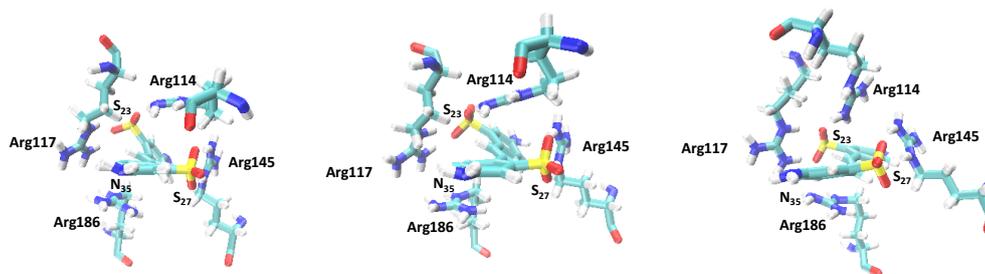


FIGURE 55 – Position du DADS²⁻ à 30 ns (gauche), 40 ns (centre) et 50 ns (droite) lors de la simulation de dynamique moléculaire au sein du site IB de la ASH.

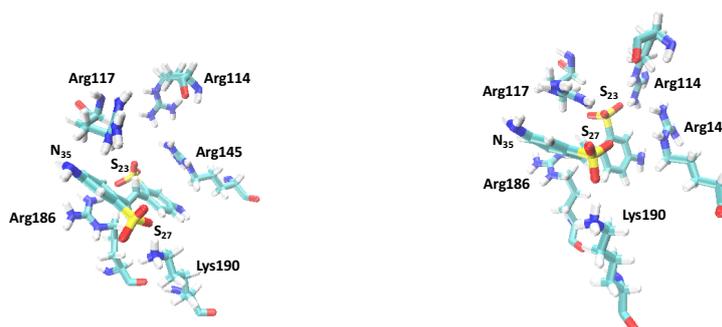


FIGURE 56 – Position du DADS²⁻ à 60 ns (gauche) et 70 ns (droite) lors de la simulation de dynamique moléculaire au sein du site IB de la ASH.

Pour conclure, le sous-domaine IB est très souple et large (site d'accueil de la 4Z,15E-bilirubine), ce qui laisse au DADS²⁻ la possibilité de se déplacer. Les interactions protéine-ligand sont principalement de nature électrostatique, notamment, par la présence des quatre arginines autour du ligand.

7.4.2 Sous-domaine IIA

Docking moléculaire

62 poses ont été générées lors du *docking* moléculaire du DADS²⁻ au sein du premier site de Sudlow. La matrice des RMSD indique deux grandes familles de poses (Figure 57, gauche) se situant en dehors de la poche hydrophobe. Étonnement, une seule pose est localisée à l'intérieur de la poche, comme dans le cas de la warfarine (Figure 57, droite). La localisation de la majorité des poses à l'extérieur de la poche hydrophobe peut être expliquée par la forte polarité du DADS²⁻ et de la présence de résidus polaires supplémentaires au sein d'une hélice du domaine III, se trouvant à proximité de l'entrée de la poche et permettant une stabilisation des deux côtés du ligand.

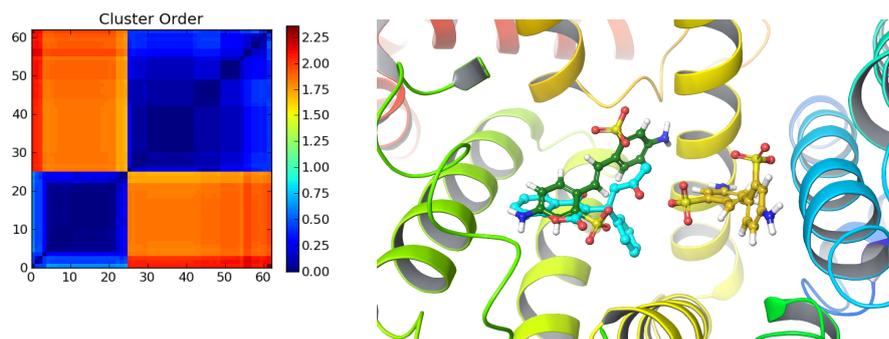


FIGURE 57 – Matrice des RMSD (\AA) entre les différentes poses du DADS^{2-} au sein du sous-domaine IIA de la ASH (gauche). Représentation de la pose du DADS^{2-} à l'intérieur (vert) et à l'extérieur (jaune) de la poche hydrophobe, ainsi que la structure cristalline de la warfarine (bleu) (droite).

Compte tenu de la spécificité du sous-domaine IIA et des résultats fournis par les calculs de *docking* moléculaire, nous avons décidé de réaliser deux simulations de MD. La première à partir d'une structure où le DADS^{2-} est à l'intérieur de la poche hydrophobe, et une deuxième où le ligand est localisé à l'extérieur de celle-ci. La structure de départ pour la simulation du DADS^{2-} dans la poche hydrophobe est présentée à la Figure 58, où un des groupement sulfonate est stabilisé par un pont salin avec l'Arg222 (2,72 \AA) et où une des fonctions amines terminales crée une liaison hydrogène avec le carbonyle de la Ser192 (2,74 \AA). La structure sélectionnée pour la simulation de MD en-dehors de la poche hydrophobe est présentée à la Figure 59. Cette pose fut retenue car elle est la structure représentative de nombreuses poses (37). La Lys436 crée un pont salin avec un premier groupement sulfonate (2,60 \AA) tandis que le deuxième groupement sulfonate est impliqué dans un réseau de ponts salins et liaisons hydrogènes avec la Lys195 (2,74 \AA), la Lys199 (2,93 \AA) et l'Arg218 (2,69 \AA). La fonction amine terminale et le cycle aromatique de la même partie de la molécule crée, respectivement, une liaison hydrogène avec la fonction carbonyle de la Lys195 (2,96 \AA) et une interaction de type π -stacking (*T-shaped*) avec le Trp214 (caractérisée par une distance de 4,81 \AA et un angle de 65°).

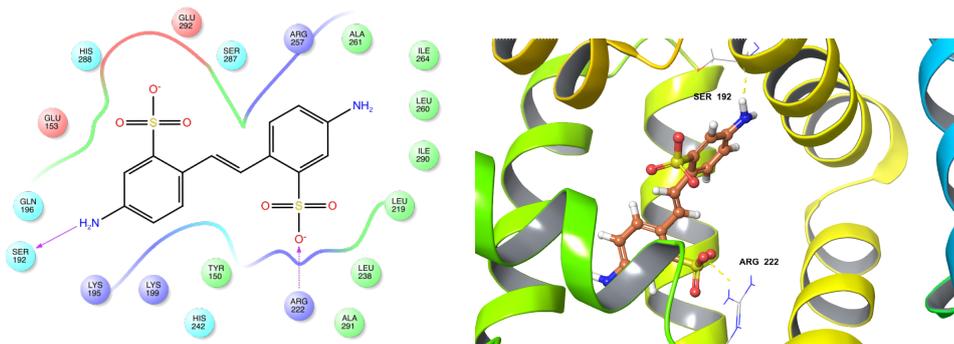


FIGURE 58 – Schématisation en 2D (gauche) et représentation en 3D (droite) des principales interactions entre la pose sélectionnée du DADS²⁻ à l'intérieur de la poche hydrophobe et les résidus du sous-domaine IIA de la ASH.

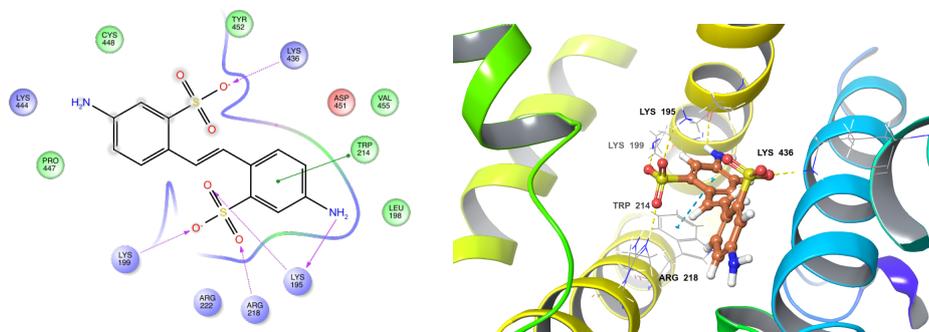


FIGURE 59 – Schématisation en 2D (gauche) et représentation en 3D (droite) des principales interactions entre la pose sélectionnée du DADS²⁻ à l'extérieur de la poche hydrophobe et les résidus du sous-domaine IIA de la ASH.

Simulations de dynamiques moléculaires

La première simulation de MD effectuée au sein du sous-domaine IIA présente la molécule de DADS²⁻ à l'intérieur de la poche hydrophobe. Les carbones α de la protéine montre des RMSD fluctuant, en moyenne, entre 2 et 3 Å (Figure 60). De légères augmentations du RMSD sont à signaler entre 10 et 20 ns, et en fin de simulation. En ce qui concerne le DADS²⁻, les valeurs de RMSD sont comprises dans le même intervalle, ce qui révèle que le ligand reste au sein de la poche hydrophobe et voit ses degrés de libertés (translation et rotation) limités par son environnement.

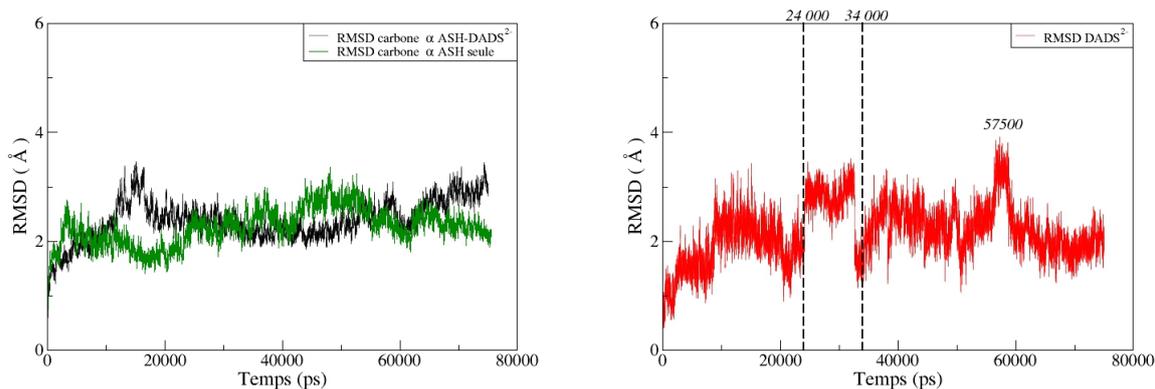


FIGURE 60 – RMSD des carbones α de la ASH complexée (noir) et seule (vert) (gauche), et celui du DADS²⁻ (rouge) (droite) lors de la simulation de dynamique moléculaire du DADS²⁻ situé à l'intérieur de la poche hydrophobe du site IIA de la ASH.

La position du DADS²⁻ au sein de la poche hydrophobe se situe près de l'entrée de la poche hydrophobe et du groupe de résidus polaires. Son orientation est présentée à la Figure 61.

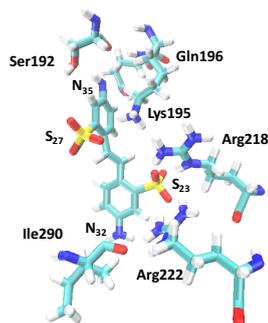


FIGURE 61 – Position initiale du DADS²⁻ lors de la simulation de dynamique moléculaire du DADS²⁻ situé à l'intérieur de la poche hydrophobe du site IIA de la ASH.

Lors des 24 premières nanosecondes, l'amine terminale, portée par l'atome N₃₅ du DADS²⁻, crée une liaison hydrogène avec le carbonyle de la Ser192 (Figure 62, gauche) et une deuxième avec le carbonyle terminal de la Gln196 (Figure 62, droite).

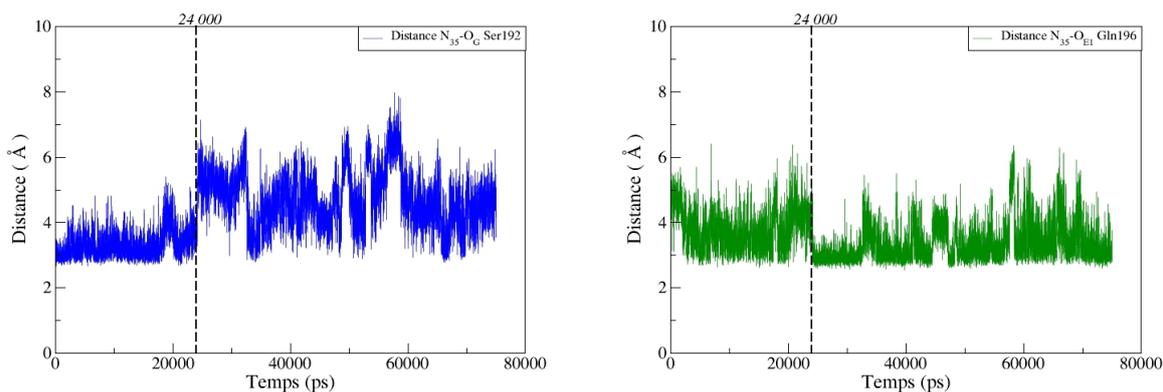


FIGURE 62 – Distances entre l’amine terminale N_{35} du $DADS^{2-}$ et la fonction OH de la Ser192 (gauche), et la fonction CO de la liaison peptidique de la Gln196 (droite) lors de la simulation de dynamique moléculaire du $DADS^{2-}$ situé à l’intérieur de la poche hydrophobe du site IIA de la ASH.

La fluctuation géométrique du ligand, observée vers 57,5 ns sur la Figure 60 (page 103), est provoquée par la rupture temporaire de la liaison hydrogène entre le Gln196 et le $DADS^{2-}$, entraînant une légère translation de ce dernier au sein de la poche hydrophobe. Le groupement sulfonate avec l’atome S_{23} (cycle aromatique A) est fortement stabilisé par les amines terminales de l’Arg218 et de l’Arg222 (Figure 63) *via* des ponts salins et liaisons hydrogènes.

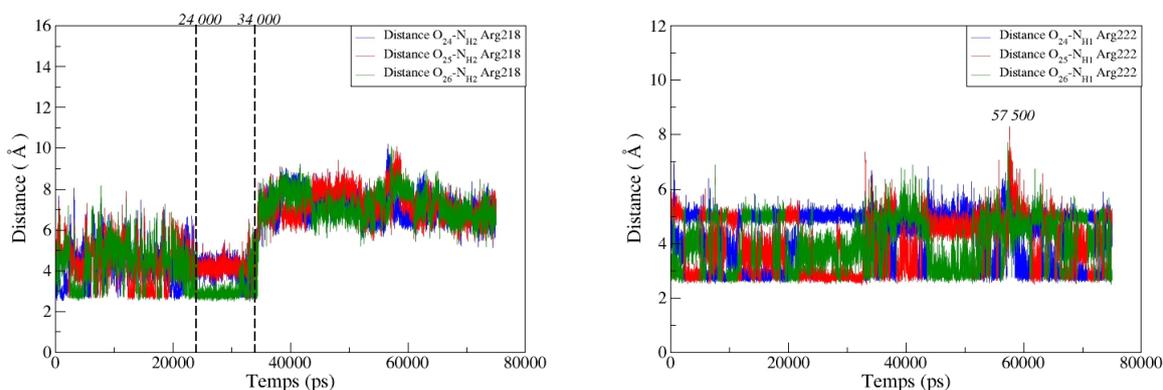


FIGURE 63 – Distances entre les oxygènes du groupement sulfonate S_{23} et l’amine terminale de l’Arg218 (gauche), et de l’Arg222 (droite) lors de la simulation de dynamique moléculaire du $DADS^{2-}$ situé à l’intérieur de la poche hydrophobe du site IIA de la ASH.

Dans la période de 24 à 34 ns, le ligand subit une variation importante au niveau de son RMSD (Figure 60, page 103), qui est la conséquence de quelques changements dans ses modes d’interactions avec les résidus de la poche hydrophobe. La liaison hydrogène avec le carbonyle de la Ser192 est rompue à 24 ns, ce qui permet à l’amine terminale de l’atome N_{35} de se rapprocher de

la Gln196. La Figure 64 illustre ce changement d'interaction ainsi que la déformation géométrique de la moitié supérieure (S_{27} et N_{35}) du $DADS^{2-}$.

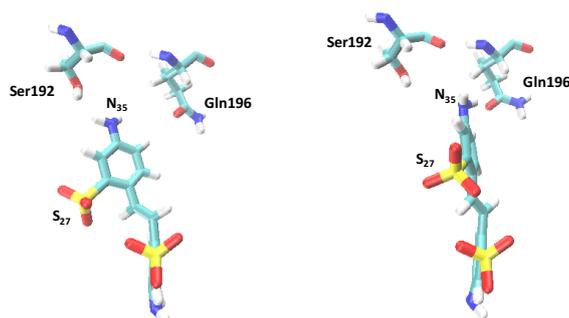


FIGURE 64 – Position du $DADS^{2-}$ à 22 ns (gauche) et 25 ns (droite) lors de la simulation de dynamique moléculaire du $DADS^{2-}$ situé à l'intérieur de la poche hydrophobe du site IIA de la ASH.

Dans le même laps de temps, l'Arg218 se décale pour se positionner entre les deux sulfonates et ainsi les ponter (Figure 65, gauche). Une des deux amines terminales est en interaction avec l'atome O_{28} tandis que l'amine primaire, en position ϵ , possède une liaison hydrogène avec le groupement sulfonate porté par l'atome S_{27} . La nouvelle conformation du ligand permet à la Lys195 de se rapprocher vers ce groupement sulfonate. Durant quelques nanosecondes, l'Arg218 et la Lys195 stabilisent la même fonction sulfonate (Figure 65, centre), jusqu'à la rupture de la liaison hydrogène de l'amine primaire de l'Arg218, qui entraîne un mouvement de ce résidu, qui devient alors libre (Figure 65, droite).

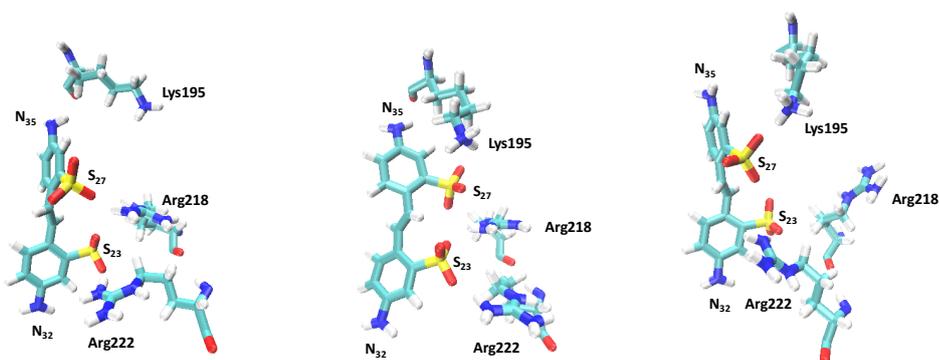


FIGURE 65 – Position du $DADS^{2-}$ à 24 ns, 32 ns (centre) et 35 ns (droite) lors de la simulation de dynamique moléculaire du $DADS^{2-}$ situé à l'intérieur de la poche hydrophobe du site IIA de la ASH.

Quelques nanosecondes après ce décrochage du ligand, l'Arg218 interagit fortement par pont salin avec l'Asp451, qui est situé au sein d'une hélice du domaine III, durant le reste de la simulation (Figure 66).

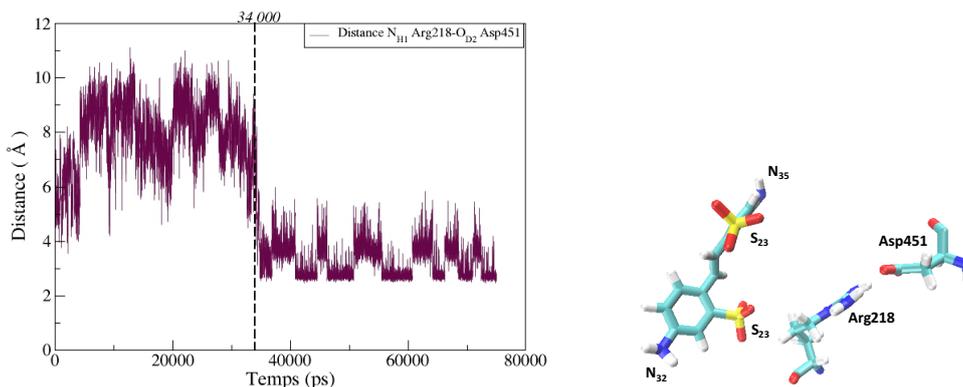


FIGURE 66 – Distances entre l’amine terminale de l’Arg218 et la fonction carboxylate de l’Asp451 (gauche) et illustration du pont salin créé entre l’Arg218 et l’Asp451 à partir de 34 ns (droite) lors de la simulation de dynamique moléculaire du DADS²⁻ situé à l’intérieur de la poche hydrophobe du site IIA de la ASH.

Après 34 ns, le mouvement du DADS²⁻, induit par les changements d’interactions, amène l’amine terminale portée par l’atome N₃₂ près du carbonyle de l’Ile290 pour créer une nouvelle liaison hydrogène (Figure 67, gauche). Avec cette dernière interaction, le ligand est totalement stabilisé au sein de la poche et garde cette conformation pendant le reste de la simulation (Figure 67, droite).

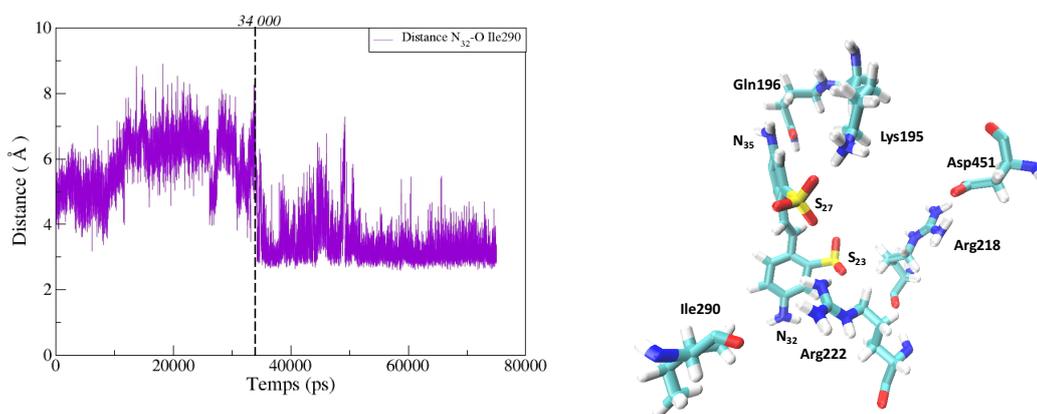


FIGURE 67 – Distances entre l’amine terminale N₃₂ du DADS²⁻ et la fonction carbonyle de l’Ile290 (gauche), et illustration de la conformation du ligand à partir de 34 ns (droite) lors de la simulation de dynamique moléculaire du DADS²⁻ situé à l’intérieur de la poche hydrophobe du site IIA de la ASH.

La deuxième simulation de MD dans ce site d’accueil a été effectuée en partant de la meilleure pose de *docking* où le DADS²⁻ se situe en dehors de la poche hydrophobe (Figure 59, page 102). Contrairement à la première simulation de MD, nous constatons que le ligand fluctue beaucoup plus en dehors de la poche hydrophobe (Figure 68).

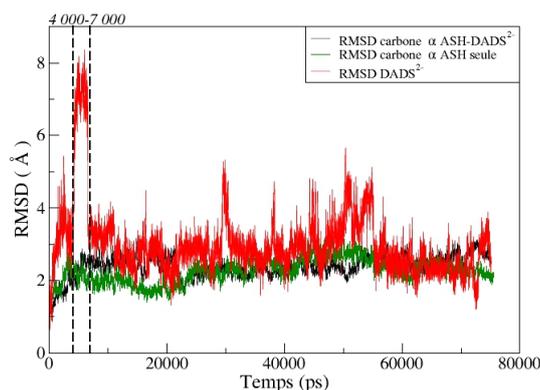


FIGURE 68 – RMSD des carbones α de la ASH complexée (noir), seule (vert) et du DADS^{2-} (rouge) lors de la simulation de dynamique moléculaire du DADS^{2-} situé à l'extérieur de la poche hydrophobe du site IIA de la ASH.

Initialement, la première moitié du DADS^{2-} (cycle aromatique A) est stabilisée par liaisons hydrogènes et ponts salins entre le groupement sulfonate S_{23} et la Lys195, la Lys199 et l'Arg218. La Figure 69 montre les distances entre ces trois résidus et les oxygènes de la fonction sulfonate S_{23} . Il est à remarquer que chacun des résidus interagit, à chaque instant, avec un oxygène différent lors de la période 0-55 ns. Un quatrième résidu, le Trp214, stabilise le cycle aromatique A par une interaction π - π de type *T-shaped*.

Les oxygènes du groupement sulfonate S_{27} sont en interaction avec la Lys195 et la Lys436 *via* des ponts salins et liaisons hydrogènes (Figure 70). Nous constatons donc que la Lys195 joue donc un rôle très important dans ce sous-domaine car elle pontre les deux fonctions sulfonates entre-elles.

Lors de cette simulation, deux événements majeurs ont lieu entre le ligand et son environnement. Le premier se déroule entre 4 et 7 ns où le DADS^{2-} montre une très forte variation géométrique ($\text{RMSD} \approx 8 \text{ \AA}$), qui se traduit par une réorientation momentanée de sa position au sein du sous-domaine IIA. La Figure 71 présente le repositionnement du ligand durant les 3 ns. Lors de ce mouvement, le DADS^{2-} reste en interaction avec les résidus à proximité, à l'exception de la Lys199 (Figure 69, haut et droite). Le décrochage de la Lys436 à partir de 7 ns avec le deuxième groupement sulfonate permet au ligand de revenir dans sa position initiale.

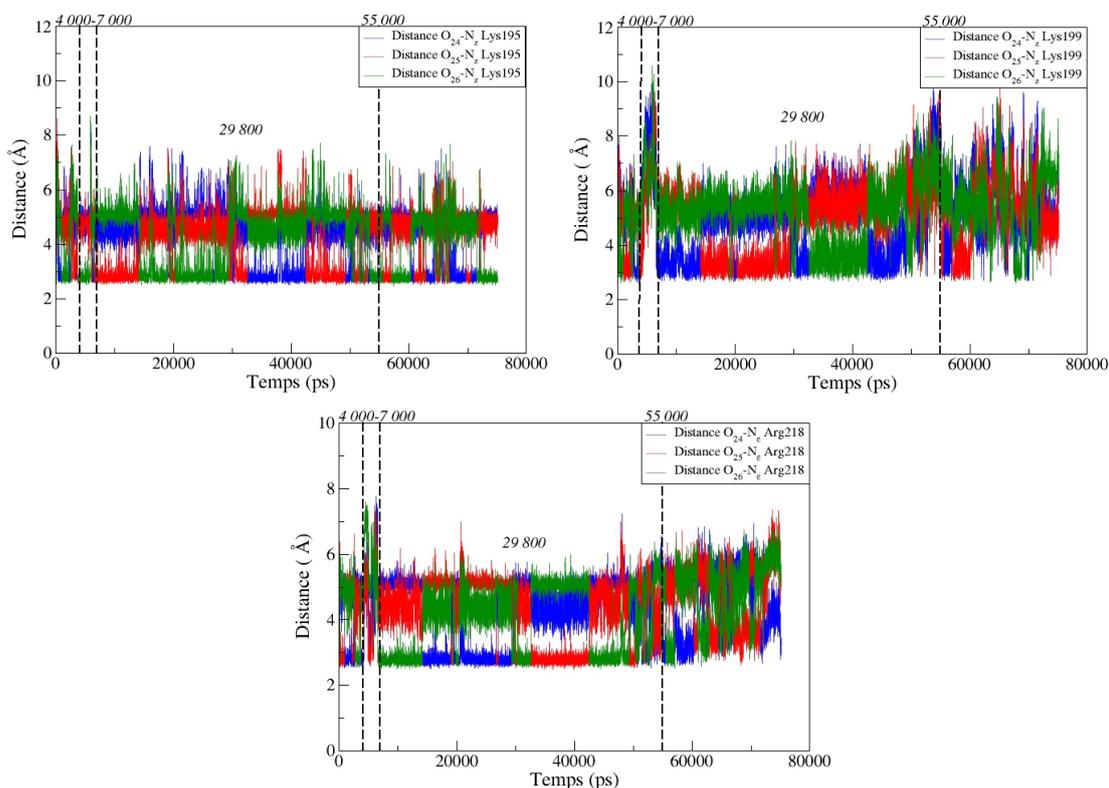


FIGURE 69 – Distances entre les oxygènes du groupement sulfonate S_{23} et l’ammonium terminal de la Lys195 (haut et gauche), l’amine terminale de la Lys199 (haut et droite) et l’amine terminale de l’Arg218 (bas) lors de la simulation de dynamique moléculaire du $DADS^{2-}$ situé à l’extérieur de la poche hydrophobe du site IIA de la ASH.

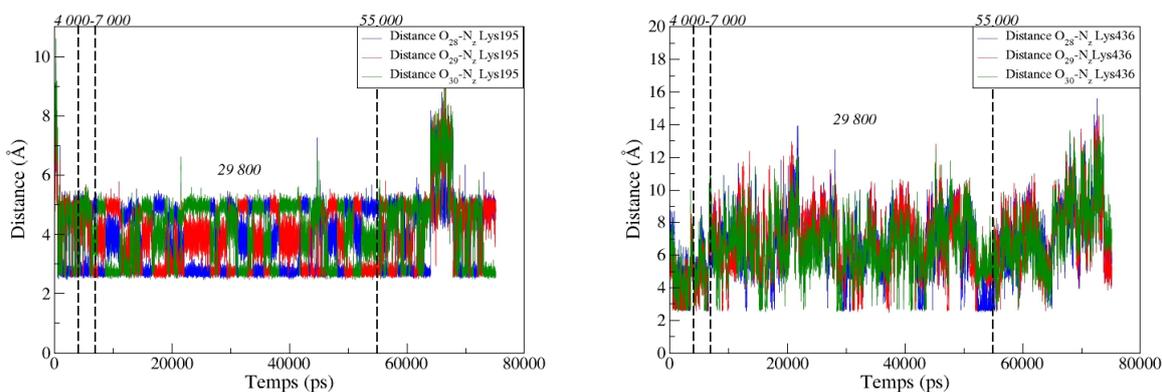


FIGURE 70 – Distances entre les oxygènes du groupement sulfonate S_{27} et l’ammonium terminal de la Lys195 (gauche) et l’ammonium terminal de la Lys436 (droite) lors de la simulation de dynamique moléculaire du $DADS^{2-}$ situé à l’extérieur de la poche hydrophobe du site IIA de la ASH.

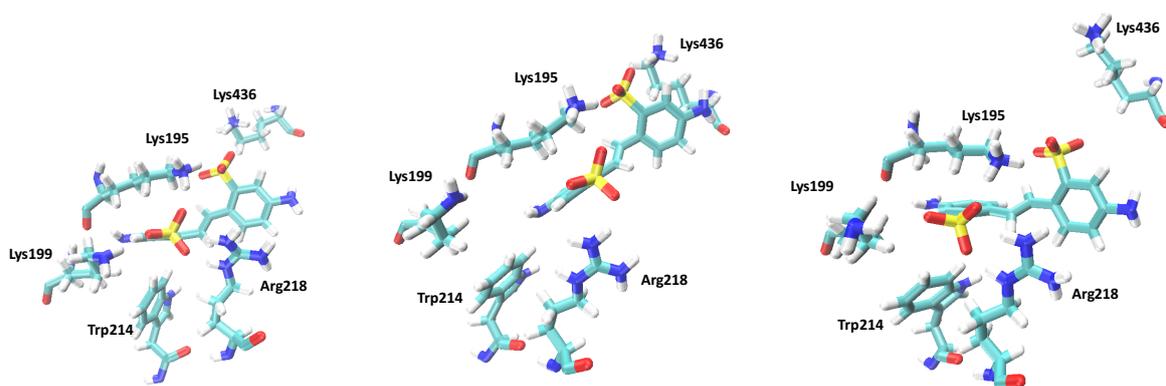


FIGURE 71 – Position du DADS²⁻ à 4 ns (gauche), 6 ns (centre) et 7 ns (droite) lors de la simulation de dynamique moléculaire du DADS²⁻ situé à l'extérieur de la poche hydrophobe du site IIA de la ASH.

Le second événement est localisé aux alentours des 55 ns où le RMSD du DADS²⁻ semble diminuer (Figure 68, page 107), indiquant que la molécule tend vers sa conformation initiale. La diminution du RMSD est induite par deux résidus, le Trp214 et l'Arg218. Comme indiqué précédemment, le Trp214 forme une interaction π - π de type *T-shaped* avec le cycle aromatique A du ligand. Cependant, vers 52 ns, l'Arg218 vient se placer en-dessous du cycle aromatique du DADS²⁻ pour créer une interaction π -cation et concurrencer alors l'interaction impliquant le Trp214. Les deux résidus se situent en-dessous du cycle aromatique jusqu'à 60 ns où le Trp214 va finalement changer de conformation et sortir du plan pour laisser la place à l'Arg218. Les RMSD du Trp214 et de l'Arg218 confirment bien les changements de conformation de ces résidus (Figure 72).

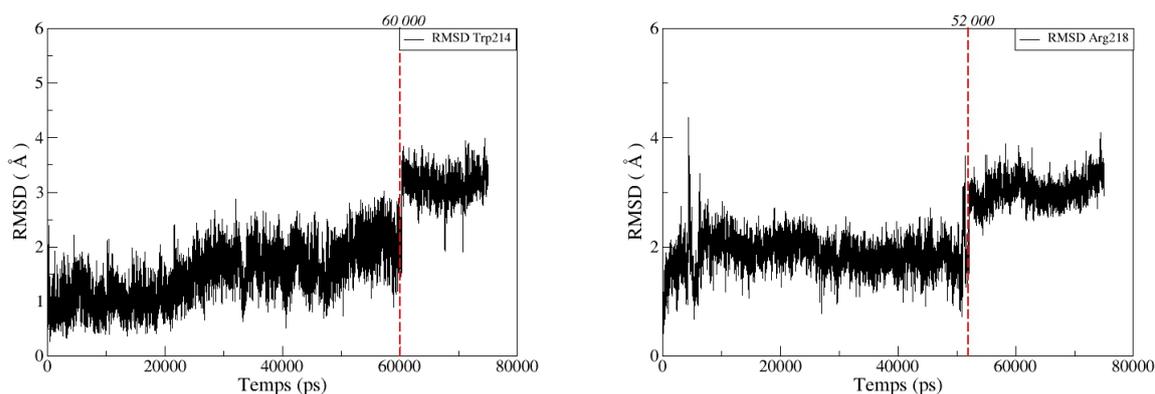


FIGURE 72 – RMSD des résidus Trp214 (gauche) et Arg218 (droite) lors de la simulation de dynamique moléculaire du DADS²⁻ situé à l'extérieur de la poche hydrophobe du site IIA de la ASH.

La Lys199 assiste le changement de conformation du Trp214 puisqu'elle se positionne de façon à bloquer ce résidu et l'empêcher de revenir en dessous du cycle aromatique par gêne stérique. La

Figure 73 illustre les différentes étapes du changement de conformation de la Lys199, le Trp214 et l'Arg218 en fonction du DADS²⁻.

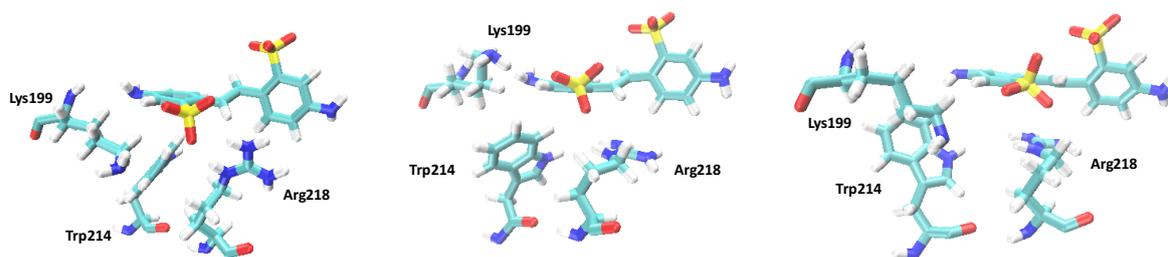


FIGURE 73 – Position du DADS²⁻ à 48 ns (gauche), 52 ns (centre) et 62 ns (droite) lors de la simulation de dynamique moléculaire du DADS²⁻ situé à l'extérieur de la poche hydrophobe du site IIA de la ASH.

Les deux simulations effectuées au sein du sous-domaine IIA indiquent que les interactions prédominantes sont de nature électrostatique, comme dans le sous-domaine IB. Nous constatons que le DADS²⁻ adopte des conformations similaires dans la poche hydrophobe, illustrant ainsi la contrainte stérique amenée par les résidus l'entourant, tandis qu'à l'extérieur de la poche, le ligand possède une plus grande relaxation géométrique.

7.4.3 Sous-domaine IIAIIB

Docking moléculaire

Les résultats du *docking* moléculaire du DADS²⁻ dans le sous-domaine IIAIIB génèrent 75 poses. Deux grands groupes de poses, divisés eux-mêmes en plusieurs sous-groupes, se distinguent comme le montre la matrice des RMSD des poses (Figure 74, gauche).

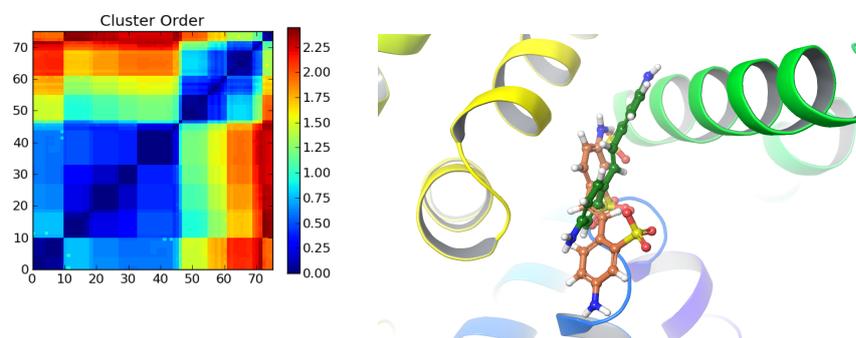


FIGURE 74 – Matrice des RMSD (Å) entre les différentes poses du DADS²⁻ au sein du sous-domaine IIAIIB de la ASH (gauche) et représentation des deux structures représentatives issues du *docking* moléculaire (droite).

Le DADS²⁻ présente peu d'interactions dans le sous-domaine IIAIIB car peu de résidus polaires y sont présents. Les deux critères que nous avons utilisés pour choisir la meilleure structure

ont été la fonction de score GE et le nombre d'interactions avec les résidus. La première sélection énergétique nous a permis de sélectionner deux structures du ligand (Figure 74, droite). La structure 1 « verte » a un GE de $-63,5 \text{ kJ.mol}^{-1}$ tandis que la structure 2 « orange » a un GE de $-61,9 \text{ kJ.mol}^{-1}$. Le nombre d'interactions guida notre sélection entre les deux poses. La conformation 2 possède deux interactions supplémentaires avec la protéine par rapport à la pose 1 qui semblent être, chimiquement, très intéressantes pour la stabilité du DADS^{2-} dans ce site d'accueil, ce qui nous amène à la privilégier. Les interactions qui caractérisent la pose 1 sont les suivantes : la Lys351 ponte les deux groupements sulfonates *via* des ponts salins (2,68 et 3,30 Å) ; les deux atomes d'oxygènes restant du deuxième sulfonate forment chacun une liaison hydrogène avec la fonction amine de la Leu481 (2,87 Å) et Val482 (2,78 Å) (Figure 75).

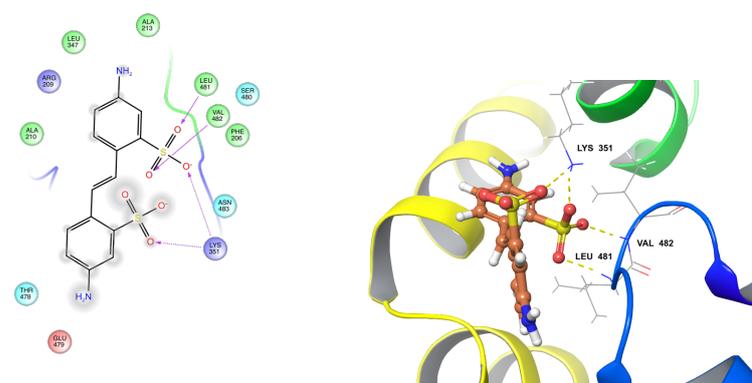


FIGURE 75 – Schématisation en 2D (gauche) et représentation en 3D (droite) des principales interactions entre la pose sélectionnée du DADS^{2-} et les résidus du sous-domaine IIAIIB de la ASH.

Simulation de dynamique moléculaire

La simulation du DADS^{2-} dans le sous-domaine IIAIIB indique que la protéine ne se déforme pas de manière significative lorsqu'une molécule de DADS^{2-} se situe dans cette région (Figure 76). En ce qui concerne le ligand, nous constatons un déplacement important entre 0 et 5 ns. Ensuite, lors du reste de la simulation, sa géométrie fluctue entre 3 et 6 Å. Ceci indique que le DADS^{2-} s'est probablement éloigné de la structure de départ et qu'il possède peu de contrainte géométrique.

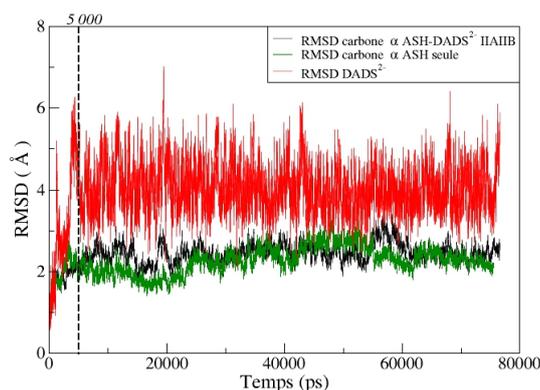


FIGURE 76 – RMSD des carbones α de la ASH complexée (noir), seule (vert) et du DADS²⁻ (rouge) lors de la simulation de dynamique moléculaire au sein du site IIAIB de la ASH.

La Leu481 et la Val482 établissent des liaisons hydrogènes avec le DADS²⁻ respectivement à hauteur de 41 et 40 % du temps de la simulation. Plus précisément, ces interactions ont lieu entre les oxygènes d'un même groupe sulfonate (S_{23}) et les fonctions amines de ces deux résidus. La Figure 77 illustre la distance entre les trois oxygènes du groupement sulfonate et les atomes d'azote, donneur de liaisons hydrogènes. Nous constatons qu'un seul oxygène à la fois est impliqué dans les deux liaisons hydrogènes potentielles et que la rotation du groupement sulfonate n'augmente pas la distance entre l'atome donneur et accepteur.

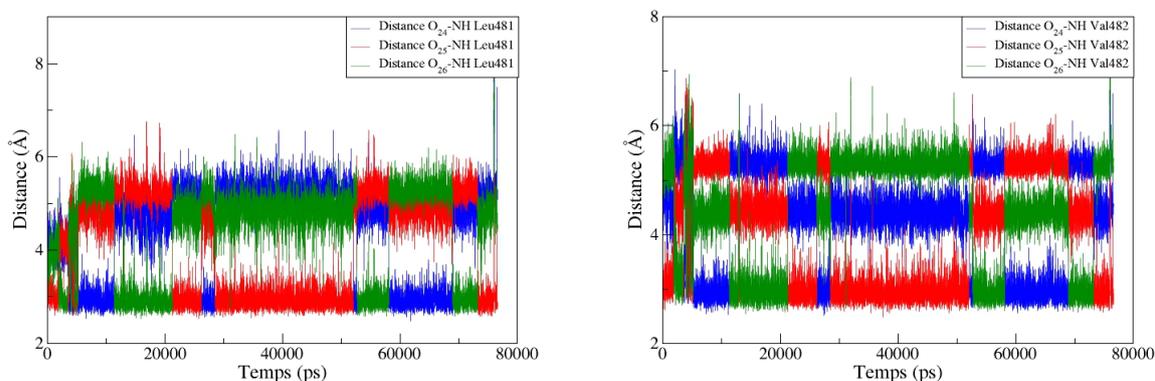


FIGURE 77 – Distances entre les trois oxygènes du groupement sulfonate S_{23} et la fonction NH des liaisons peptidiques de la Leu481 (gauche) et de la Val482 (droite) lors de la simulation de dynamique moléculaire au sein du site IIAIB de la ASH.

La structure issue du *docking* moléculaire montrait l'implication de la Lys351 pour stabiliser les deux sulfonates du DADS²⁻. Lors du début de la dynamique, la Lys351 est effectivement proche du ligand mais rapidement (après 1 ns), celle-ci s'éloigne pour créer un pont salin avec la Glu354. Vers 5 ns, le ligand se repositionne au sein du site d'accueil car la Thr478 le stabilise par liaison

hydrogène. La fonction NH₂ terminale du ligand vient se rapprocher de la fonction carbonyle de la Thr478 (Figure 78).

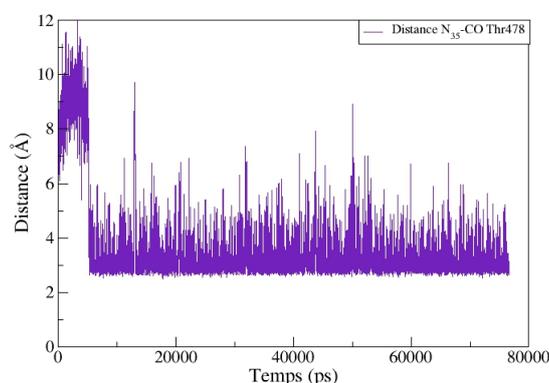


FIGURE 78 – Distance entre l'amine terminale N₃₅ du DADS²⁻ et la fonction CO de la liaison peptidique de la Thr478 lors de la simulation de dynamique moléculaire au sein du site IIAIIB de la ASH.

La Figure 79 illustre le changement de conformation du DADS²⁻ dans le sous-domaine IIAIIB qui s'effectue lors des cinq premières nanosecondes. Malgré une position qui est très exposée au solvant, le DADS²⁻ reste en interaction avec la protéine tout au long de la simulation et ne se décroche pas d'elle.

Au final, malgré la faible présence de résidus polaires et d'une position exposée au solvant, le sous domaine IIAIIB montre qu'il peut accepter une molécule de DADS²⁻.

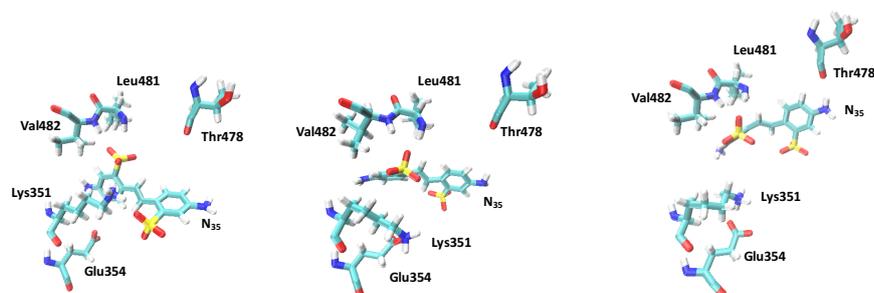


FIGURE 79 – Position initiale (gauche), à 4 ns (centre) et à 6 ns (droite) du DADS²⁻ lors de la simulation de dynamique moléculaire au sein du site IIAIIB de la ASH.

7.4.4 Sous-domaine IIIA

Docking moléculaire

Dans le deuxième site de Sudlow, 25 poses différentes ont été déterminées et la matrice RMSD montre 4 sous-groupes différents (Figure 80, gauche). Cependant, la différence entre toutes les

structures est très minime, avec un écart en RMSD inférieur à 0,1 Å et la superposition de l'ensemble des 25 poses confirme effectivement qu'elles sont toutes pratiquement identiques (Figure 80, droite). Il est évident, dans ce cas, que la petite taille de la poche hydrophobe contraint le DADS²⁻ à adopter une seule conformation.

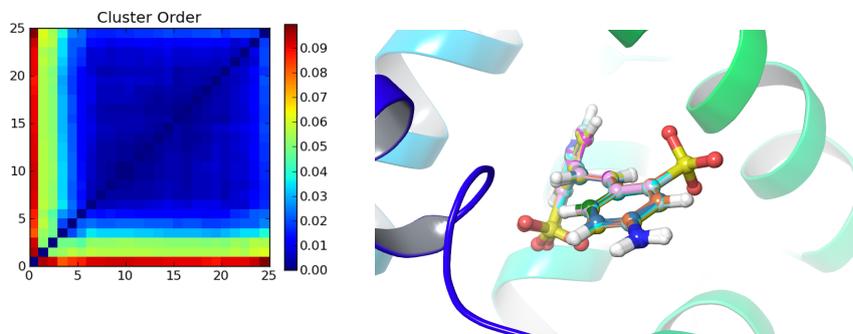


FIGURE 80 – Matrice des RMSD (Å) entre les différentes poses du DADS²⁻ au sein du sous-domaine IIIA de la ASH, obtenue par le module *clustering of conformers* (gauche) et superposition des 25 poses issues des calculs de *docking* moléculaire (droite).

Les interactions spécifiques du DADS²⁻ avec son environnement sont illustrées à la Figure 81. Le groupement sulfonate localisé le plus dans la poche est stabilisé par deux résidus : le premier est la Tyr411 avec une liaison hydrogène avec l'un des oxygènes (2,72Å), et le second est l'Arg410 avec un pont salin vers l'oxygène chargé négativement. L'Arg410 s'oriente de façon à constituer une interaction π -cation avec le cycle aromatique, positionné à l'entrée de la poche hydrophobe. Cette dernière interaction est caractérisée par une distance de 3,96 Å et un angle de 71,1 °. Une dernière liaison hydrogène existe entre l'amine de la chaîne latérale de l'Asn391 et le deuxième groupement sulfonate (2,75 Å).

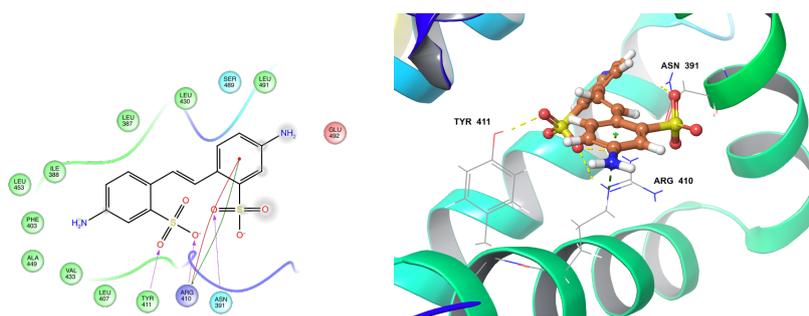


FIGURE 81 – Schématisation en 2D (gauche) et représentation en 3D (droite) des principales interactions entre la pose sélectionnée du DADS²⁻ et les résidus du sous-domaine IIIA de la ASH.

Simulation de dynamique moléculaire

Lors de la simulation de MD dans le sous-domaine IIIA, le DADS²⁻ présente plusieurs comportements. Comme l'indique la Figure 82, le déplacement du ligand peut être analysé séparément durant quatre périodes distinctes : 0-30, 30-44, 44-54 et 54-75 ns.

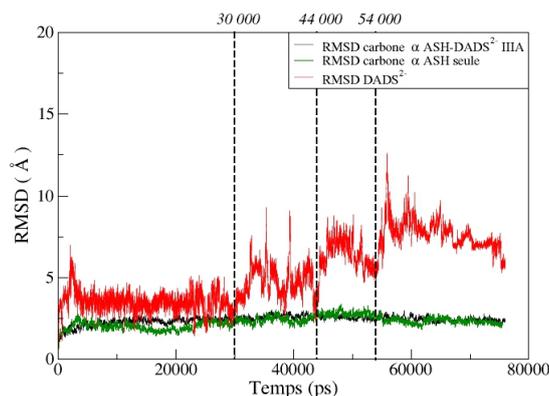


FIGURE 82 – RMSD des carbones α de la ASH complexée (noir), seule (vert) et du DADS²⁻ (rouge) lors de la simulation de dynamique moléculaire au sein du site IIIA de la ASH.

Durant les trente premières nanosecondes, une moitié du DADS²⁻ se situe au sein de la poche hydrophobe, tandis que l'autre se trouve à l'entrée et est donc plus exposée au solvant, comme initialement dans la pose issue du *docking*. De nombreuses interactions avec les résidus polaires présents permettent de maintenir le ligand. La Tyr411, la Lys414 et la Ser489 stabilisent très bien le groupement sulfonate porté par le soufre 23 (S₂₃) à travers un pont salin et plusieurs liaisons hydrogènes. L'Arg410 stabilise par un pont salin le deuxième groupement sulfonate, porté par le soufre 27 (S₂₇). Une interaction π -cation avec le cycle aromatique portant la deuxième fonction sulfonate peut aussi avoir lieu avec l'Arg410. La Figure 83 illustre les différentes interactions et conformations adoptées par le ligand et le site d'accueil à des intervalles de temps différents au cours de la simulation de MD. Remarquons que différents résidus interagissent de manière spécifique avec le ligand, comparé à la structure du *docking* moléculaire.

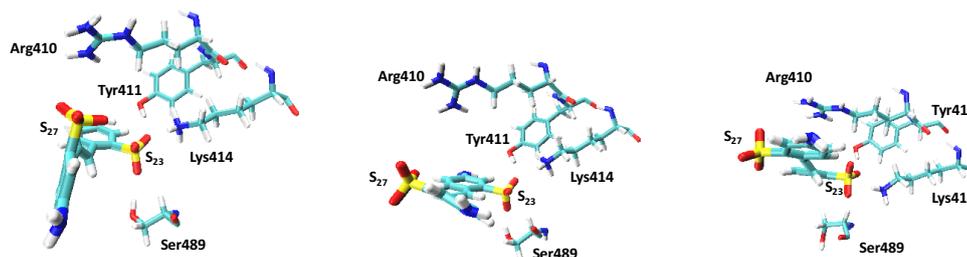


FIGURE 83 – Position à 10 ns (gauche), 20 ns (centre) et 27 ns (droite) du DADS²⁻ lors de la simulation de dynamique moléculaire au sein du site IIIA de la ASH.

Lors de la deuxième période (30-44 ns), le DADS²⁻ montre une grande variation au niveau de son RMSD. L'analyse des angles dièdres τ_1 et τ_2 du ligand (Figure 84) indique une variation vers 30 ns. En effet, une modification des valeurs moyennes des deux angles dièdres apparaît. Ils fluctuent autour de valeurs positives après 30 ns tandis que précédemment elles étaient négatives.

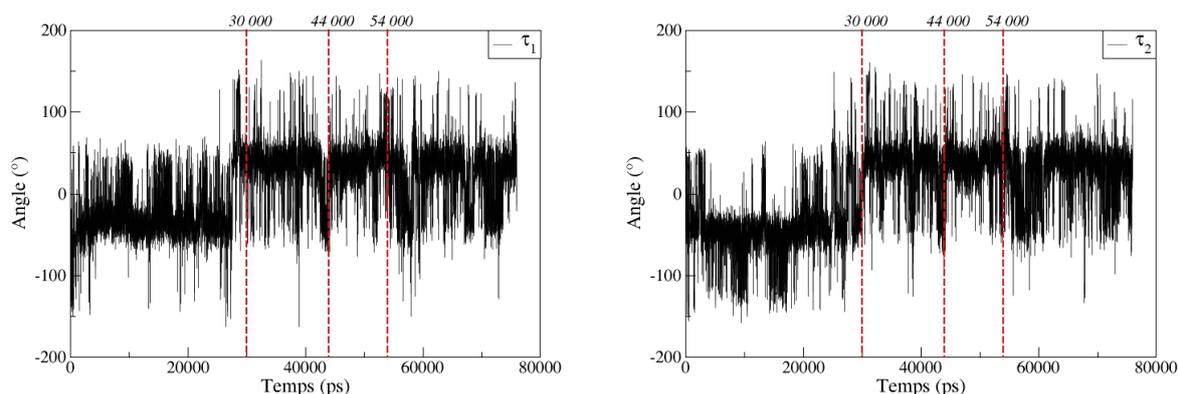


FIGURE 84 – Angles dièdres τ_1 (gauche) et τ_2 (droite) du DADS²⁻ lors de la simulation de dynamique moléculaire au sein du site IIIA de la ASH.

La variation des angles dièdres τ_1 et τ_2 est induite par la disparition de l'interaction entre le groupement sulfonate S₂₇ et l'Arg410, offrant une plus grande liberté de mouvement au niveau de cette moitié de la molécule. De plus, le DADS²⁻ commence à se mouvoir légèrement en dehors de la poche hydrophobe, le groupement sulfonate S₂₃ étant cependant toujours en interaction avec les résidus polaires, Arg410, Tyr411, Lys414 et Ser489.

Pendant les dix nanosecondes de la troisième période (44-54 ns), le DADS²⁻ adopte une nouvelle conformation au sein de la poche hydrophobe puisque le ligand migre à nouveau vers la sortie. Les interactions avec l'Arg410 et la Tyr411 disparaissent complètement et la Lys414 vient s'intercaler entre les deux sulfonates afin de les ponter *via* des ponts salins (Figure 85, gauche). Les distances entre l'azote terminal de la Lys414 et les deux atomes de soufre (S₂₃ et S₂₇) sont représentées à la Figure 85 (droite). Pendant la période 44-54 ns, les deux distances sont presque égales, environ 4 Å. Le déplacement du ligand positionne la fonction amine terminale du DADS²⁻, porté par l'atome N₃₂, proche de la Ser489 où une liaison hydrogène stabilise la position obtenue.

Pendant la dernière période (54-75 ns), le DADS²⁻ sort complètement de la poche hydrophobe (Figure 86, gauche). L'interaction pontante de la Lys414 disparaît et le ligand se retrouve comme libre à la surface de la protéine. La distance entre l'atome N₃₂ du DADS²⁻ et la Ser489 illustre très bien le déplacement du ligand par rapport à la poche hydrophobe (Figure 86, droite) à partir de 54 ns. La migration du ligand hors de la poche hydrophobe du sous-domaine IIIA indique clairement que ce dernier n'est pas apte à accueillir la molécule de DADS²⁻.

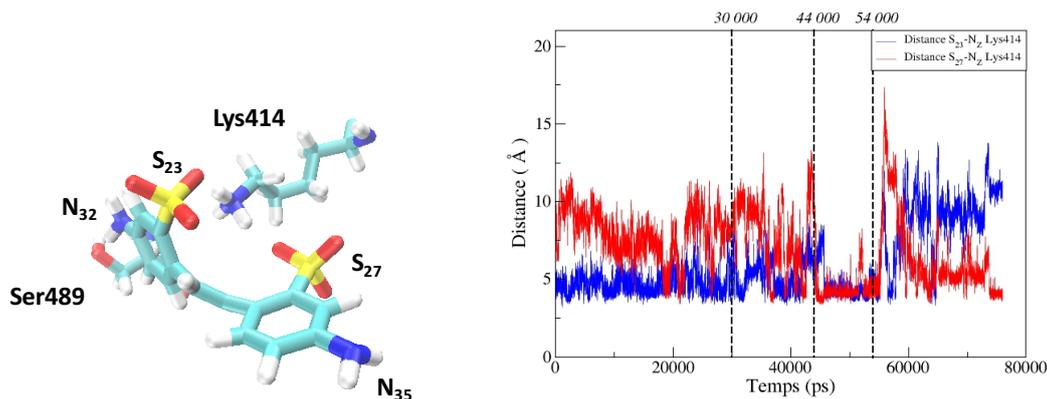


FIGURE 85 – Illustration de la Lys414 pontant les deux groupements sulfonates du DADS²⁻ (gauche), et distances entre les atomes S₂₃ (bleu) et S₂₇ (rouge) du DADS²⁻ avec l'ammonium terminal de la Lys414 (droite) lors de la simulation de dynamique moléculaire au sein du site IIIA de la ASH.

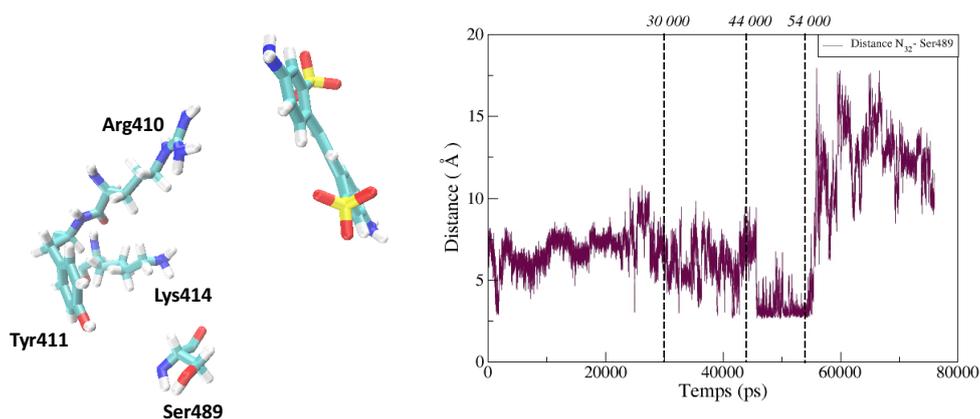


FIGURE 86 – Illustration de la migration hors de la poche hydrophobe du DADS²⁻ à 56 ns (gauche), et distances entre l'amine terminale N₃₂ du DADS²⁻ et la fonction OH de la Ser489 (droite) lors de la simulation de dynamique moléculaire au sein du site IIIA de la ASH.

7.5 Discussions

L'étude des différentes simulations de MD du DADS²⁻ au sein des différents sites de la ASH a permis d'identifier les résidus importants présents dans chaque site afin de caractériser les interactions essentielles au complexe protéine-ligand. À partir de ces analyses, des conclusions et tendances ressortent.

Les interactions du DADS²⁻ avec son environnement sont principalement de type électrostatique, ce qui semble logique au vu de la nature chimique de ce ligand. Les deux fonctions sulfonates sont principalement impliquées à travers des ponts salins et liaisons hydrogènes avec des résidus

polaires ; les fonctions amines terminales sont elles aussi sollicitées en tant que groupement donneur de liaison hydrogène, en particulier avec des fonctions carbonyles du squelette carboné de la protéine, et certains sites sont favorables à la présence d'interactions π -cation avec le ligand afin de le stabiliser davantage. Au regard des différents sites de la ASH étudiés et des résultats obtenus, nous pouvons établir que les sites IB et IIA apparaissent comme d'excellents candidats pour pouvoir accueillir le DADS²⁻. En effet, ces deux sites ont montré que l'environnement qui les compose permet de fortement stabiliser le DADS²⁻, notamment par la présence de nombreux résidus polaires chargés de type lysines et arginines. Le site IIIA semble ne pas être adapté pour accepter le DADS²⁻ au sein de sa poche hydrophobe. La simulation de MD effectuée dans cette région de la protéine a montré que le ligand est peu enclin à rester dans la poche et migre en dehors peu après 50 ns. La similitude structurale (poche hydrophobe) avec le site IIA et la différence de comportement du DADS²⁻ dans ces deux sites est peut-être surprenante. Cependant, deux éléments peuvent expliquer cette dissimilitude. Le premier est la taille de la poche hydrophobe car celle du site IIA possède un volume plus important dans le cas du site IIA. Le deuxième est la présence dans la poche du site IIA de deux groupes de résidus polaires (un à l'entrée et l'autre dans une des sous-chambres), permettant une meilleure stabilisation du ligand, alors qu'au sein du site IIIA, seul un groupe de quatre résidus polaires à l'entrée de la poche est présent. Ces deux différences mettent en avant la limite pour la poche hydrophobe du site IIIA d'accueillir le DADS²⁻. Le site IIAIB se positionne pour être un site de fixation secondaire du DADS²⁻ car il ne présente que peu de résidus polaires et sa localisation au sein de la protéine l'expose fortement au solvant. Cependant, lors de la simulation de MD nous n'avons pas observé de décrochage du DADS²⁻ de la protéine ASH, ce qui nous conduit à penser que le DADS²⁻ peut être accueilli par le site IIAIB.

Nous avons effectué des calculs MM/GBSA dans le but d'obtenir des énergies de complexation ASH-DADS²⁻ relatives pour chaque site étudié et ainsi compléter notre analyse. Les valeurs moyennes des énergies de complexation relatives sont les suivantes : -19,6 kcal.mol⁻¹ pour le site IB ; -3,4 et -15,6 kcal.mol⁻¹, respectivement, à l'intérieur et à l'extérieur de la poche hydrophobe du site IIA ; -9,9 kcal.mol⁻¹ pour le site IIAIB ; et -7,0 kcal.mol⁻¹ pour le site IIIA. De manière générale, l'ensemble des sites montrent des $\Delta G_{MM/GBSA}$ négatifs lors des simulations de MD. Ceci nous laisse à penser que le complexe ASH-DADS²⁻ est énergétiquement favorable, en considérant l'approximation de négliger l'entropie des systèmes. Dans le cas de la simulation du DADS²⁻ dans le sous-domaine IB, nous distinguons très clairement que ce site est le plus enclin à accueillir le DADS²⁻ (-19,6 kcal.mol⁻¹) et que la formation de la cage d'arginines autour du ligand favorise la formation du complexe protéine-ligand. Les deux simulations du DADS²⁻ dans le sous-domaine IIA donnent lieu à des valeurs d'énergies totalement différentes. La position du ligand à l'intérieur de la poche hydrophobe du sous-domaine IIA est moins favorable énergétiquement par rapport à sa position à l'extérieur (une différence de 12 kcal.mol⁻¹). De plus, la présence du DADS²⁻ dans la poche hydrophobe présente la valeur de $\Delta G_{MM/GBSA}$ la plus faible parmi les simulations réalisées (-3,4 kcal.mol⁻¹). Malgré le peu de résidus polaires présents au sein du site IIAIB, ce site présente la troisième meilleure énergie de complexation relative (-9,9 kcal.mol⁻¹).

L'exposition du ligand au solvant (l'eau) permet au complexe protéine-ligand d'être stable énergétiquement. Le cas DADS²⁻ dans le sous-domaine IIIA est un peu plus compliqué au vu de son départ de la protéine car il est indéniable que la poche hydrophobe du sous-domaine IIIA n'est pas adaptée pour accueillir le DADS²⁻. De plus, nous pouvons être critique sur les résultats obtenus (7,0 kcal.mol⁻¹) à cause de la migration du ligand en dehors de la poche. Nous avons exploré un scénario possible parmi tant d'autres et des simulations supplémentaires *via* des répliqua sont nécessaires pour étudier d'avantage ce site d'accueil.

7.6 Conclusions

Dans ce chapitre, nous nous sommes intéressés aux interactions entre le DADS²⁻ et la protéine de transport ASH. En effet, l'interaction avec la protéine de transport et donc la biodisponibilité du médicament sont des facteurs importants à considérer comme explicité en Introduction de cette thèse. Le complexe ASH-DADS²⁻ n'a pas encore été, à ce jour, cristallisé et aucune information sur la position de la molécule d'intérêt au sein de la ASH n'est disponible. Nous avons à l'aide de la mécanique moléculaire tenté de répondre à cette problématique. Les structures initiales du complexe ASH-DADS²⁻ ont été obtenues en s'appuyant sur la méthode de *docking* moléculaire semi-flexible. A l'aide de structures cristallographiques existantes de la ASH complexée avec d'autres ligands aromatiques, nous avons testé un protocole pour obtenir des complexes protéine-ligand cohérents et nous l'avons validé au vu des très bons accords avec l'expérience. Ensuite, nous avons *docké* de manière semi-flexible le DADS²⁻ au sein de quatre sites d'accueils probables de la ASH. Une (à deux) structure(s) complexée(s) de la ASH avec le ligand (selon les sites), a (ont) été sélectionnée(s). Ces structures complexées ont servi de point de départ pour les dynamiques moléculaires (75 ns). Nous avons constaté que la dynamique moléculaire (*Molecular Dynamics*, MD) apporte des informations complémentaires et essentielles sur l'évolution du ligand au sein de chaque site. En effet, la position du ligand au sein des sites obtenue issue du *docking* moléculaire semi-flexible fluctue lors des simulations de MD, notamment, dans le cas du sous-domaine IIIA où les premiers résultats de *docking* ne présageaient pas le départ du DADS²⁻ hors de la poche hydrophobe. Il est donc indispensable de compenser la limite du *docking* moléculaire semi-flexible par ces simulations qui relaxent le système (absence de contraintes).

L'analyse des simulations de MD pour les quatre sites potentiels d'accueil de la molécule de DADS²⁻ a permis d'identifier, sur des bases géométriques et énergétiques, deux sites d'accueil principaux : les sous-domaines IB et IIA. Le premier site présente de nombreuses interactions polaires et électrostatiques, aidant fortement à la stabilisation géométrique et énergétique du ligand. La formation d'une cage d'arginines autour du ligand est clairement identifiée comme un élément important pour son affinité avec ce site d'accueil. La combinaison de ponts salins et d'interactions π -cation se présente donc comme un mode d'interaction très favorable énergétiquement pour le complexe protéine-ligand. Le deuxième site principal qui pourrait accueillir le ligand serait le sous-domaine IIA. Dans ce site, le ligand est en forte interaction mais il se trouve à l'extérieur de

la poche hydrophobe. En effet, la simulation de MD du ligand situé à l'extérieur de cette dernière présente un profil énergétique favorable à la formation du complexe tandis que c'est moins le cas lorsque le DADS²⁻ se trouve à l'intérieur de la poche hydrophobe. De plus, ce sous-domaine présente un intérêt tout particulier dû à la position du seul tryptophane de la ASH (Trp214) par rapport au ligand où des interactions de type *T-shaped* sont présentes. Dans ce contexte, des mesures expérimentales de FRET (Förster Resonance Energy Transfer) pourraient confirmer la présence du DADS²⁻ près de ce résidu. Nous tenons également à souligner que le sous-domaine IIAIB peut être considéré comme un site secondaire potentiel car, malgré le peu d'interactions possibles avec les résidus le composant, la simulation de MD a montré qu'il pourrait accueillir le DADS²⁻.

En plus des expériences de FRET, des mesures de spectres d'absorption de la ASH seul et/ou complexée avec le DADS²⁻ pourraient être obtenues. Des calculs de spectres d'absorption du complexe ASH-DADS²⁻ par des méthodes hybrides alliant la mécanique quantique et la mécanique moléculaire (QM/MM) pourraient être corrélés avec les résultats expérimentaux ; le déplacement du spectre dû au phénomène de complexation serait un indicateur supplémentaire pour déterminer la (les) position(s) du DADS²⁻ au sein de la ASH.

Conclusion générale et perspectives

Durant notre thèse de doctorat, nous nous sommes intéressés à une famille de composés, les stilbènes disulfoniques (SD), connus pour inhiber la protéine Rad51 qui joue un rôle majeur dans la recombinaison par homologie de l'ADN. La protéine Rad51 est responsable de la résistance et de la réparation des cellules cancéreuses face aux traitements anti-cancéreux actuels. Cependant, elle est relativement complexe puisqu'elle polymérise, et sa structure cristallographique n'a été résolue que fin 2016. Nous avons donc, après une étude physico-chimique des composés, envisagé leur complexation avec la protéine de transport albumine sérique humaine (ASH). En effet, il est également important de considérer la biodisponibilité et la caractérisation avec ASH est cruciale.

Pour réaliser ce travail de thèse en chimie théorique, nous avons choisi les méthodes les plus adaptées en fonction de différents environnements dans lequel le(s) composé(s) d'intérêt se trouva(en)t. D'un côté, la mécanique quantique (*Quantum Mechanics*, QM) et plus particulièrement la théorie de fonctionnelle de densité (*Density Functional Theory*, DFT) et sa version dépendante du temps (*Time-dependent-DFT*, TD-DFT) pour étudier le pK_a , la structure conformationnelle et les spectres optiques de composés dans différents solvants. Les approches de mécanique moléculaire (*Molecular Mechanics*, MM) ont ensuite été nécessaires pour décrire le complexe ASH-SD allant d'une étape de *docking* moléculaire à la dynamique moléculaire (*Molecular Dynamics*, MD).

Le contexte biologique et les acteurs principaux du travail ont été présentés dans la première partie du manuscrit. Le premier chapitre se concentre sur la protéine ASH, sa découverte, sa structure tridimensionnelle et son rôle au sein de l'organisme. Le deuxième chapitre reprend des résultats importants obtenus sur la ASH, alliant la modélisation moléculaire à l'expérience ainsi qu'une présentation non exhaustive des études théoriques menées sur l'interaction entre ASH et certains médicaments. Le troisième chapitre est quant à lui consacré aux composés de la famille des SD qui nous ont particulièrement intéressés durant la thèse. Leurs applications en tant qu'inhibiteur de systèmes biologiques y sont montrées ainsi que leur grande réactivité.

La deuxième partie de la thèse est axée sur l'étude des propriétés physico-chimiques des composés SD par des méthodes reposant sur l'utilisation de calculs quantiques *ab initio*. Le Chapitre 4 reprend les théories et principes dont sont issues les méthodes QM utilisées lors de cette thèse. Les théories de la DFT et la TD-DFT y sont présentées ainsi que les modèles et formalismes employés pour la prise en compte des effets de solvants. Le chapitre suivant expose les résultats de notre étude sur les propriétés physico-chimiques des molécules SD en solution en combinant des mesures expérimentales réalisées au laboratoire et de nos calculs théoriques. Dans un premier temps,

nous nous sommes concentrés sur la détermination de l'état de protonation en solution aqueuse des SD. La mise en place de mesures de spectres d'absorptions à différentes valeurs de pH et la résolution d'un cycle thermodynamique de type Börn-Haber *via* des calculs de type DFT nous a permis de déterminer (i) qu'un seul état de protonation est présent en solution et (ii) qu'il s'agit de la forme totalement déprotonée (SD^{2-}). Cette donnée structurale est cruciale en vue de notre étude sur les complexes protéine-ligand. Dans un deuxième temps, nous nous sommes penchés sur les propriétés optiques des SD^{2-} en solution. Les spectres d'absorptions expérimentaux indiquent un effet solvatochrome (déplacement bathochromique) entre les solvants eau et DMSO pour les molécules $DIDS^{2-}$ et $DADS^{2-}$. Une conclusion importante est que la simulation de cet effet par les calculs DFT et TD-DFT n'a été possible que par l'utilisation du modèle de solvation (SMD) car ce dernier permet de reproduire le déplacement bathochromique observé expérimentalement. Des contributions vibroniques sur les spectres d'absorptions expérimentaux du $DIDS^{2-}$ dans le DMSO ont été constatés et nous avons été capable de reproduire la forme caractéristique des bandes d'absorption.

Dans la troisième et dernière partie de notre travail, nous nous sommes intéressés au traitement du complexe protéine-ligand, et notamment, le complexe ASH- $DADS^{2-}$, à travers des calculs de type MM. Le Chapitre 6 présente les principes de la MM, de la MD et des calculs d'énergie libre relative (MM/GBSA). Le Chapitre 7 présente les travaux sur le complexe ASH-ligand que nous avons menés. La première partie du chapitre concerne la validation de notre méthode de *docking* moléculaire, à travers des tests effectués sur trois molécules 4Z,15E-bilirubine-IX α , la warfarine et l'ibuprofène, pour lesquelles une structure cristallographique existe sous leur forme complexée avec la ASH. Les résultats ont montré que notre protocole de *docking* moléculaire permettait de retrouver des poses de ces ligands dans les différents sites d'accueil de la ASH très similaires aux structures cristallographiques (RMSD < 1 Å). Le même protocole pouvait donc être utilisé pour obtenir des structures de complexe ASH- $DADS^{2-}$. Nous avons ensuite docké le $DADS^{2-}$ dans quatre sites (IB, IIA, IIAIIB et IIIA) d'accueil potentiel de la ASH et sélectionné les meilleures poses afin de les utiliser comme structure de départ pour les simulations de MD des complexes. Après une étape de minimisation et d'équilibration des systèmes au sein d'une boîte d'eau, des simulations de 75 ns ont été produites pour chaque site potentiel. Nous constatons que les interactions ASH- $DADS^{2-}$ impliquaient principalement des résidus polaires et chargés pour former de fortes interactions électrostatiques avec le $DADS^{2-}$, *via* principalement ses deux fonctions sulfonates. Parmi les sites explorés, les sites IB et IIA (en dehors de la poche hydrophobe) se présentent comme étant les sous-domaines les plus favorables à l'accueil du $DADS^{2-}$. Ces sites sont composés de nombreux résidus polaires, aidant à la stabilisation du ligand, notamment, dans le site IB où une cage à arginine apparaît comme mode de liaison avec le $DADS^{2-}$. Le site IIA est aussi très intéressant de par la présence de l'unique résidu de type tryptophane de la ASH pour des mesures expérimentales de type FRET (*Förster Resonance Energy Transfer*). En revanche, les sites d'accueil IIA et IIIA, sites souvent privilégiés pour de nombreux médicaments, sont caractérisés par des poches hydrophobes et ne semble pas être adéquat à accueillir le $DADS^{2-}$. La poche

du sous-domaine IIA présente une énergie de complexation relativement faible et celle du sous-domaine IIIA ne permet pas de conserver le ligand dans son enceinte car ce dernier migre en-dehors de la poche durant la simulation.

Cette partie concluant notre travail de thèse, nous aimerions, cependant, préciser que des travaux ont également été réalisés sur d'autres SD (DNDS²⁻ et DATDS²⁻) ainsi que sur un autre inhibiteur prometteur, le B02 (molécule au motif quinazolinone). Les résultats sur ASH-B02 sont sensiblement différents que ceux présentés dans notre manuscrit puisque les interactions protéine-ligand sont guidées par des interactions de van der Waals ; les deux poches hydrophobes (sous-domaines IIA et IIIA) sont donc plus à même d'accueillir le B02.

En perspective de nos travaux, il serait intéressant de prévoir des mesures de spectres optiques de la protéine ASH libre et complexée avec les molécules de DADS²⁻, DNDS²⁻ et DATDS²⁻ pour comparer des simulations de propriétés optiques des complexes formés (ASH-SD²⁻), réalisés à l'aide de méthodes hybrides alliant la mécanique quantique et la mécanique moléculaire (QM/MM). Cela permettrait de déterminer la (les) localisation(s) des ligands au sein de la protéine ASH en comparant les spectres théoriques avec les données obtenus par l'expérience. Des mesures de FRET expérimentales pourraient être également programmées et, à l'issue des résultats obtenus, des mutations de la ASH de type *Tryptophane-Scanning* seront envisagées. Des analyses par spectrométrie de masse sont, par ailleurs, planifiées en juillet afin de répondre à une hypothèse selon laquelle les composés DIDS²⁻, DASDSS²⁻ et SITS²⁻ se lieraient de manière covalente avec la protéine ASH. Si les conclusions de l'expérience ne montrent pas l'existence de liaisons covalentes entre ces composés et la protéine ASH, alors il conviendrait d'envisager des simulations de MD des trois composés au sein de ASH. Cependant, un travail méthodologique préliminaire sur l'obtention des paramètres de champs de force des fonctions -NCS (DIDS²⁻ et SITS²⁻) et N₃ (DASDSS²⁻) nous paraît indispensable avant les simulations de MD.

Liste des abréviations

ADME	Absorption, Distribution, Métabolisme et Excrétion
ADN	Acide désoxyribonucléique
ADNdb	ADN double brin
ADNsb	ADN simple brin
AMBER	Assisted Model Building with Energy Refinement
ASB	Albumine sérique bovine
ASH	Albumine sérique humaine
BO	Born-Oppenheimer
CASPT2	Complete Active Space Perturbation Theory 2
CAS-SCF	Complete Active-Space Self Consistent Field
CGENFF	CHARMM General Force Field
CGTO	Constrained GTO
CHARMM	CHemistry at Harvard Macromolecular Mechanics
CLOA	Combinaison Linéaire d'Orbitales Atomiques
cLR	Corrected Linear Response
DADS	Acide 4,4'-diaminostilbène-2,2'-disulfonique
DATDS	Acide 4,4'-diazidostilbène-2,2'-disulfonique
DATDS	Acide 4,4'-diacétamidostilbène-2,2'-disulfonique
DFT	Density Functional Theory
DIDS	Acide 4,4'-diisothiocyanostilbène-2,2'-disulfonique
DMSO	Diméthylsulfoxyde
DNDS	Acide 4,4'-dinitrostilbène-2,2'-disulfonique
EPR	Enhance Permability and Retention

eq	Équilibre
FA	Fatty Acid
FC	Franck-Condon
FRET	Förster Resonance Energy Transfer
GAFF	Generalysed AMBER Force Field
GE	Glide Emodel
GGA	Generalysed Gradient Approximation
gue	Gaz uniforme d'électrons
GTO	Gaussian Type Orbital
HF	Hartree-Fock
HPLC	High Pressure liquid Chromatography
HsRad51	<i>Homo sapiens</i> Rad51
IE	Implicite/Explicite
KS	Kohn-Sham
L	Ligand
LDA	Local Density Approximation
LJ	Lennard-Jones
LR	Linear Response
LSDA	Local Spin Density Approximation
MD	Molecular Dynamics
MeOH	Méthanol
MM	Molecular Mechanics
MM/GBSA	Molecular Mechanics/Generalysed Born Surface Area
MS-CASPT2	Multi-state CASPT2
neq	non-équilibre
PCM	Polarisable Continuum Model

PDB	Protein Data Bank
pKa	Constante d'acidité
P	Protéine
PL	Protéine-Ligand
QM	Quantum Mechanics
RH	Recombinaison Homologue
RMN	Résonance Magnétique Nucléaire
RMSD	Root mean-Square Deviation
RSH	Range Separated Hybrid
S₀	État fondamental
S₁	Premier état excité
SAS	Solvent Accessible Surface
SASA	Solvent Accessible Surface Area
SD	Stilbène disulfonique
SES	Solvent Exclusion Surface
SITS	acide 4-acétamido-4'-isothiocyanostilbène-2,2'-disulfonique
SMD	Solvation Model based on solute electron Density
SS	State Specific
STO	Slater Type Orbital
TD-DFT	Time-dependent DFT
TF	Thomas-Fermi
TFD	Thomas-Fermi-Dirac
vdW	van der Waals
XCF	eXchange-Correlation Functional
XP	Extra Precision

Bibliographie

- [1] V. Besse, Proposition de loi visant à faire de la prévention et du dépistage du cancer une grande cause nationale 2017.
- [2] J. L. Biedler, *Cancer Research* **54**, 666 (1994).
- [3] N. S. Nagathihalli and G. Nagaraju, *BBA-Reviews on Cancer* **1816**, 209 (2011).
- [4] A. Shinohara, H. Ogawa, and T. Ogawa, *Cell* **69**, 457 (1992).
- [5] H. Han, D. J. Bearss, L. W. Browne, R. Calaluce, R. B. Nagle, and D. D. Von Hoff, *Cancer Res.* **62**, 2890 (2002).
- [6] L. T. Hansen, C. Lundin, M. Spang-Thomsen, L. N. Petersen, and T. Helleday, *Int. J. Cancer* **105**, 472 (2003).
- [7] A. P. Wiegmans, F. Al-Ejeh, N. Chee, P.-Y. Yap, J. J. Gorsk, L. D. Silva, E. Bolderson, G. Chenevix-Trench, R. Anderson, P. T. Simpson, S. R. Lakhani, and K. K. Khanna, *Oncotarget* **5**, 3261 (2014).
- [8] S. Vispé, C. Cazaux, C. Lesca, and M. Defais, *Nucleic Acids Res.* **26**, 2859 (1998).
- [9] J. A. Hannay, J. Liu, Q.-S. Zhu, S. V. Bolshakov, L. Li, P. W. Pisters, A. J. Lazar, D. Yu, R. E. Pollock, and D. Lev, *Mol. Cancer Ther.* **6**, 1650 (2007).
- [10] J. Nomme, Y. Takizawa, S. F. Martinez, A. Renodon-Cornière, F. Fleury, P. Weigel, K. I. Yamamoto, H. Kurumizaka, and M. Takahashi, *Genes Cells* **13**, 471 (2008).
- [11] B. Budke, H. L. Logan, J. H. Kalin, A. S. Zelivianskaia, W. Cameron McGuire, L. L. Miller, J. M. Stark, A. P. Kozikowski, D. K. Bishop, and P. P. Connell, *Nucleic Acids Res.* **40**, 7347 (2012).
- [12] J. Zhu, L. Zhou, G. Wu, H. Konig, X. Lin, G. Li, X.-L. Qiu, C.-F. Chen, C.-M. Hu, E. Goldblatt, R. Bhatia, A. R. Chamberlin, P.-L. Chen, and W.-H. Lee, *EMBO Molecular Medicine* **5**, 353 (2013).
- [13] T. Ishida, Y. Takizawa, T. Kainuma, J. Inoue, T. Mikawa, T. Shibata, H. Suzuki, S. Tashiro, and H. Kurumizaka, *Nucleic Acids Res.* **37**, 3367 (2009).

- [14] D. Velic, Ph.D. thesis, Université de Nantes, 2016.
- [15] Y. Matsumura and H. Maeda, *Cancer Res.* **46**, 6387 (1986).
- [16] R. K. Jain, *Cancer Res.* **48**, 2641 (1988).
- [17] T. G. Burke and Z. Mi, *J. Med. Chem.* **37**, 40 (1994).
- [18] F. Fleury, A. Ianoul, M. Berjot, A. Feofanov, A. J. Alix, and I. Nabiev, *FEBS Letters* **411**, 215 (1997).
- [19] G. Ferraro, L. Massai, L. Messori, and A. Merlino, *Chem. Commun.* **51**, 9436 (2015).
- [20] U. Kragh-Hansen, *Dan. Med. Bull.* **37**, 57 (1990).
- [21] T. Peters, in *All About Albumin. Biochemistry, Genetics, and Medical Applications*, edited by I. Academic Press (Elsevier Science, San Diego, 1996).
- [22] M. Fasano, S. Curry, E. Terreno, M. Galliano, G. Fanali, P. Narciso, S. Notari, and P. Ascenzi, *IUBMB Life* **57**, 787 (2005).
- [23] S. Curry, *Drug Metab. Pharmacok.* **24**, 342 (2009).
- [24] G. Fanali, A. di Masi, V. Trezza, M. Marino, M. Fasano, and P. Ascenzi, *Mol. Aspects Med.* **33**, 209 (2012).
- [25] T. Coar, in *The Aphorisms of Hippocrates with a Translation into Latin and English*, edited by Longman (Royal College of Physicians, London, 1822).
- [26] W. Pagel, in *Paracelsus : An Introduction to Philosophical Medicine in the Era of the Renaissance*, edited by K. Medical and S. Publishers (Karger, New York, 1982).
- [27] C. Leake, in *Exercitatio Anatomica De Motu Cordis Et Sanguinis In Animalibus (Anatomical Studies on the Motion of the Heart and Blood) by William Harvey : An English Translation with Annotations*, edited by C. C. Thomas (Springfield, Illinois, 1929).
- [28] G. Kauder, *Arch. Exp. Pathol. Pharmacol.* **20**, 411 (1886).
- [29] T. Svedberg, *Science* **79**, 327 (1934).
- [30] E. J. Cohn, L. E. Strong, W. L. Hughes, D. J. Mulford, J. N. Ashworth, M. Melin, and H. L. Taylor, *J. Am. Chem. Soc.* **68**, 459 (1946).
- [31] J. Ghuman, P. A. Zunszain, I. Petitpas, A. A. Bhattacharya, M. Otagiri, and S. Curry, *J. Mol. Biol.* **353**, 38 (2005).
- [32] T. W. Evans, *Aliment. Pharmacol. Ther.* **16**, 6 (2002).

- [33] M. M. Duarte, J. B. Rocha, R. N. Moresco, T. Duarte, I. B. Da Cruz, V. L. Loro, and M. R. Schetinger, *Clin. Biochem.* **42**, 666 (2009).
- [34] D. Gupta and C. G. Lis, *Nutr. J.* **9**, 69 (2010).
- [35] M. Koga and S. Kasayama, *Endocrine J.* **57**, 751 (2010).
- [36] S. Eftihia, P. Georgiadou, and V. Voudris, *Clin. Chem. Lab. Med.* **49**, 177 (2011).
- [37] M. Roche, P. Rondeau, N. R. Singh, E. Tarnus, and E. Bourdon, *FEBS Lett.* **582**, 1783 (2008).
- [38] B. Meloun, L. Moravek, and V. Kostka, *FEBS Lett.* **58**, 134 (1975).
- [39] J. E. Foster, in *Plasma albumin, in the plasma proteins*, edited by E. Putnam (Academic Press, New York, 1960), Vol. 1, pp. 179–239.
- [40] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, and K. Kobayashi, *Protein Eng.* **12**, 439 (1999).
- [41] G. Sudlow, D. J. Birkett, and D. N. Wade, *Mol. Pharmacol.* **11**, 824 (1975).
- [42] G. Sudlow, D. J. Birkett, and D. N. Wade, *Clin. Exp. Physiol. Pharmacol.* **2**, 129 (1975).
- [43] G. Sudlow, D. J. Birkett, and D. N. Wade, *Mol. Pharmacol.* **12**, 1052 (1976).
- [44] F. Karush, *J. Am. Chem. Soc.* **72**, 2705 (1950).
- [45] D. Koshland, *Proc. Natl. Acad. Sci.* **44**, 98 (1958).
- [46] D. Carter, X. He, S. Munson, P. Twigg, K. Gernert, M. Broom, and T. Miller, *Science* **244**, 1195 (1989).
- [47] D. Carter and X. He, *Science* **249**, 302 (1990).
- [48] X. He and D. Carter, *Nature* **358**, 209 (1992).
- [49] S. Curry, H. Mandelkow, P. Brick, and N. Franks, *Nat. Struct. Mol. Biol.* **5**, 827 (1998).
- [50] S. Curry, P. Brick, and N. P. Franks, *Biochim. Biophys. Acta* **1441**, 131 (1999).
- [51] A. A. Bhattacharya, T. Grune, and S. Curry, *J. Mol. Biol.* **303**, 721 (2000).
- [52] I. Petitpas, T. Grune, A. A. Bhattacharya, and S. Curry, *J. Mol. Biol.* **314**, 955 (2001).
- [53] M. Wardell, Z. Wang, J. X. Ho, J. Robert, F. Ruker, J. Ruble, and D. C. Carter, *Biochem. Biophys. Res. Commun.* **291**, 813 (2002).

- [54] P. A. Zunszain, J. Ghuman, T. Komatsu, E. Tsuchida, and S. Curry, *BMC Struct. Biol.* **3**, 6 (2003).
- [55] P. A. Zunszain, J. Ghuman, A. F. McDonagh, and S. Curry, *J. Mol. Biol.* **381**, 394 (2008).
- [56] F. Yang, C. Bian, L. Zhu, G. Zhao, Z. Huang, and M. Huang, *J. Struct. Biol.* **157**, 348 (2007).
- [57] A. A. Bhattacharya, S. Curry, and N. P. Franks, *J. Biol. Chem.* **275**, 38731 (2000).
- [58] L. Zhu, F. Yang, L. Chen, E. J. Meehan, and M. Huang, *J. Struct. Biol.* **162**, 40 (2008).
- [59] K. L. Hein, U. Kragh-Hansen, J. P. Morth, M. D. Jeppesen, D. Otzen, J. V. Moller, and P. Nissen, *J. Struct. Biol.* **171**, 353 (2010).
- [60] Y. Zhang, P. Lee, S. Liang, Z. Zhou, X. Wu, F. Yang, and H. Liang, *Chem. Biol. Drug Des.* **86**, 1178 (2015).
- [61] A. J. Ryan, J. Ghuman, P. A. Zunszain, C. Chung, and S. Curry, *J. Struct. Biol.* **174**, 84 (2011).
- [62] J. E. Walker, *FEBS Lett.* **66**, 173 (1976).
- [63] J. T. Gerig, K. E. Katz, J. D. Reinheimer, G. R. Sullivan, and J. D. Roberts, *Org. Magn. Reson.* **15**, 158 (1981).
- [64] Y. Kurono, K. Ichioka, and K. Ikeda, *J. Pharm. Sci.* **72**, 432 (1983).
- [65] M. Yvon and J.-M. Wal, *FEBS Lett.* **239**, 237 (1988).
- [66] J. T. Gerig and J. D. Reinheimer, *J. Am. Chem. Soc.* **97**, 168 (1975).
- [67] J. Gerig, K. Katz, and J. Reinheimer, *Biochim. Biophys. Acta* **534**, 196 (1978).
- [68] N. Iberg and R. Fluckiger, *J. Biol. Chem.* **261**, 13542 (1986).
- [69] D. C. Carter and J. X. Ho, *Adv. Protein Chem.* **45**, 153 (1994).
- [70] N. Díaz, D. Suárez, T. L. Sordo, and K. M. Merz, *J. Med. Chem.* **44**, 250 (2001).
- [71] C. N. Pace, G. R. Grimsley, J. A. Thomson, and B. J. Barnett, *J. Biol. Chem.* **263**, 11820 (1988).
- [72] S. K. Nayak, D. Rathore, and J. K. Batra, *Biochemistry* **38**, 10052 (1999).
- [73] W. J. Wedemeyer, E. Welker, M. Narayan, and H. A. Scheraga, *Biochemistry* **39**, 4207 (2000).
- [74] E. Dumont, A. D. Laurent, P.-F. Loos, and X. Assfeld, *J. Chem. Theory. Comput.* **5**, 1700 (2009).

- [75] C. David, S. Foley, C. Mavon, and M. Enescu, *Biopolymers* **89**, 623 (2008).
- [76] G. Paris, S. Kraszewski, C. Ramseyer, and M. Enescu, *Biopolymers* **97**, 889 (2012).
- [77] M. M. Castellanos and C. M. Colina, *J. Phys. Chem. B* **117**, 11895 (2013).
- [78] D. S. Fredrickson and R. S. Gordon Jr., *J. Clin. Invest.* **37**, 1504 (1958).
- [79] T. Jaunet-Lahary, A. Goupille, D. Jacquemin, F. Fleury, J. Graton, and A. D. Laurent, *Chem-PhysChem* **17**, 2434 (2016).
- [80] S. Fujiwara and T. Amisaki, *Proteins* **64**, 730 (2006).
- [81] H. Vorum and B. Honoré, *J. Pharm Pharmacol.* **48**, 870 (1996).
- [82] S. Fujiwara and T. Amisaki, *Biophys. J.* **94**, 95 (2008).
- [83] S. Fujiwara and T. Amisaki, *Chem.Pharm. Bull.* **59**, 860 (2011).
- [84] Y.-J. Hu, Y. Ou-Yang, C.-M. Dai, Y. Liu, and X.-H. Xiao, *Biomacromolecules* **11**, 106 (2010).
- [85] N. Ibrahim, H. Ibrahim, S. Kim, J.-P. Nallet, and F. Nepveu, *Biomacromolecules* **11**, 3341 (2010).
- [86] V. D. Suryawanshi, P. V. Anbhule, A. H. Gore, S. R. Patil, and G. B. Kolekar, *Ind. Eng. Chem. Res.* **51**, 95 (2012).
- [87] B. Sudhamalla, M. Gokara, N. Ahalawat, D. G. Amooru, and R. Subramanyam, *J. Phys. Chem. B* **114**, 9054 (2010).
- [88] C. Domonkos, I. Fitos, J. Visy, and F. Zsila, *Mol. Pharm.* **10**, 4706 (2013).
- [89] C. Qin, M.-X. Xie, and Y. Liu, *Biomacromolecules* **8**, 2182 (2007).
- [90] Y. Hu, K. N. Houk, K. Kikuchi, K. Hotta, and D. Hilvert, *J. Am. Chem. Soc.* **126**, 8197 (2004).
- [91] G. Colmenarejo, A. Alvarez-Pedraglio, and J.-L. Lavandera, *J. Med. Chem.* **44**, 4370 (2001).
- [92] O. Deeb, M. C. Rosales-Hernández, C. Gómez-Castro, R. Garduño-Juárez, and J. Correa-Basurto, *Biopolymers* **93**, 161 (2010).
- [93] I. Petitpas, A. A. Bhattacharya, S. Twine, M. East, and S. Curry, *J. Biol. Chem* **276**, 22804 (2001).
- [94] B. K. Sahoo, K. S. Ghosh, and S. Dasgupta, *Biopolymers* **91**, 108 (2009).

- [95] O. K. Abou-Zied, N. Al-Lawatia, M. Elstner, and T. B. Steinbrecher, *J. Phys. Chem. B* **117**, 1062 (2013).
- [96] N. J. Gumede, P. Singh, M. I. Sabela, K. Bisetty, L. Escuder-Gilabert, M. J. Hernández, and S. Sagrado, *J. Chem. Inf. Mod.* **52**, 2754 (2012).
- [97] S. Fujiwara and T. Amisaki, *BBA-General Subjects* **1830**, 5427 (2013).
- [98] P. Daneshgar, A. A. Moosavi-Movahedi, P. Norouzi, M. R. Ganjali, A. Madadkar-Sobhani, and A. A. Saboury, *Inter. J. Biol. Macromol.* **45**, 129 (2009).
- [99] G. Fanali, V. Rampoldi, A. di Masi, A. Bolli, L. Lopiano, P. Ascenzi, and M. Fasano, *IUBMB Life* **62**, 371 (2010).
- [100] Y. Wang, X. Wang, J. Wang, Y. Zhao, W. He, and Z. Guo, *Inorg. Chem.* **50**, 12661 (2011).
- [101] T. Zohoorian-Abootorabi, H. Sane, H. Iranfar, M. R. Saberi, and J. Chamani, *Spectrochim. Acta A* **88**, 177 (2012).
- [102] E. Rass, A. Grabarz, P. Bertrand, and B.-S. Lopez, *Cancer/Radiothérapie* **16**, 1 (2012).
- [103] B. Alligand, Ph.D. thesis, Université Nantes, 2015.
- [104] S. P. Jackson and J. Bartek, *Nature* **461**, 1071 (2009).
- [105] X. Yu and J. Chen, *Mol. Cell. Biol.* **24**, 9478 (2004).
- [106] T. H. Stracker and J. H. J. Petrini, *Nat. Rev. Mol. Cell Biol.* **12**, 90 (2011).
- [107] T. Sugiyama and N. Kantake, *J. Mol. Biol.* **390**, 45 (2009).
- [108] O. Ostling and K. J. Johanson, *Biochem. Biophys. Res. Commun.* **123**, 291 (1984).
- [109] L. Q. Du, X. Q. Du, J. Q. Bai, Y. Wang, Q. S. Yang, X. C. Wang, P. Zhao, H. Wang, Q. Liu, and F. Y. Fan, *J. Cancer Res. Clin. Oncol.* **138**, 811 (2012).
- [110] J. J. Buggy, Z. A. Cao, K. E. Bass, E. Verner, S. Balasubramanian, L. Liu, B. Schultz, P. R. Young, and S. A. Dalrymple, *Mol. Cancer. Ther.* **5**, 1309 (2006).
- [111] S. Adimoolam, M. Sirisawad, J. Chen, P. Thiemann, J. M. Ford, and J. J. Buggy, *Proc. Natl. Acad. Sci. USA* **104**, 19482 (2007).
- [112] X. Zhang, H. Sun, D. C. Danila, S. R. Johnson, Y. Zhou, B. Swearingen, and A. Klibanski, *J. Clin. Endocrinol. Metab.* **87**, 1262 (2002).
- [113] C. C. Chang, W. C. Chen, T. F. Ho, H. S. Wu, and Y. H. Wei, *J. Biosci. Bioeng.* **111**, 501 (2011).

- [114] R. Pandey, R. Chander, and K. B. Sainis, *Curr. Pharm. Des.* **15**, 732 (2009).
- [115] R. Perez-Tomas, B. Montaner, E. Llagostera, and V. Soto-Cerrato, *Biochem. Pharmacol.* **66**, 1447 (2003).
- [116] Y. Li, Y. He, and Y. Luo, *Biochemistry* **48**, 6805 (2009).
- [117] Y. Wu, Y. He, I. A. Moya, X. Qian, and Y. Luo, *Mol. Cell* **15**, 423 (2004).
- [118] Y. Wu, X. Qian, Y. He, I. A. Moya, and Y. Luo, *J. Biol. Chem.* **28**, 722 (2005).
- [119] X. Qian, Y. Wu, Y. He, and Y. Luo, *Biochemistry* **44**, 13753 (2005).
- [120] M. Takaku, T. Kainuma, T. Ishida-Takaku, S. Ishigami, H. Suzuki, S. Tashiro, R. W. van Soest, Y. Nakao, and H. Kurumizaka, *Genes Cells* **16**, 427 (2011).
- [121] F. Huang, N. A. Motlekar, C. M. Burgwin, A. D. Napper, S. L. Diamond, and A. V. Mazin, *ACS Chemical Biology* **6**, 628 (2011).
- [122] F. Huang, O. M. Mazina, I. J. Zentner, S. Cocklin, and A. V. Mazin, *J. Med. Chem.* **55**, 3011 (2012).
- [123] A. B. Conway, T. W. Lynch, Y. Zhang, G. S. Fortin, C. W. Fung, L. S. Symington, and P. A. Rice, *Nat. Struct. Mol. Biol.* **11**, 791 (2004).
- [124] K. Jayathilaka, S. D. Sheridan, T. D. Bold, K. Bochenska, H. L. Logan, R. R. Weichselbaum, D. K. Bishop, and P. P. Connell, *Proc. Natl. Acad. Sci. USA* **105**, 15848 (2008).
- [125] Z. I. Cabantchik and A. Rothstein, *J. Membr. Biol.* **10**, 311 (1972).
- [126] Z. I. Cabantchik and R. Greger, *Am. J. Physiol.* **262**, C803 (1992).
- [127] C. Miller and M. M. White, *Ann. N.Y. Acad. Sci.* **341**, 534 (1980).
- [128] G. B. Makara, E. Stark, M. Karteszi, M. Palkovits, and G. Rappay, *Am. J. Physiol.* **240**, 441 (1981).
- [129] I. Parker, K. Sumikawa, and R. Miledi, *Proc. R. Soc. London, Ser. B* **233**, 201 (1988).
- [130] A. Picollo, A. Liantonio, M. P. Didonna, L. Elia, D. C. Camerino, and M. Pusch, *EMBO Rep.* **5**, 584 (2004).
- [131] K. Matulef and M. Maduke, *Biophys. J.* **89**, 1721 (2005).
- [132] I. Koszela-Piotrowska, K. Choma, P. Bednarczyk, K. Dolowy, A. Szewczyk, W. S. Kunz, L. Malekova, V. Kominkova, and K. Ondrias, *Cell. Mol. Biol. Lett.* **12**, 493 (2007).
- [133] A. M. Garcia and H. F. Lodish, *J. Biol. Chem.* **264**, 19607 (1989).

- [134] M. M. Kay, *FASEB J.* **5**, 109 (1991).
- [135] K. Okubo, D. Kang, N. Hamasaki, and M. L. Jennings, *J. Biol. Chem.* **269**, 1918 (1994).
- [136] E. Winograd, S. Eda, and I. W. Sherman, *Mol. Biochem. Parasitol.* **136**, 243 (2004).
- [137] T. Yamaguchi, Y. Ikeda, Y. Abe, H. Kuma, D. Kang, N. Hamasaki, and T. Hirai, *J. Mol. Biol.* **397**, 179 (2010).
- [138] J. M. Salhany, *Cell. Mol. Biol.* **42**, 1065 (1996).
- [139] T. J. Jentsch, V. Stein, F. Weinreich, and A. A. Zdebik, *Physiol. Rev.* **82**, 503 (2002).
- [140] M. F. Romero, C. M. Fulton, and W. F. Boron., *Pflugers Arch.* **447**, 495 (2004).
- [141] F. Jessen, C. Sjøholm, and E. K. Hoffmann, *J. Membr. Biol.* **92**, 195 (1986).
- [142] K. R. Lamont, M. G. Hasham, N. M. Donghia, J. Branca, M. Chavaree, B. Chase, A. Breggia, J. Hedlund, I. Emery, F. Cavallo, M. Jasin, J. Rüter, and K. D. Mills, *J. Exp. Med.* **210**, 1021 (2013).
- [143] A. Renodon-Cornière, P. Weigel, M. L. Breton, and F. Fleury, in *New Research Directions in DNA Repair*, edited by P. C. Chen (InTech, USA, 2013), Chap. New Potential Therapeutic Approaches by Targeting Rad51- Dependent Homologous Recombination, pp. 467–488.
- [144] D. Velic, A. Couturier, M. Ferreira, A. Rodrigue, G. Poirier, F. Fleury, and J. Masson, *Biomolecules* **5**, 3204 (2015).
- [145] C. Esnault, A. Renodon-Cornière, M. Takahashi, N. Casse, N. Delorme, G. Louarn, F. Fleury, J.-F. Pilard, and B. Chénais, *ChemPhysChem* **15**, 3753 (2014).
- [146] B. I. Greene, R. M. Hochstrasser, and R. B. Weisman, *Chem. Phys.* **48**, 289 (1980).
- [147] A. Syage, W. R. Lambert, P. M. Felker, A. H. Zewail, and R. M. Hochstrasser, *Chem. Phys. Lett.* **88**, 266 (1982).
- [148] A. Syage, P. M. Felker, and A. H. Zewail, *J. Chem. Phys.* **81**, 4685 (1984).
- [149] V. Molina, M. Merchan, and B. O. Roos, *J. Phys. Chem. A* **101**, 3478 (1997).
- [150] L. Gagliardi, G. Orlandi, V. Molina, P.-A. Malmqvist, and B. O. Roos, *J. Phys. Chem. A* **106**, 7355 (2002).
- [151] J. Quenneville and T. J. Martínez, *J. Phys. Chem. A* **107**, 829 (2003).
- [152] R. Improta, F. Santoro, C. Dietl, E. Papastathopoulos, and G. Gerber, *Chem. Phys. Lett.* **387**, 509 (2004).

- [153] R. Improta and F. Santoro, *J. Phys. Chem. A* **109**, 10058 (2005).
- [154] J. Saltiel, *J. Am. Chem. Soc.* **89**, 1036 (1967).
- [155] D. C. Todd, J. M. Jean, S. J. Rosenthal, A. J. Ruggiero, D. Yang, and G. Fleming, *J. Chem. Phys.* **93**, 8658 (1990).
- [156] S. T. Repinec, R. J. Sension, A. Z. Szarka, and R. M. Hochstrasser, *J. Phys. Chem.* **95**, 10380 (1991).
- [157] J. Rodier and A. Myers, *J. Am. Chem. Soc.* **115**, 10791 (1993).
- [158] R. J. Sension, A. Szarka, and R. Hochstrasser, *J. Chem. Phys.* **97**, 5239 (1992).
- [159] G. Orlandi and W. Siebrand, *Chem. Phys. Lett.* **30**, 352 (1975).
- [160] D. H. Waldeck, *Chem. Rev.* **91**, 415 (1991).
- [161] B. O. Roos, in *The Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations In Advances in Chemical Physics : Ab Initio Methods In Quantum Chemistry Parts 2*, edited by K. P. Lawley (John Wiley & Sons, Inc., Hoboken, NJ, USA, 1987).
- [162] K. Andersson, P.-A. Malmqvist, and B. O. Roos, *J. Chem. Phys.* **96**, 1218 (1992).
- [163] B. G. Levine and T. J. Martinez, *Annu. Rev. Phys. Chem.* **58**, 613 (2007).
- [164] S. P. Kwasniewski, M. Deleuze, and J. P. Francois, *Int. J. Quant. Chem.* **80**, 672 (2000).
- [165] S. P. Kwasniewski, L. Claes, J. P. Francois, and M. S. Deleuze, *J. Chem. Phys.* **118**, 7823 (2003).
- [166] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [167] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [168] E. Runge and E. K. U. Gross, *Phys. Rev. Lett.* **52**, 997 (1984).
- [169] M. E. Casida, in *In Recent Advances in Density Functional Methods, Part I*, edited by D. P. Chong (World Scientific, Singapore, 1995), pp. 155–192.
- [170] J. Catalan and J. Saltiel, *J. Phys. Chem. A* **105**, 6273 (2001).
- [171] W. G. Han, T. Lovell, T. Q. Liu, and L. Noodleman, *Chem. Phys. Chem.* **3**, 167 (2002).
- [172] C. Angeli, R. Improta, and F. Santoro, *J. Chem. Phys.* **130**, 174307 (2009).
- [173] Y. Lei, L. Yu, B. Zhou, C. Zhu, Z. Wen, and S. H. Lin, *J. Phys. Chem. A* **118**, 9021 (2014).

- [174] Y. Lei, S. Wu, C. Zhu, Z. Wen, and S.-H. Lin, *Int. J. Photoenergy* **2014**, 132149 (2014).
- [175] A. J. Neukirch, L. C. Shamberger, E. Abad, B. J. Haycock, H. Wang, J. Ortega, O. V. Prezhdo, and J. P. Lewis, *J. Chem. Theory Comput.* **10**, 14 (2014).
- [176] N. Minezawa and M. S. Gordon, *J. Phys. Chem. A* **115**, 7901 (2011).
- [177] A. Charaf-Eddin, A. Planchat, B. Mennucci, C. Adamo, and D. Jacquemin, *J. Chem. Theory Comput.* **9**, 2749 (2013).
- [178] P. Jakobsen and R. W. Horobin, *Stain Technol.* **64**, 301 (1989).
- [179] K. Matulef, A. E. Howery, L. Tan, W. R. Kobertz, J. D. Bois, and M. Maduke, *ACS Chem. Biol.* **3**, 419 (2008).
- [180] C. J. Cramer, *Essentials of Computational Chemistry : Theories and Models (2nd Edition)* (John Wiley & Sons, West Sussex, England, 2004).
- [181] F. Jensen, *Introduction to Computational Chemistry*, 2 ed. (John Wiley & Sons, Chichester, England, 2006).
- [182] A. R. Leach, *Molecular modelling : principles and applications*, 2 ed. (Pearson Prentice Hall, Edinburgh, 2009).
- [183] X. Rocquefelte, Ph.D. thesis, Université Nantes, 2001.
- [184] P.-F. Loos, Ph.D. thesis, Université Henri Poincaré, Nancy, 2008.
- [185] A. D. Laurent, Ph.D. thesis, Université Henri Poincaré, Nancy, 2010.
- [186] E. Fermi, *Rend. Accad. Naz. Lincei* **6**, 376 (1927).
- [187] L. H. Thomas, *Proc. Cambridge Phil. Roy. Soc.* **26**, 542 (1927).
- [188] P. A. M. Dirac, *Proc. Camb. Phil. Roy. Soc.* **26**, 376 (1930).
- [189] S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200 (1980).
- [190] J. P. Perdew and Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).
- [191] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- [192] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [193] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **78**, 1396 (1997).
- [194] A. D. Becke, *J. Chem. Phys.* **107**, 8554 (1997).
- [195] C. Adamo and V. Barone, *J. Chem. Phys.* **108**, 664 (1998).

- [196] C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- [197] B. Miehlich, A. Savin, H. Stoll, and H. Preuss, *Chem. Phys. Lett.* **157**, 200 (1989).
- [198] J. P. Perdew, *Phys. Rev. B* **33**, 8822 (1986).
- [199] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Phys. Rev. B* **46**, 6671 (1992).
- [200] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Phys. Rev. B* **48**, 4978 (1993).
- [201] J. P. Perdew, K. Burke, and Y. Wang, *Phys. Rev. B* **54**, 16533 (1996).
- [202] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, *Phys. Rev. Lett.* **82**, 2544 (1999).
- [203] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003).
- [204] T. van Voorhis and G. E. Scuseria, *J. Chem. Phys.* **109**, 400 (1998).
- [205] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994).
- [206] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- [207] A. D. Boese and J. M. L. Martin, *J. Chem. Phys.* **121**, 3405 (2004).
- [208] C. Adamo and V. Barone, *J. Chem. Phys.* **110**, 6158 (1999).
- [209] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.* **120**, 215 (2008).
- [210] T. Yanai, D. P. Tew, and N. C. Handy, *Chem. Phys. Lett.* **393**, 51 (2004).
- [211] J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **10**, 6615 (2008).
- [212] A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).
- [213] J. Tomasi, B. Mennucci, and R. Cammi, *Chem. Rev.* **105**, 2999 (2005).
- [214] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **113**, 6378 (2009).
- [215] R. van Leeuwen, *Phys. Rev. Lett.* **82**, 3863 (1999).
- [216] R. Cammi, S. Corni, B. Mennucci, and J. Tomasi, *J. Chem. Phys.* **122**, 104513 (2005).
- [217] M. Caricato, B. Mennucci, J. Tomasi, F. Ingrosso, R. Cammi, S. Corni, and G. Scalmani, *J. Chem Phys.* **124**, 124520 (2006).
- [218] R. Improta, V. Barone, G. Scalmani, and M. J. Frisch, *J. Chem. Phys.* **125**, 054103 (2006).

- [219] R. Improta, G. Scalmani, M. J. Frisch, and V. Barone, *J. Chem. Phys.* **127**, 074504 (2007).
- [220] J. C. Slater, *Phys. Rev.* **36**, 57 (1930).
- [221] S. F. Boys, *Proc. R. Soc. A* **200**, 542 (1950).
- [222] S. Huzinaga, *J. Chem. Phys.* **42**, 1293 (1965).
- [223] T. H. Dunning, *J. Chem. Phys.* **55**, 716 (1971).
- [224] T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- [225] A. Schafer, C. Huber, and R. Ahlrichs, *J. Chem Phys.* **100**, 5829 (1994).
- [226] A. K. Wilson, T. van Mourik, and T. H. Dunning, *J. Mol. Struct.* **388**, 339 (1996).
- [227] F. Weigend, F. Furche, and R. Ahlrichs, *J. Chem. Phys.* **119**, 12753 (2003).
- [228] W. J. Hehre, R. F. Stewart, and J. A. Pople, *J. Chem. Phys.* **51**, 2657 (1969).
- [229] W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.* **56**, 2257 (1972).
- [230] J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Am. Chem. Soc.* **102**, 939 (1980).
- [231] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *J. Chem. Phys.* **72**, 650 (1980).
- [232] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, *J. Chem. Phys.* **77**, 3654 (1982).
- [233] M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.* **80**, 3265 (1984).
- [234] A. Maddy, *Biochim. Biophys. Acta* **88**, 390 (1964).
- [235] P. Rothbarth, H. Tanke, N. Mul, J. Ploem, J. Vliegthart, and R. Ballieux, *J. Immunol. Meth.* **19**, 101 (1978).
- [236] M. Ramjeesingh, A. Gaarn, and A. Rothstein, *BBA-Biomembranes* **641**, 173 (1981).
- [237] J. M. Salhany, *Blood Cells, Molecules, and Diseases* **27**, 127 (2001).
- [238] M. Born, *Ber. Dtsch. Phys. Ges.* **21**, 13 (1919).
- [239] F. Haber, *Ber. Dtsch. Phys. Ges.* **21**, 750 (1919).
- [240] I. A. Topol, G. J. Tawa, S. K. Burt, and A. A. Rashin, *J Chem. Phys.* **111**, 10998 (1999).
- [241] T. R. Tuttle Jr., S. Malaxos, and J. V. Coe, *J. Phys. Chem. A* **106**, 925 (2002).
- [242] C. C. R. Sutton, G. V. Franks, and G. da Silva, *J. Phys. Chem. B* **116**, 11999 (2012).

- [243] R. Casanovas, J. Ortega-Castro, J. Donoso, J. Frau, and F. Munoz, *Phys. Chem. Chem. Phys.* **15**, 16303 (2013).
- [244] X. Wu, Q. Zhou, and K. Xu, *Acta Pharmacol. Sin.* **30**, 501 (2009).
- [245] A. D. Laurent and D. Jacquemin, *Int. J. Quantum Chem.* **113**, 2019 (2013).
- [246] C. Hansch, A. Leo, and R. W. Taft, *Chem. Rev.* **91**, 165 (1991).
- [247] F. J. A. Ferrer, R. Improta, F. Santoro, and V. Barone, *Phys. Chem. Chem. Phys.* **13**, 17007 (2011).
- [248] J. Tatchen and E. Pollak, *J. Chem. Phys.* **128**, 164303 (2008).
- [249] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.* **117**, 5179 (1995).
- [250] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *J. Comput. Chem.* **26**, 1668 (2005).
- [251] A. D. M. Jr., D. Bashford, M. Bellott, J. R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).
- [252] N. Foloppe and A. D. MacKerell, Jr., *J. Comput. Chem.* **21**, 86 (2000).
- [253] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- [254] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell Jr., *J. Comput. Chem.* **31**, 671 (2010).
- [255] J. Huang and A. D. MacKerell Jr., *J. Comput. Chem.* **34**, 2135 (2013).
- [256] G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou, and T. A. Halgren, *J. Comput. Chem.* **23**, 1515 (2002).
- [257] P. Ren and J. W. Ponder, *J. Comput. Chem.* **23**, 1497 (2002).
- [258] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, *J. Chem. Theory Comput.* **9**, 4046 (2013).
- [259] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, *J. Phys. Chem. A* **105**, 9396 (2001).

- [260] D. W. Hofmann, L. Kuleshova, and B. D'Aguanno, *Chem. Phys. Lett.* **448**, 138 (2007).
- [261] B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* **27**, 1208 (1957).
- [262] J. A. McCammon, B. R. Gelin, and M. Karplus, *Nature* **267**, 585 (1977).
- [263] Y. Duan and P. A. Kollman, *Science* **282**, 740 (1998).
- [264] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan, and B. Towles, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, SC '09* (ACM, New York, NY, USA, 2009), pp. 39 :1–39 :11.
- [265] K. Lindorff-Larsen, P. Maragakis, S. Piana, and D. E. Shaw, *J. Phys. Chem. B* **120**, 8313 (2016).
- [266] J. Verlet, *J. Phys. Rev.* **159**, 98 (1967).
- [267] R. W. Hockney, *Methods Comput. Phys.* **9**, 136 (1970).
- [268] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- [269] W. G. Hoover, *A. Rev. Phys. Chem.* **34**, 103 (1983).
- [270] S. Nosé, *Mol. Phys.* **52**, 255 (1984).
- [271] W. G. Hoover, *Phys. Rev. A* **31**, 1695 (1985).
- [272] R. Kubo, M. Toda, and N. Hashitsume, in *Statistical Physics II : Nonequilibrium Statistical Mechanics*, edited by Springer (2nd ed., New York, 1991).
- [273] H. C. Andersen, *J. Chem. Phys.* **72**, 2384 (1980).
- [274] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, *J. Chem. Phys.* **103**, 4613 (1995).
- [275] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- [276] M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.* **112**, 8910 (2000).
- [277] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.* **91**, 6269 (1987).
- [278] C.-Y. Yang, H. Sun, J. Chen, Z. Nikolovska-Coleska, and S. Wang, *J. Am. Chem. Soc.* **131**, 13709 (2009).

- [279] G. Rastelli, A. D. Rio, G. Degliesposti, and M. Sgobba, *J. Comput. Chem.* **31**, 797 (2010).
- [280] J. M. Hayes, V. T. Skamnaki, G. Archontis, C. Lamprakis, J. Sarrou, N. Bischler, A.-L. Skaltsounis, S. E. Zographos, and N. G. Oikonomakos, *Proteins* **79**, 703 (2011).
- [281] T. Hou, J. Wang, Y. Li, and W. Wang, *J. Chem. Inf. Model.* **51**, 69 (2011).
- [282] T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks, *J. Med. Chem.* **47**, 1750 (2004).
- [283] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin, *J. Med. Chem.* **47**, 1739 (2004).
- [284] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrín, and D. T. Mainz, *J. Med. Chem.* **49**, 6177 (2006).
- [285] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, and J. H. Jensen, *J. Chem. Theory Comput.* **7**, 2284 (2011).
- [286] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, *J. Chem. Theory Comput.* **7**, 525 (2011).
- [287] K. S. Watts, P. Dalal, R. B. Murphy, W. Sherman, R. A. Friesner, and J. C. Shelley, *J. Chem. Inf. Model.* **50**, 534 (2010).
- [288] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**, 11225 (1996).
- [289] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, *J. Comput. Chem.* **26**, 1781 (2005).

Thèse de Doctorat

Titouan JAUNET-LAHARY

Modélisation et simulation de nouveaux inhibiteurs de Rad51 au sein d'une protéine de transport

Modelling and simulation of new Rad51 inhibitors within a carrier protein

Résumé

La protéine RAD51 est l'un des acteurs majeurs des mécanismes de la résistance et de la reconstruction des cellules cancéreuses. La modulation de son activité ouvre une nouvelle voie dans la lutte contre le cancer. De nouveaux inhibiteurs potentiels de RAD51 ont été recensés, notamment, les dérivés du stilbène disulfonique (SD).

Dans cette thèse, nous étudions ces inhibiteurs dans différents milieux (solvant et protéine). La première partie de cette thèse est dédiée à l'étude des propriétés physico-chimiques en solution des dérivés SD. À l'aide de spectres optiques expérimentaux et de simulations par calculs quantiques (DFT et TD-DFT), nous avons montré que les SD se présentent sous leur forme dianionique en condition physiologique et la prise en compte du couplage vibronique est cruciale pour simuler des bandes des spectres d'absorption. La seconde partie se focalise sur la modélisation du complexe SD^{2-} en interaction avec la protéine de transport albumine sérique humaine (ASH), qui joue un rôle prépondérant dans le transport de molécule au sein du corps humain. Elle apparaît être un excellent candidat pour acheminer des inhibiteurs SD^{2-} vers les cellules cancéreuses. Sans donnée cristallographique du complexe $ASH-SD^{2-}$, la modélisation moléculaire est le seul outil prédictif utilisable pour obtenir des données structurales, et nous avons associé ici des méthodes de *docking* moléculaire et de dynamique moléculaire. Cette méthodologie nous a permis (i) d'identifier le(s) résidu(s) important(s), (ii) d'évaluer les énergies d'interaction par des calculs MM-GBSA et (iii) d'identifier les sites d'accueil les plus adéquats pour les composés de type SD^{2-} .

Mots clés

Rad51, Albumine Sérique Humaine (ASH), Stilbène Disulfonique (SD), DFT et TD-DFT, Docking, Dynamique Moléculaire

Abstract

The Rad51 protein is responsible for the resistance and rebuild mechanisms of cancer cells. The modulation of its activity therefore opens a new path for cancer therapy. Recently, disulfonic stilbene (SD) derivatives were found to be a new RAD51's inhibitor.

In this PhD thesis, the SD behavior is studied in various environments (in solution and within protein). In the first part, the physico-chemical properties of SD molecules in solution are explored in a joint experimental and theoretical (DFT and TD-DFT) investigation. The latter demonstrates the dianionic nature of SD compounds in physiological conditions and indicates that accounting for vibronic coupling is crucial to reproduce the bandshape of the absorption spectra. The second part is focused on the modeling of SD^{2-} complexes within a carrier protein, the human serum albumin (HSA), which is essential for the drug transport process through the human body. HSA appears to be an excellent candidate to carry SD^{2-} compounds into cancer cells. Without any crystallographic structure of $HSA-SD^{2-}$ complex, molecular modelling is the only predictive tool to obtain structural data. We performed molecular docking and molecular dynamic methodology to (i) identify the key residues, (ii) investigate the complex energies by MM-GBSA calculations and (iii) determine the most potent binding sites to host SD^{2-} derivatives.

Key Words

Rad51, Human Serum Albumin (HSA), Disulfonic Stilbene (SD), DFT and TD-DFT, Docking, Molecular Dynamics